Journal of the Scholarship of Teaching and Learning, Vol. 19, No. 5, December 2019, pp.77-104.
doi: 10.14434/josotl.v19i5.25856

Not All Flipped Classes are the Same: Using Learning Science to Design Flipped Classrooms

Alyssa P. Lawson
University of California, Santa Barbara
a_lawson@ucsb.edu

Caylor R. Davis
California State University, Los Angeles
cdavis38@calstatela.edu

Ji Y. Son
California State University, Los Angeles
json2@calstatela.edu

Abstract: *The flipped classroom has recently become a popular method used in both higher education and K-12 classrooms, yet research has not consistently demonstrated clear benefits of flipping a classroom. Also, any benefits seen might not be from the flipped design itself, but instead from the individual aspects of a flipped classroom (e.g., more active learning in class, more feedback on homework). This suggests that research focused on how to develop the activities and components of the flipped classroom, instead of simply flipping the traditional in-class and out-of-class activities, is critical to flipped pedagogy. These activities (both in and out of the classroom) should be designed based on a theory of learning. We draw upon the Practicing-Connections (PC) Hypothesis, supported by contemporary theories and research in the cognitive sciences, to design instructional activities to promote learning by having students practice making connections between concepts and situations. This paper examines a flipped class based on a theory-driven design versus a basic flipped classroom. The results of this work offer suggestions as to what dimensions of flipping may be important and how to design and evaluate flipped classrooms based on theories of learning.*

Keywords: flipped classroom, active learning, learning theory, instructional design, transfer

The underlying assumption of traditional lectures is this: the delivery of content by a knowledgeable speaker will impact all levels of students' learning including remembering, understanding, applying, analyzing, evaluating, and creating. A recent meta-analysis of STEM courses revealed that the traditional lecture is, on many counts, a poorer pedagogical method compared to active learning strategies (Freeman et al., 2014). Although there is a great range of what is considered "active learning," there is rapidly growing interest in the flipped classroom (or inverted classroom). Although a basic definition of the flipped classroom is one where the activities traditionally conducted in class (i.e., lecture) and out of class (i.e., homework) swap places (e.g., Lage, Platt, & Treglia, 2000), there are cogent arguments that the flipped model's effectiveness comes from being able to utilize class time for active learning that engages higher order cognitive processes (Bergmann & Sams, 2012; Burke & Fedorek, 2017). The theory and study presented here is an effort to move past efficacy tests of the flipped classroom and towards research that is intended to help instructors *improve* the implementation of the flipped model. We will start with a review of these efficacy tests, then develop a theory of implementation based on research on learning and cognition.

Effectiveness of the Flipped Classroom

Past research on how effective the flipped classroom is for student achievement has shown mixed results. A popular method of testing the flipped model design is to compare achievement in a flipped classroom to prior versions of the same course that used more traditional pedagogies (Baepler, Walker, & Driessen, 2014; Galway, Corbett, Takaro, Tairyan, & Frank, 2014; He, Swenson, & Lents, 2012; Hill & Nelson, 2011; Wilson, 2013). Most, but not all, of these types of studies show that flipping a classroom helps students perform better on tests in a variety of domains (e.g., math: Wilson, 2013; chemistry: Baepler et al., 2014; He et al., 2012). Some of these studies show no significant differences between the two methods of teaching on student success. For example, in learning about ecosystems, students in a flipped classroom design had similar exam results to past students who had been taught in more traditional ways (Hill & Nelson, 2011). Graduate students in a flipped public health class showed no significant improvement when compared to prior students in a standard classroom (Galway et al., 2014).

Studies using more controlled methods have also been inconsistent in demonstrating the effectiveness of flipped classrooms. A study compared two separate sections of students enrolled in the same course in the same time period during a one week period; one section was designed in a flipped format for the instruction week and the other was largely traditional during the same week (Deslauriers, Schelew, & Wieman, 2011). Students in the flipped section performed significantly better on questions related to topics taught during the flipped section than students in the traditional lecture. However, other studies, using similar experimental designs, found no differences between flipped classrooms and traditional classrooms (Jensen, Kummer & Godoy, 2014; Yong, Levy, & Lape, 2015). These studies demonstrate the difficulties inherent in bridging research and instructional practice: most instructors are interested in how well students learn material that is complex, and mastered over the course of many weeks, but this prolonged study is often difficult to examine in brief, highly controlled experimental designs. Even when these studies are conducted, often with great difficulty and commitment, the results are difficult to translate back into instructional practice.

Despite the conflicting evidence on how flipped classrooms affect student performance and learning (e.g., test scores), there have been more consistent results concerning student perceptions of the flipped classroom design and student engagement. Many studies report students having positive experiences with course materials and positive attitudes towards flipped classrooms in general (Galway et al., 2014; Gilboy, Heinerich, & Pazzaglia, 2015; He et al., 2012; Hill & Nelson, 2011; Nouri, 2016; Wilson, 2013). Students' overall engagement and attendance increased when introduced to the flipped classroom design (Baepler et al., 2014; Deslauriers et al., 2011). Student reactions tended to be positive towards the flipped method, although there are documented cases of negative attitudes towards the use of technology (Jensen et al, 2014). However, this mostly positive perception may be in part a bias in the published research; researchers and instructors who have made the effort to flip a course may be less likely to go through the rigors of publication to report students' negative flipped experiences.

There is also evidence that the different components of the flipped classroom are independently beneficial to students. Meta-analyses reveal the benefits of using videos and online materials not in the context of a flipped classroom (Cohen, Ebeling, & Kulik,1981; Means, Toyama, Murphy, Bakia, & Jones, 2009). In the majority of studies investigated by Cohen et al. (1981), instruction that included visually-based instruction (e.g., videos) resulted in better student achievement than traditional classrooms. Videos that involve interactive participation by participants - including being able to stop, move forward, and move backwards - allowed students to gain more information than videos that moved from start to finish with no interaction (Schwan & Riempp, 2004) or from textbooks with the same information (Merkt, Weigand, Heier, & Schwan 2011). Interactive activities, the key face-to-face component in a flipped classroom, have also shown to benefit learning independent of flipped classroom designs (Schwan & Riempp, 2004; Merkt, et al., 2011).

As more researchers and instructors get involved in designing flipped classrooms, it is clear that the efficacy studies that examine whether the flipped model has any effect over traditional methods has little direct impact on how we design instruction and course structures. The best these studies do is convince instructors to try out a flipped design. As technology improves and the limitations of lectures are documented, the scholarship of teaching and learning in flipped classrooms needs a shift in research focus: from "should we flip the classroom?" to "*how* should we flip the classroom?" What aspects of a flipped classroom are important? What are critical components of the design? To begin to get at the *how* question, we must move towards comparisons of flipped classrooms to other versions of flipped classrooms. The research we report here examines how two different implementations of a flipped statistics class (taught by the same instructor) can lead to markedly different effects. Our results indicated that the focus on flipping itself may be misplaced because the details of *how* a class is flipped are vitally important.

Using Learning Science to Design the Flipped Classroom

Often the design of flipped classrooms is heavily influenced by available technologies (e.g., video lectures) and forms (e.g., problem-based or team-based learning). The question of *how* a flipped class should be designed should be driven by theories of how learning works rather than how flipped courses are typically implemented. As psychologists, we propose developing design principles for flipping courses based on research in the cognitive sciences.

As defined by Lage et al. (2000), a basic flipped class simply inverts lecture and homework such that the lectures are moved out of the classroom and textbook problems (typically assigned as homework) comprise the in-class activities. The great opportunity provided by the flipped format is that face-to-face time does not have to be spent on traditional lectures delivering content. Face-to-face time can instead be used for time intensive, higher-order cognitive processing, such as creating conceptual connections, evaluating the appropriateness of concepts, analyzing the similarities and differences of various approaches, and applying concepts to novel circumstances. As has been argued by several flipped researchers and practitioners (Bergmann & Sams, 2012; Burke & Fedorek, 2017), this feature, the increase in actively *practicing connections and high level thinking*, is key to the flipped model. In the next section, we briefly summarize the latest theories and evidence in the cognitive sciences to develop this design hypothesis. Later, we present an in-depth look at a single lesson to contrast a basic flipped class with a Practicing-Connections (PC) flipped class designed to maximize the use of in class time for higher order cognitive practices. Finally, we report the results of a quasi-experimental study to examine the PC hypothesis.

Developing the Practicing-Connections (PC) Hypothesis

Although most educators commonly believe that procedures need to be practiced to achieve fluency, practice is often not considered part of acquiring concepts and conceptual skills such as thinking, reasoning, and analyzing. Most often, a lecture is intended to communicate a well thought out sequence (e.g., a proof or a narrative about scientific discovery) and hearing/comprehending the lecture is considered equivalent to thinking about the concepts. However, modern theories of cognitive science (e.g., Damasio, 1989; Barsalou, 1999; Glenberg, 1997), and in particular those associated with the "embodied cognition" perspective, posit that concepts are actually more like processes where mental representations are dynamically tailored to the current needs of some situation

(Barsalou, 2003). If concepts are actually more like skills, then in order to "learn a concept" we would need to practice tailoring mental representations to fit different situations appropriately. This is the practicing connections (PC) hypothesis: the act of connecting concepts to appropriate structures, situations, and contexts *is* the process of learning (Son, Ramos, DeWolf, Loftus, & Stigler, 2018). To flip a course using this design hypothesis, we would use both in- and out-of-class activities to help students engage in practicing connections.

From research on expertise, we know that experts often have a lot of practice, (see Chase & Simon's 10-years-of-training observation, 1973) but also their practice is varied in a particular way. The development of expertise seems to include practice that is highly responsive to outcomes, errors, and violated expectations (Chase & Ericsson, 1981; VanLehn, 1989). In these studies, people who actively engage in trying different strategies and improving their previous iterations seem to develop more skill and expertise. Deliberate practice is defined as repeated exercises where the goal is improvement, (unlike a test where the goal is achievement or repeated practice where the goal is fluency) and these exercises must be varied, more complex over time, and should not result in early automatization (Ericsson, Krampe, & Tesch-Romer, 1993). If the goal of a classroom is to produce expertise in thinking, then any "practice" should be deliberate practice. In order to practice making connections in this deliberate way, students should engage in specific tasks invented to get feedback on connections, overcome conceptual errors, and tailor concepts for multiple contexts. The practice of making connections cannot mean just making the same connections repeatedly but this activity should vary and increase in complexity.

If we want deliberate practice in making connections to be the main focus of in-class time, we must change our instructional practices to be aligned with this goal. Unfortunately, there are several barriers that make this difficult. We will focus on barriers that have been researched in context of math learning, but these limitations are often present in other domains as well. One major barrier is a maladaptive belief on the part of students, that learning math is about memorizing steps and practicing procedures (Garofalo, 1989; Stigler, Givvin, & Thompson, 2010; Schoenfeld, 1989), not about sense-making and reasoning. In response to the question, "What does it mean to be good at math?" community college students commonly gave responses such as, "In math, sometimes you have to just accept that that's the way it is and there's no reason behind it" (Stigler et al., 2010). Perhaps in part because students expect it, US teachers for the most part teach math as a set of isolated procedures that need to be memorized and practiced (Stigler & Hiebert, 1999). A contributing factor might be that instructors themselves may implicitly think of learning math in a manner similar to students. Many US teachers cannot explicitly explain connections, for example, between fraction division and the contexts in which such an operation would be useful (Ma, 1999).

Beyond just beliefs about mathematics, maladaptive beliefs about the nature of intelligence may also pose barriers (Blackwell, Trzesniewski, & Dweck, 2007; Dweck, 2007). The mindset students hold about their intelligence and/or math ability can affect their ability to perform well while learning and prevent them from absorbing new information. A student who believes that intelligence, generally, or math ability, in particular, are fixed, and therefore unchangeable, might be less inclined to engage in a process of improving their mathematical thinking.

But beliefs aren't the only barriers. Thinking and connecting are processes that require a lot of cognitive resources. In other words, thinking is hard. Research has demonstrated that people who are anxious (Beilock & Carr, 2005), impoverished (Mani, Mullainathan, Shafir, & Zhao, 2013), or otherwise saddled with a cognitive load (Sweller, 1988) are less able to engage in thinking. And to close out this less-than-exhaustive list, a final barrier to practicing thinking is that it takes up a lot of time. In order to make room in the classroom to practice thinking, some learning must be off-loaded to out-of-class time in order to "cover" the required curriculum.

The PC hypothesis is borne from theories of cognition that emphasize the importance of practicing connections and considers the cognitive barriers to such practice to drive the design of a flipped classroom. A basic flipped statistics class would simply move the lecture out of class (e.g., through video lectures) and bring textbook practice problems into class time. In contrast, a PC flipped class attempts to design activities both in and out of the classroom to facilitate the deliberate practice of making connections and address barriers to such thinking.

Based on research in cognitive science (for more details on the specific studies and theoretical approaches that have informed this hypothesis see Fries, Son, & Stigler, in preparation), the PC hypothesis suggests that we must create three types of learning opportunities for students: (1) making explicit connections between concepts, (2) productive struggle (Kapur, 2008) in making those connections, and (3) deliberate practice with connections where the tasks are varied and increasing in complexity over time. In a PC flipped class, instructors and instructional designers must identify the explicit connections for students to make between concepts and representations, contexts, and core conceptual structures rather than come up with a set of disparate learning outcomes. The design of activities (both in and out of the classroom) should scaffold and promote making those connections rather than those connections being told to students. Finally, the connections must be practiced in increasingly difficult and varied circumstances rather than being made only once or practiced in a way that promotes fluency.

Designing a Class around a PC Hypothesis

So how did we translate these three learning opportunities into a design of an actual statistics class? First, we made learning explicit connections the central goal of the class. All of the in-class activities and exams were re-designed to be authentic research situations where students would have to *practice* adapting concepts to fit the particular situation at hand. The out-of-class materials were designed to provide pre-training (Mayer, 2001) of the individual concepts that would be connected through the in-class activities. This was achieved by having students watch online videos while filling out guided notes that went along with the videos (all guided notes and links to the corresponding videos are available here: http://bit.ly/statsteach).

To provide opportunities for productive struggle, we utilized instructional technologies such as clickers and writing worksheets to give students low stakes opportunities to make connections and receive feedback. We also implemented cycles of productive struggle into the instructional routine; at the beginning of class, students could make connections from a more limited set of choices, but at the end of the class session students were challenged to produce their own explanations with less scaffolding.

Deliberate practice was implemented by designing connecting activities in a progressive sequence to be more complex and varied in systematic ways over time. One way this was implemented was that the in class activities were cumulative, such that every statistical technique students previously learned was folded into the next week's activities. For instance, if students learned about *t*-tests in one week, then in the lessons on new information that followed, students would still have to consider whether a *t*-test would be appropriate for some situations. This is unlike many math and statistics courses wherein, if the focus of the week is regression, students only practice creating regression models. In the PC flipped course, students constantly had to adapt their prior learning as they were tasked with selecting the appropriate analysis from their newly acquired repertoire of statistical tests.

At the same time, the class was designed to remove or address barriers that prevent students from focusing on making connections. Activities in class explicitly addressed cultural beliefs about mathematics and their mindsets about the nature of intelligence. We also off-loaded peripheral course material (that must be "covered" by the class) to outside class time to give students time and space in class to engage in making connections. Additionally, we provided social supports to encourage making connections in low-stakes situations so that students could learn what it felt like to be wrong and to see the value of testing out an idea.

As with all translation of research into practice, the details matter immensely. Even though the design philosophy is based on research and theory, every lesson made many design decisions based on a number of assumptions. Because we know these details matter to instructors, we have made all in-class activities available (as worksheets) at this website: bit.ly/statsteach. In the following section, we will provide a detailed look into the design of one PC flipped lesson sequence to see how a theory of implementation might guide actual implementation. To start, we will provide a contrast: the same lesson taught in the basic flipped class and the PC class.

Basic versus Practicing-Connections Lesson Design

The aim of this work is to advance a theory of implementation based on research. We enter into this as both researchers, deeply versed in the laboratory research of our field (cognitive sciences), and instructors, grappling with the problem of coordinating teaching actions over time and across a whole class. Even though we will delve deeply into just one lesson sequence (out-of-class and in-class) in each of the flipped models (basic and PC) here, this in-depth example is meant to illustrate the design demands on instructors trying to teach productively with the flipped model.

The lesson sequence described here will tackle something covered in all statistics courses: the arithmetic mean. In both the basic and PC flipped lesson sequence, students were assigned to watch the same lecture videos that week (a 15-minute lecture on visualizing distributions in plots, http://bit.ly/lecturevideo_visualizations, a 32-minute lecture on central tendency in distributions, http://bit.ly/lecturevideo_centraltendency, and a 31-minute lecture on measures of variability, http://bit.ly/lecturevideo_variability). In the basic flipped course, there was a 10-minute open note quiz at the beginning of each class meant to encourage students to watch the videos and take notes (the quiz is provided in Appendix A). Most of the questions were drawn from verbatim statements made in the videos.

In the PC flipped course, instead of merely encouraging compliance, that is, watching the assigned videos, we conceptualized the out-of-class activities as "pre-training" (Mayer, 2001), where students would gain initial familiarity with the names and characteristics of key concepts. That would facilitate the in-class focus on creating connections between these concepts. The PC flipped class replaced the quiz with guided notes, which are outlines of the lecture with blank spaces for students to fill in with presented content, and have been shown to facilitate students' recall of complex information (Austin, Lee, Thibeault, Carr & Bailey, 2002; Neef, McCord, & Ferreri, 2006). These guided notes were due at the start of class and checked for completion, and then utilized by the students as a reference during in-class activities (The associated notes for this lesson are numbered 03a, 03c, and 04a and are available here: http://bit.ly/StatsTeach_Notes).

Note that both the basic and PC flipped designs require a commitment to active learning during the in-class activities. In the basic flipped course, to learn about the arithmetic mean, students completed chapter problems from a well-developed textbook from accomplished statistics education researchers (Watkins, Scheaffer, & Cobb, 2nd edition). Some of these problems were modified so that students could practice data analysis using statistical software. The full set of questions are provided in Appendix B.

If we zoom in further at just the parts of the lesson aimed at teaching students about the arithmetic mean, we can see that students were asked a variety of calculation oriented questions from calculating the mean and median from small sets of numbers (e.g., problem 1a: 1, 2, 3; problem 1d: 1, 2, 297), adding a new data point and adjusting the mean, and confirming that the residuals from the mean ($x_i - \bar{x}$) add up to 0. They were also asked conceptual questions related to content in the textbook and the videos. For example, one conceptual question asked students to "explain" a critical feature of the mean: "Use the fact that the median is the halfway point and the mean is the balance point to explain why an outlier typically affects the mean more than the median." These questions from the textbook were selected to help students both calculate the mean and consider the implications of using the mean (e.g., the mean is a balancing point of a distribution, the residuals from the mean add up to 0, the mean is more affected by outliers).

Here we pause to consider the way that the basic flipped course was designed. The instructor (the third author) had learning outcomes and selected questions from the textbook that seemed to target those particular learning outcomes. Does the basic flipped lesson engage in *practicing connections*? Some of the questions, even in the basic flipped class, could have helped students make connections between concepts (there were plenty of questions that contrasted the mean with the median). But there were many connections that were not made. For example, the questions that asked students to consider the mean as a balancing point and that the residuals sum to 0 were two separate questions. In the PC flipped course, we will return to connecting these implications of the mean to each other. In the basic flipped course, there was some opportunity for struggle, but because these concepts were not scaffolded, most of the students either copied explanations from the textbook or waited until the class discussion to write down their responses. And finally, there was no opportunity for deliberate practice -- to engage with concepts like *balancing* and *residuals from the mean* in more complex and new ways.

Turning to the in-class activities of the PC flipped course, the lesson on the arithmetic mean was also guided by an in-class worksheet. The worksheet was developed by the instructor (found in Appendix C), and included a series of scenario-based questions. Most of the questions were conceptual inquiries, with minimal questions requiring procedural calculations (as these were assumed to be covered by the preparatory out-of-class activities). All of the questions were designed to assist in practicing connections between the central concepts related to the mean and, particularly, focusing on the principal goal of the lesson, which is to understand why the mean is so useful for describing central tendency. To achieve this, questions about the residuals adding up to zero and the mean being the balance point of those residuals were asked in context of one another, and also set in multiple and seemingly dissimilar scenarios (about measuring height by rounding to the nearest foot instead of inch, about cow weight guesses on the internet).

To examine the ways in which the PC flipped lesson was able to encourage practicing connections, we can take a closer look at some of the specific sequences in the lesson. To begin, there are several opportunities to practice making explicit connections between concepts, as the students were guided through varied scenarios that, on the surface, do not appear to be connected, but all have the same underlying concepts embedded within them. For instance, to practice connecting the concept of the mean as the balance point of the residuals, the students begin with a small data set (5, 5, 5, 10, 20) and are given the mean (9) and the median (5). They are then asked to explain how the mean and the median are both trying to describe the "middle" of this distribution in different ways. Furthermore, when explaining how the mean is the "middle" they are asked to consider all the *differences* from the data points to the mean. As they compute all of those differences, it becomes apparent that the negative differences are balanced out by the positive differences and sum to zero, and hence, they begin to connect the idea of the mean as a measure of central tendency with the feature that it uniquely balances the residuals around it. As seen in Figure 1a, we have depicted the data set as a dotplot and shown how the negative residuals balance the positive residuals. Because this is a small concrete data set, students are able to engage in productive struggle as they grapple with the questions initially on their own and with their peers with minimal guidance from the instructor.

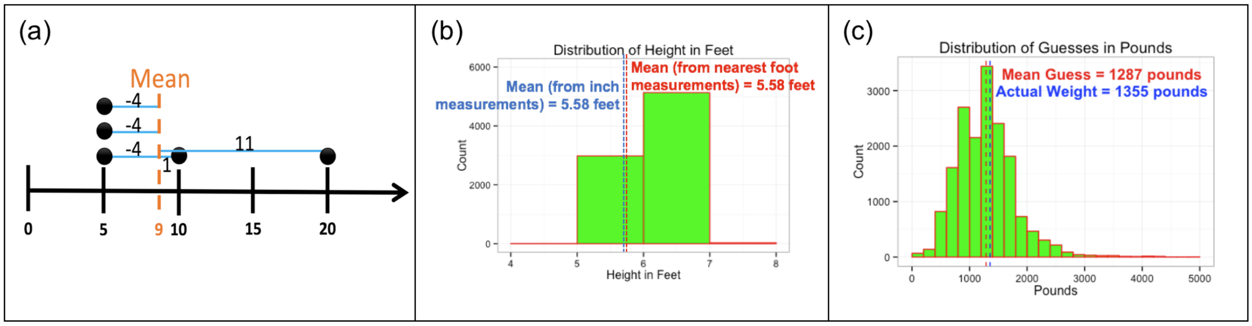


Figure 1. (a) This figure is similar to one that eventually gets developed by students in the course while explaining how the mean is the middle of a distribution.
(b) This figure accompanies the question: "Why is the mean of such awful measurements -- heights rounded to the nearest foot -- so similar to the mean of better measurements -- heights rounded to nearest inch?"
(c) This figure shows the distribution of guesses of how much a cow (in a photograph) weighs. Students are asked to explain why the mean of these guesses are so similar to the actual weight of the cow. All three activities require students to adapt their idea of the mean balancing the residuals to different circumstances in increasingly challenging ways.

This exercise is followed by an example of "measuring heights badly," where students first get practice with trying to visually mark their estimate of the mean on a histogram of heights in inches by considering what point along the x-axis would balance the distribution on either side like a fulcrum. They are then asked whether they would consider measuring height in feet, as in rounding to the nearest foot, to be as good of a measurement as measuring height by rounding to the nearest inch or centimeter. Students will typically respond by saying that rounding to the nearest foot is not as precise of a measurement so it would not be a good way to measure height. The class then tests it out by taking a few students' heights and rounding them to the nearest foot, and then comparing the average of their heights in feet to the average of their heights in inches. Then students are shown an entire distribution made up of these nearest foot measurements and nearest inch measurements. They are shown that the means of the two distributions are quite similar (see Figure 1b). Upon seeing how similar the two means are, and discussing how the times we rounded up were balanced by the times we rounded down, they are encouraged to connect this to the concept of the mean as a balance point of the residuals, as explicitly delineated in the previous example. This provides deliberate practice via an unexpected encounter with this concept, but allows for an explicit connection to be made between the experience and the major learning objective.

Following this class exercise in the worksheet, a new scenario is introduced: a survey that had been conducted by National Public Radio's (NPR) Planet Money. The survey had asked visitors to guess the weight of a cow pictured on their website. The students are also asked to view the picture and take a guess at the weight of the cow. They are also asked to consider things such as: "How accurate could we get by guessing (especially since most people do not know much about cows)? What would you consider accurate?" Students are shown a histogram of the over 17,000 guesses and that the resulting average guess, 1,287 pounds, was pretty close to the actual weight of the cow, 1,355 pounds (see Figure 1c). The students are asked to grapple with the question of why the average guess was so close even though the guesses were individually so terrible. This is now their third attempt to connect the idea of the mean balancing the residuals to a situation: the overestimations were balanced out by the underestimations. Students are asked to make deep connections between these diverse examples of the mean balancing the residuals: with simple sets of numbers, with terrible measurements, and with guesses from uninformed respondents.

Thus, the goal of the in-class worksheet is to guide students in achieving the advanced learning objectives through practicing conceptual connections across multiple contexts. Yet, pertinently, the connections are not easy for students to make across different contexts. They are presented as seemingly unrelated scenarios with fundamental dissimilarities (e.g., sources of "error" differ), and thus the connections come as a surprise. The practice feels difficulty and thus is an opportunity for deliberate practice. Such practice enhances students' capacity to generalize to new contexts. Embedded throughout the lesson, there are also moments to engage in productive struggle where students struggle answering the questions on their own or with peers, try leveraging what they do understand in a novel situation, and test bad ideas along with good ones. In sum, this lesson incorporates the three features of the Practicing-Connections hypothesis: (1) making explicit connections between concepts, (2) productive struggle in making those connections, and (3) deliberate practice with connections where the tasks are varied and increase in complexity over time. These qualities, combined with a sharp focus on the advanced learning objective, grants students a chance to succeed in achieving that learning outcome; not to repeat rote calculations of the mean or adding up the residuals and getting 0, but to understand more coherently how these concepts connect to a variety of contexts.

Comparing Outcomes from Basic versus Practicing-Connections Approaches

Now we will zoom out from the specific lesson and evaluate these designs at a broader level by comparing the student outcomes for each of these flipped design approaches: basic versus PC. We will compare measures such as exam scores, shifts in mindset about intelligence and math ability, and growth in general quantitative reasoning. The section that follows will attempt to quantitatively assess the efficacy of each of these two approaches so that we can draw more comprehensive conclusions about how to successfully modify a flipped classroom model based on theories of learning.

Method

Course Description and Participants

The course that was flipped was called "*Inferential Statistics for Psychology*" and was considered an upper level statistics course that "provide[d] an introduction to inferential statistics. Topics included: hypothesis testing, probability, parametric (z, t, and F tests, correlation/regression) and nonparametric (Chi-square) procedures, the written presentation of statistical results, and the use of statistical software packages" (from the course catalog). Allotted class time included four lecture hours and three laboratory hours per week. The course is a core course in the Psychology major, meaning that all students in the major are required to take this course. Psychology students commonly cite the difficulty of statistics, lack of use in future career, and dislike of math as contributing to their negative attitudes towards statistics courses (Griffith, Adams, Gu, Hart, & Nichols-Whitehead, 2012).

The third author taught two sections of the course (the basic flipped course in the Fall quarter and the PC flipped course in following Spring quarter of the same academic year). This statistics course is traditionally taught as a lecture and lab course with 24 students; the basic flipped course was taught in this format. The PC flipped course increased enrollment to almost 48 students in a lecture with two separate lab sections (24 students each). In total, 72 participants were involved in the study. The demographics of the course reflected the base rates of the major and the university.

To enroll in this class, students were required to pass a prerequisite course (descriptive statistics). Combined across the basic and PC flipped courses, 80% of the students reported taking the prerequisite more than one academic year prior to the current class.

Design and Procedures

In this quasi-experimental design, we examined three outcomes between the basic flipped and PC flipped courses. The two types of flipped courses used the same four lab tests and final lecture exam. There were 10 weeks in the term with finals week (we call it Week 11). The lab tests occurred during the lab section at Week 4, 7, 9, and 11. The final exam occurred during Week 11 during the lecture section. Students in both classes were also administered questions about their implicit theories of intelligence and math ability (popularly called mindsets, Dweck, 2007) at both Week 1 and 11. Finally, the PC flipped class was assessed on a measure of quantitative reasoning at Week 1 and 11.

Measures

We used exam scores, a mindset survey, and a test of quantitative reasoning as measures of performance in this study.

Exams. To assess student performance, we examined test grades. There were 5 tests in all: 4 lab tests and 1 essay final exam. The lab tests, administered in a computer lab, were designed to mimic what experts do when confronted with a data set. Students were provided with a data set and a brief description of why and how the data were collected and the variables that were measured. The data and context were all novel situations that had never before been discussed in class, so these were all transfer tests rather than tests of memory. Students were asked 3-5 research questions (e.g., "Does caffeine affect athletic performance?") and were asked to answer them using the provided data set in APA format. These lab tests examined students' ability to connect statistical knowledge to novel contexts. Because they were transfer tests, students were allowed to use any notes or other materials they wanted to bring to the computer lab. This also mimics data analysis in valid situations (most researchers have access to all their notes and the internet when conducting data analysis). Examples of these tests can be found at bit.ly/statsteach. The four lab tests were worth 2%, 4%, 7%, and 14% of the grade for the course. The final exam consisted of 16 short essay questions that asked students to explain various statistical concepts. Example short essay questions are available in Appendix D. The final exam was worth 25% of the course grade.

Mindsets. Students' mindsets about their ability to grow in intelligence and math ability were measured both before and after completing their respective class (the basic flipped or PC flipped course). There were six statements that students rated on a six-point scale (levels of agreement/disagreement). Half of these statements were about intelligence and the other half were about math ability. Intelligence growth mindset was assessed by students' agreement with statements such as "You can always substantially change how intelligent you are" and "You can learn new things, but you can't really change your basic intelligence" (Blackwell, Trzesniewski, & Dweck, 2007). Similarly, math ability growth mindsets were assessed by questions such as "You can change your basic math ability level considerably" and "To be honest, you can't really change how much math ability you have." Negative statements that indicated a fixed mindset were reverse coded such that higher numbers meant more agreement with growth mindset. See Appendix E for all statements (more coding information is provided here: http://bit.ly/mathmindsetquestions).

Quantitative Reasoning. Lastly, quantitative reasoning (based on Stigler et al., 2010) was measured through a 21-question test. This test included almost no calculations, but asked students questions designed to target their understanding of symbols, reasoning with operations, and judgment of magnitude. This test was only administered in the PC flipped classroom design at the beginning and the end of the course. This quantitative reasoning test did not affect students' grades in the course. A copy of the test can be downloaded here: http://bit.ly/quantitativereasoningtest.

Results

Exam Scores

There were four lab tests and a final exam. We conducted ANOVAs to compare test scores between the two flipped classes (basic versus PC) and the results are shown in Figure 2. PC flipped students performed significantly better on the first lab test, $F(1, 70) = 12.01$, $p = .001$, $\eta^2 = .15$ and the final exam, $F(1, 69) = 5.72$, $p = .02$, $\eta^2 = .07$. The two flipped classes were not statistically distinguishable on the second, $F(1, 69) = 3.21$, $p = .08$, third, $F(1,70) = 1.578$, $p = .21$, and fourth lab tests, $F(1, 69) = .11$, $p = .74$.

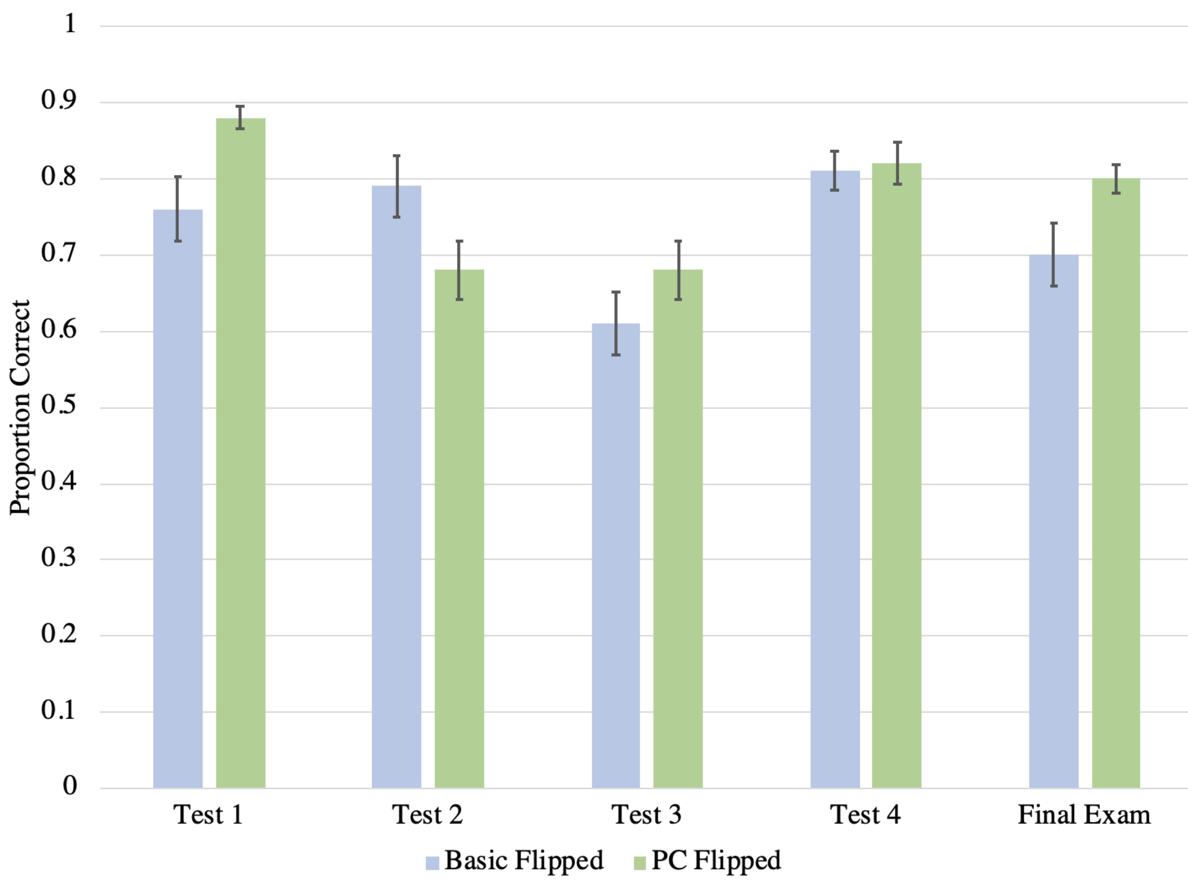


Figure 2. The mean percent correct on each exam for both basic flipped and PC flipped classes.

Mindsets

We calculated mindset change scores by subtracting the pre-class score from the post-class score. A positive score indicates more growth mindset after the course while a negative score indicates a more fixed mindset. A 2 (Flipped Class: basic vs. flipped) x 2 (Domain: intelligence vs. math ability), mixed repeated-measures ANOVA revealed that the type of flipped class had a significant effect on mindset, $F(1,62) = 9.61$, $p = .003$, $\eta^2 = .13$. The PC flipped class experienced a more significant positive shift towards growth mindset in both intelligence and math ability. There was no main effect of domain, $F(1, 62) = 3.35$, $p = .07$, nor was the interaction significant, $F(1,62) = .02$, $p = .89$.

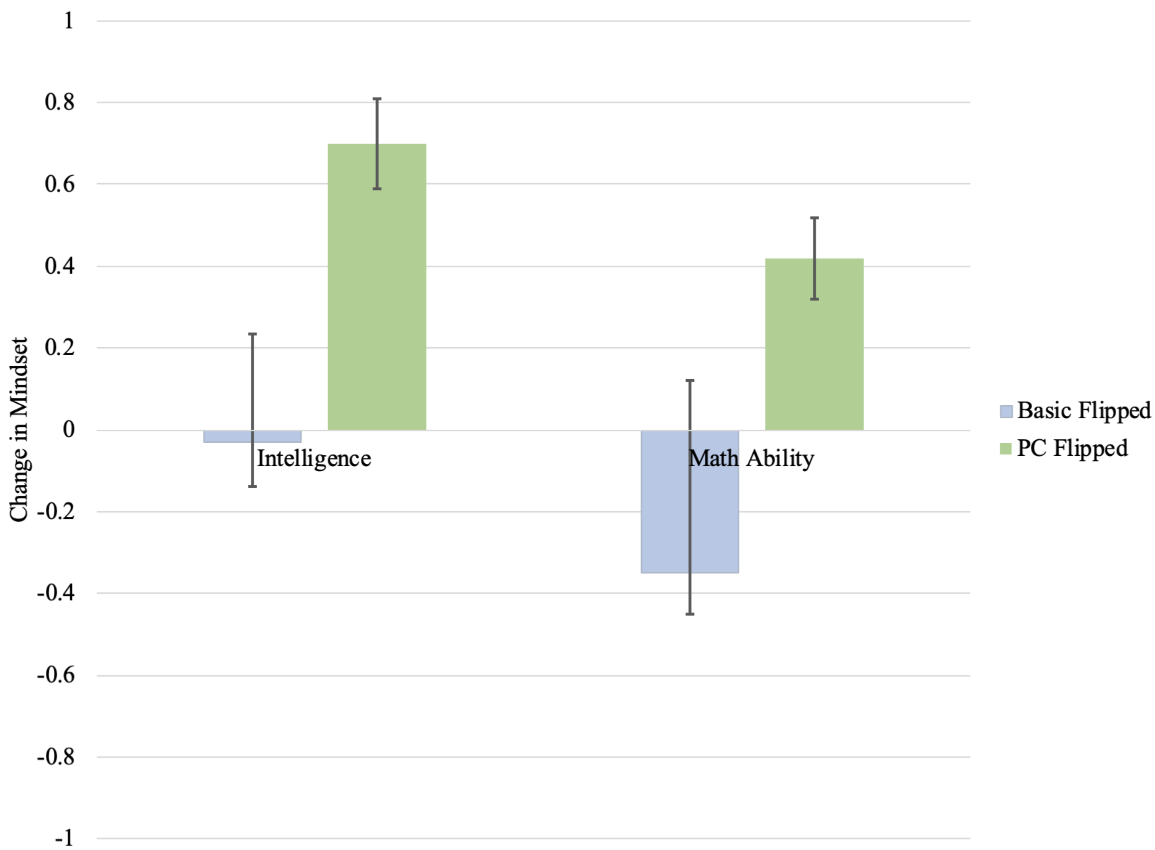


Figure 3. The change in intelligence and math ability mindset for both basic flipped and PC flipped classes.

Quantitative Reasoning

A repeated-subjects ANOVA revealed a significant difference between the beginning of class scores and the end of class scores on the quantitative reasoning exam in the PC flipped class, $F(1, 47) = 12.74$, $p = .001$, $\eta^2 = .21$. This indicates a reliable shift in quantitative reasoning on a test that was not directly related to the course content between the start and finish of the class.

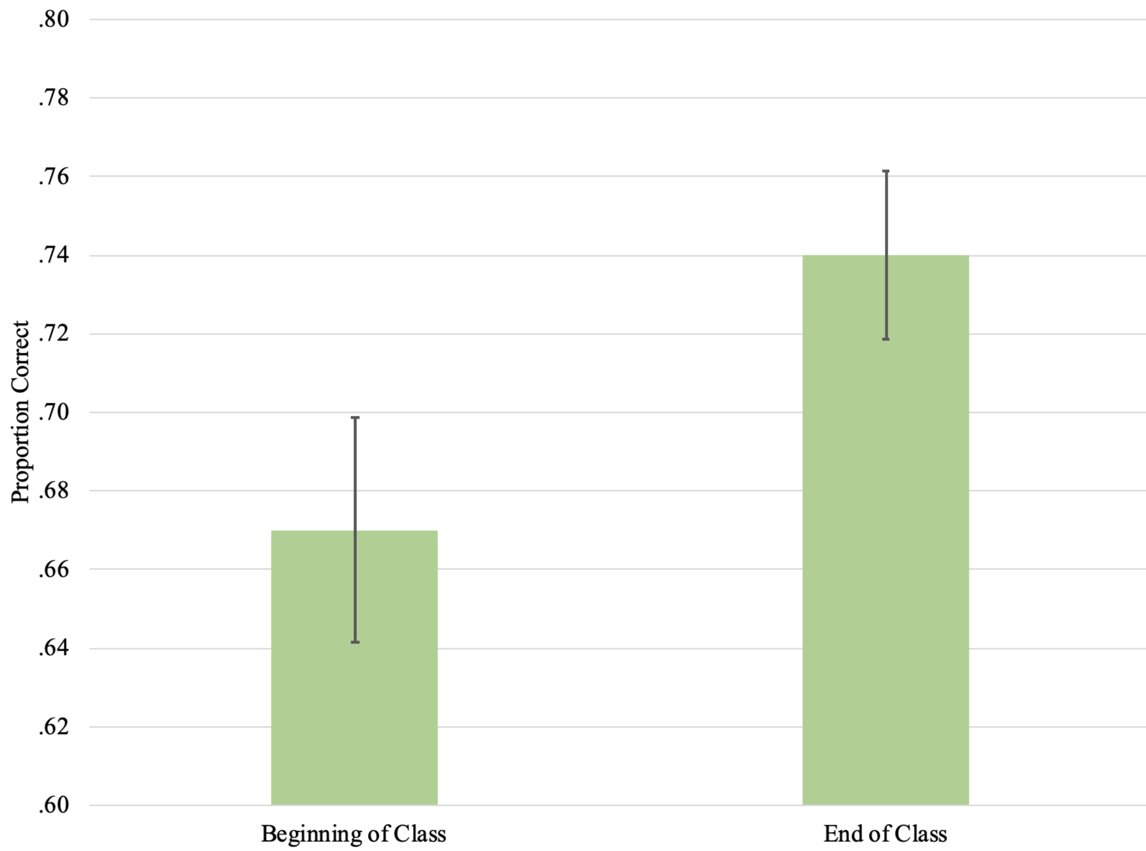


Figure 4. The mean percent correct on quantitative reasoning exam for PC class only.

Discussion

Just as there are better and worse lectures, there are better and worse flipped classrooms. In many ways, efficacy studies compare a nascent flipped classroom to a well-developed culture of lecturing. The time has come to develop a more nuanced view that any pedagogical model is quite broad and provides only the barest outline for possible learner-teacher interaction. There are lots of details that need to be worked out and many of those details may hold the key to better more consistent implementation of flipped models.

In this research, we attempted to develop, implement, and test a set of design principles for flipped pedagogy based on findings in cognitive science. Socioemotional (i.e., mindsets) and cognitive (i.e., course exams and quantitative reasoning) aspects of learning are both important to address in math classes, but mindsets are rarely assessed in a traditional statistics course. The PC flipped class design helped students develop growth mindsets about intelligence and math relative to the basic flipped course. This is an important finding because these mindsets can influence persistence and resiliency in learning and testing (Yeager, & Dweck, 2012). Mindsets may be particularly important in courses that engage students in struggling with difficult concepts.

Although there was a distinct difference in the mindset change between the two flipped classes, the achievement tests only partially demonstrated the benefits of the PC flipped class over the basic one. On a subset of the lab tests and the final exam, which asked students to explain concepts from the course, students in the PC flipped classroom design performed significantly better. Taking a look at the lab tests that did not show a significant difference, we suspect that students needed more practice connecting computer-based activities (using SPSS) with the core concepts of the course. The

PC hypothesis provides a framework to look for missing connections and develop lessons to help students forge those connections in future revisions of course materials. We speculate that the use of software should be interleaved with core concepts earlier in the learning materials. Thus, when students are engaged in lab skills, they will also coordinate them with conceptual knowledge. However, by the final exam, students in the PC flipped course were significantly better at applying concepts from the course because they had practiced making connections rather than actively practicing textbook problems.

Furthermore, students in the PC flipped class increased their quantitative reasoning ability between the beginning of the class and the end of the class. Although this type of reasoning was not addressed directly in class, they were able to show an increased understanding of symbols, reasoning with operations, and judgements on magnitude. Overall, their general ability in mathematics reasoning grew despite not addressing this type of learning directly. However, this result alone is difficult to interpret because the quantitative reasoning test was not administered to the basic flipped class.

What We Learn from Implementing the PC Hypothesis

One of the valuable lessons we can learn from implementing the PC hypothesis is that it provides a framework for an instructional designer's judgment. Whenever an instructor sets out to redesign a whole class (e.g., trying to flip a class), there are innumerable design decisions involved. Because the PC hypothesis focuses on the core instructional goal of engaging students in deliberate practice of making explicit connections, its simplicity, modifiability, and emphasis on deep learning provides a theoretically motivated basis for making many of these decisions. Any usable research-based framework should supply guidance for adapting course materials for different contexts, methods of delivery, and student populations.

The broader goal of using theories of cognitive science to influence course design has been attempted by a number of frameworks. Notable implementation theories include the Knowledge-Learning-Instruction (KLI) (Koedinger, Corbett, & Perfetti, 2012) and Interactive, Constructive, Active, and Passive (ICAP) framework (Chi & Wylie, 2014). These theories offer learning activities best for eliciting a specific cognitive process. Such frameworks offer valuable insight on how to capitalize on the active learning time that a flipped classroom affords and to appropriately align the activities when learning is more passive, particularly, outside of the classroom (since this is an equally important component of a flipped course design). Frameworks like these offer what instructors are often looking for, which is: given some learning outcome, what activities and materials does research suggest will best lead to achieving that outcome?

However, even these highly respected, broadly modifiable, and theoretically motivated frameworks are not often used by instructional designers nor faculty. We speculate that one barrier might be that it takes a significant amount of work to figure out how each learning goal should be achieved in the appropriate way. If these frameworks are used, they are often used with a few lessons or learning outcomes, but not effectively implemented throughout the course. Additionally, since most instructional designers are not cognitive or learning scientists, they may find it more helpful to see an example of a full class implemented with such a framework than to abstractly hear about the framework.

The PC hypothesis takes these considerations and as an accompaniment to this scholarly article, we provide an implementation example in a full set of course materials (available at http://bit.ly/statsteach). These materials can be used in classrooms without the instructor having a background in learning sciences or having a deep understanding of the motivating theoretical framework itself. The theory-driven design makes it possible for materials themselves to embody the

PC hypothesis rather than relying on the hypothesis to be adopted by individual instructors. This makes it possible for instructors to become interested in the learning science-based approach after seeing and using the materials. Even as instructors are gradually learning about the design framework, they can potentially reap the benefits of the embedded emphasis on having students practice making connections.

Although we began with introductory statistics, this framework is general enough for any course. This framework will hopefully engage educators and designers in developing their own judgment for modifying their materials to promote interconnected, flexible knowledge. For instance, a chemistry professor might use the PC hypothesis to consider the key connections that need to be made in their domain. There is considerable research on how connecting multiple representations of nanoscale particles (e.g., connecting particle models to symbols and macroscale phenomena) can help students gain a more coherent understanding of chemistry (Prain, Tytler, & Peterson, 2009; Treagust, Chittleborough, & Mamiala, 2003). The PC hypothesis provides a framework that puts practicing those connections at the focus of instructional design and provides the perspective for varying the connecting activities over time to sustain a level of challenge for the students.

Theories of Implementation for the Flipped Classroom

Recently, there has been an increasing interest in the flipped classroom. However, many of these high quality research programs are difficult to translate into instructional design flexible enough to implement in a different course and in a variety of educational settings because many of them culminate with a test of efficacy ("does the flipped version work significantly better?"). We need more than just tests of efficacy, we need a theoretically-driven design perspective that we can incrementally improve upon. The broader goal of this modest work is to develop a theory of *implementation* so that classrooms can be flipped to maximize student learning.

Our hope as researchers, instructors, and designers of educational materials is to begin with a hypothesis that is able to guide practitioners. If someone reads about a flipped classroom in statistics that was "effective", they would still be left with the question of how to turn the insights from that particular instantiation into concrete actionable designs for their own course content, student population, and institution. We have begun with a few design principles (the PC hypothesis) that we believe can guide the design of course content and instructional routines. Our research goal was to document the instantiation of these design principles to further our understanding of how to flip a course effectively.

Along the way, we have learned that teaching and learning always take place in a system. Flipping the course is only one part of that overall system. Teaching the content of a course cannot be divorced from how we *measure* learning and how we *conceptualize* learning. Our measures of learning must be as current as our teaching techniques. The exams and quantitative reasoning measures used in this study were designed to measure transferable learning and adapting concepts to novel problems and contexts. If we had used traditional exams that ask students about definitions and carrying out procedures (e.g., carry out a t-test), we would not have the data we need to adapt our teaching toward the goal of transferable learning. But we have much further to go in developing assessments that truly measure adaptive understanding of novel situations. Any work in developing better theories and techniques of teaching will be dependent on effective and sensitive assessments.

Another salient lesson is that we need to broaden how we conceptualize learning. As cognitive scientists, we put a heavy emphasis on concepts and mental work. As content experts, we care more about students learning content than about changing attitudes. But all too often, "learning of concepts" is equated to "memorizing" and so practicing concepts is just repeating those concepts ineffectively. The notion of practice is not applied to attitudes and understanding of concepts. The PC hypothesis attempts to directly address those underlying ideas about learning and practice. The notion of "practice" should be applied to help students make explicit connections in systematically varied and increasingly complex ways. Also, attitudes and socioemotional stances must be practiced. A talk on persevering through difficult problems is less effective than practicing those attitudes in situ. The benefit of the flipped classroom is the apportionment of more class time to practice such perseverance.

Closing the Loop of Theory and Practice

In order to move educational practice forward in a data-driven manner, there needs to be more of a connection between developing learning theories and implementing these abstract ideas in a course. An approach like ours is an attempt to create a bridging type of theory, a set of implementation guidelines that connect basic research in cognition with practice in the classroom. Such an approach can reveal where theories are lacking. For example, although we know that productive struggle routines, where a student attempts a problem before hearing a lecture, can work in a 1-hour experiment (Kapur, 2012, 2014), in a real-life classroom, some students might just learn to wait until the lecture and may cease working on the problems independently once they learn that a lecture will be forthcoming. An implementation theory also reveals how common instructional practice can benefit from basic research. For example, the transfer literature (e.g., Gick & Holyoak, 1980, 1983; Gentner, Loewenstein, & Thompson, 2003) has an extensive set of findings that can help us develop assessments that can measure flexible learning.

There are several other approaches to create bridges between basic research and instruction that have influenced this work (Koedinger, 2002; Merrill & Twitchell, 1994). For instructional practice to move forward, we need studies like this one that attempt to design a whole course, from start to finish, based on these principles. Depending on the field and content area, different approaches might provide different benefits. Instructors and practitioners need to see more examples of these theories implemented in classrooms. Ultimately, theory and practice should work in a loop where theories help guide instructional design, and implementation can illuminate issues in need of better theories.

Conclusion

By incorporating the PC hypothesis into a flipped classroom, we found that there were improvements in mindsets, course learning, and quantitative thinking. These results are promising, but there is also room for improvement. Our framework illuminated areas in our course that were still lacking connections for students (e.g., a disconnect between computer-based SPSS experience and core statistical concepts). Not only does this framework help identify areas that may be lacking, but it also guides us in our understanding of what we could do better in future classes (e.g., make more explicit connections between computer-based experiences and core statistical concepts). In our research, we attempted to improve upon the basic flipped course through a design framework grounded in modern theories of learning. Merely implementing a flipped method classroom will not automatically affect student learning. There is nothing magical about the format of flipping itself. It is the process of incrementally improving the flipped classroom that can incrementally improve learning.

As John Dewey said in *Sources of Science Education* (1929), "No conclusion of scientific research can be converted into an immediate rule of educational art. The value of the science… resides in the enlightenment and guidance it supplies to observation and judgment of actual situations as they arise (p. 19, 32)." Much of instructional design must be guided by the expertise of educators about their context, their content domain, and their students. The PC hypothesis allows that expertise to work together with what we know from research on teaching and learning to result in guiding design decisions as they arise.

Acknowledgements

We had valuable guidance and support from Assistant Vice Chancellor Gerard Hanley of the California State University system and his amazing Course Redesign with Technology team, Cal State LA's Center for Excellence in Teaching and Learning, the Cal State LA Learning Lab, and the UCLA Teaching and Learning Lab. We thank Jim Stigler, Cat Haras, and Beverly Bondad-Brown for comments on prior drafts of this manuscript.

Appendix

Appendix A. Quiz (corresponding with assigned videos) for Basic Flipped Class

1. For this set of numbers [1,1,1,6,6] what is the:

   Mode? _________________

   Median? _________________

   Mean? _________________

2. What does this symbol (Σ) indicate?

3. What is the symbol for the mean of the sample?

4. What is the symbol for the mean of the population?

5. There were 11 people in a room who made an average salary of 25K per year. Another person walked in the room. That person was unemployed and made $0 per year. Which of the following statements below is true?
   a. The mean for all 12 people is smaller than the mean of 11 people.
   b. The mean for all 12 people is larger than the mean of 11 people because the sum of 12 numbers is bigger than the sum of 11 numbers.
   c. The mean for all 12 people is larger than the mean of 11 people because 12 is bigger than 11.
   d. None of the above can be verified without knowing what each person's salary is.

6. In this set of numbers, the mean is equal to the median [1,1,3,5,5]. Change one number in the set so that the mean is greater than the median.

   ______ ______ ______ ______ ______

7. Range and inter-quartile range are usually used with which of the following measures of central tendency?
   a. Mean

b. Median
c. Mode
d. None of the above

8. Standard deviation and variance are usually used with which of the following measures of central tendency?
a. Mean
b. Median
c. Mode
d. None of the above

9. Which of the following statements are true?
a. The sum of all squared deviations from the mean = 0.
b. The absolute value of a deviation from the mean = 0.
c. The sum of all deviations from the mean = 0.
d. All of the above are true.

10. What is the symbol for sample standard deviation?

11. What is the symbol for population standard deviation?

12. What is the "sum of squares"?
a. Sum of squared values
b. Sum of squared standard deviations
c. Sum of squared distances from the mean
d. Sum of means
e. Sum of squared means

13. What is the conceptual difference between s and S?
a. s is the estimate of the sample standard deviation from the population distribution and S is the sample standard deviation.
b. s is the estimate of the population standard deviation from the sample distribution and S is the sample standard deviation.
c. s is the sample standard deviation and S is the population standard deviation.
d. s is the population variance and S is the sample variance.

14. Which of the following statements is generally true?
a. $S = s$
b. $S > s$
c. $S < s$
d. $S = \sigma$

Appendix B. Basic Flipped Lesson on Arithmetic Mean

1. Find the mean and median of each ordered list, and contrast their behavior. (Feel free to use excel!)
    a. 1, 2, 3 Mean __________ Median __________
    b. 1, 2, 6 Mean __________ Median __________
    c. 1, 2, 9 Mean __________ Median __________
    d. 1, 2, 297 Mean __________ Median __________

2. Use the fact that the median is the halfway point and the mean is the balance point to explain why an outlier typically affects the mean more than the median.

3. Five children (about 4 ft tall) are standing together when their teacher (who is 6 ft tall) joins the group. What is the new mean height? The new median height?

4. Verify that the sum of the deviations from the mean is 0 for the numbers 1, 2, 4, 6, 9. Find the standard deviation. Explain what that number means.

5. Without computing, match each list of numbers in the left column with its standard deviation in the right column. Check any answers you aren't sure of by computing [you can use excel functions if you wish].
    a. 1 1 1 1 i. 0
    b. 1 2 3 ii. 1.581
    c. 1 2 2 iii. 1
    d. 10 20 20 iv. 0.577
    e. 1 2 3 4 5 v. 5.774

6. The mean of a set of seven values is 25. Six of the values are 24, 47, 34, 10, 22, and 28. What is the 7th value?

7. The sum of a set of values is 84, and the mean is 6. How many values are there?

Essay Question Reflection (Connect to Explanations in Lecture Videos)

8. In inferential statistics, we usually are talking about s (rather than S or σ) when we refer to standard deviation. Why is this the case?What is the conceptual difference between s and σ? (The formulas are not necessary for this answer.)
9. How are the concepts of mean and standard deviation similar? How are they different?
10. How does the formula for s relate to the definition of standard deviation?
11. In the formula for s, why do we divide by the term n-1 rather than n?

Lab practice: Introduction to SPSS

We will be looking at a data set introduced on pg. 33 (Display 2.17): Life expectancies in countries around the world. Use SPSS to explore the shape, center, and spread of this distribution.

1. How to convert excel into spss data file
    a. Download **lifeexpectancies.xlsx** (either from "Notes" in course website or from google drive folder) and convert to an .sav file.

   b. What is the difference between an .sav file and an .spv file?
   c. In general, which one do you need to turn in?
2. How to create a quick histogram in SPSS (use "Graphs → Legacy dialogs → Histograms")
   a. Why does this histogram look different than the one on pg. 33 (Example 2.3)?
3. How to get common descriptive statistics in SPSS (there are a few different ways but let's explore "Analyze → Descriptive statistics → Frequencies")
   a. Mean, median, mode
   b. Standard deviation and variance
   c. Minimum and maximum
   d. Range and interquartile range (well… at least the quartile cutoffs)

Appendix C. PC Flipped Lesson on Arithmetic Mean

Summary Statistics vs. Data Visualizations

Data often comprises many, many numbers. Imagine a study with 10,000 participants. It's hard to keep in mind the results of each participant. Summary measures (such as the mean) are easier to keep in mind. Data visualizations (such as histograms) are easier to explore and think about.

Sometimes people get hung up on the summary measures; these summaries can hide telling patterns. Sometimes people think: all we have to know is the mean and that will be enough information about the *distribution.*

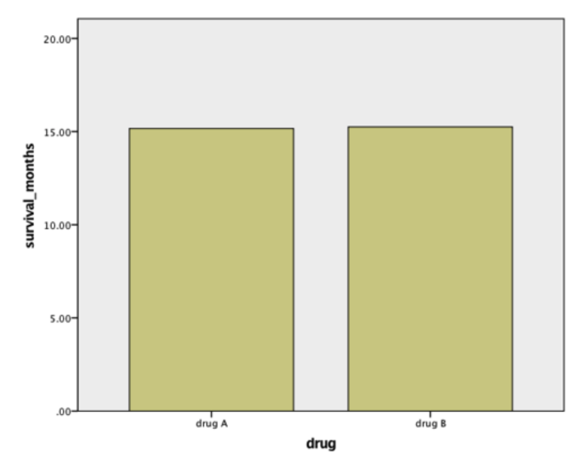
Let's say you are trying to choose between two cancer drugs. Which of these is a better drug?
Here are two cancer drugs. Which is a better drug? (Assume the number means additional months lived.)

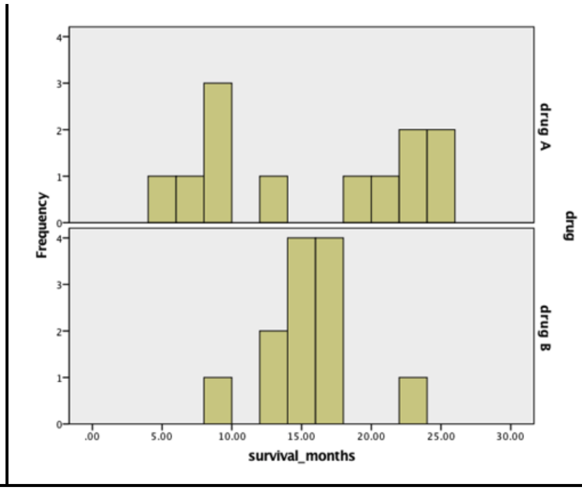
Drug A: 23, 21, 5, 9, 18, 24, 6, 8, 8, 25, 23, 12 (mean is 15.17)
Drug B: 17, 16, 16, 13, 15, 22, 17, 13, 15, 15, 9, 15 (mean is 15.25)

If we just looked at the means, we would think that drug B might be slightly better (the difference expressed as .08 of a month turns out to be about 2.5 days).
1. Why is drug A better than drug B?
2. Why is drug B better than drug A?
3. Imagine someone looked at the means and summarized the two drugs like this: "Drug B basically helps you live 2.5 days longer than drug A." Is this an accurate statement? Why not?

Histograms vs. Bar charts

Figure 1.	Figure 2.





4. Which figure depicts the bar chart? Which is the histogram? How did you know?
5. Which one would be more helpful to a person who is trying to figure out which drug to take? Why?
6. Now remember that the mean for drug A is 15.17 and 15.25 for drug B. On each of the figures above, put a mark where the means should go. Where do the means go on a bar chart? Where does it go on the histogram? How are they different?
7. Let's say the pharmaceutical company tells people, "Well, on average, people live about 15 more months with this drug."
    a. Is that an accurate statement for each drug?
    b. Which drug is better described by the mean? Why?

Figure 3.

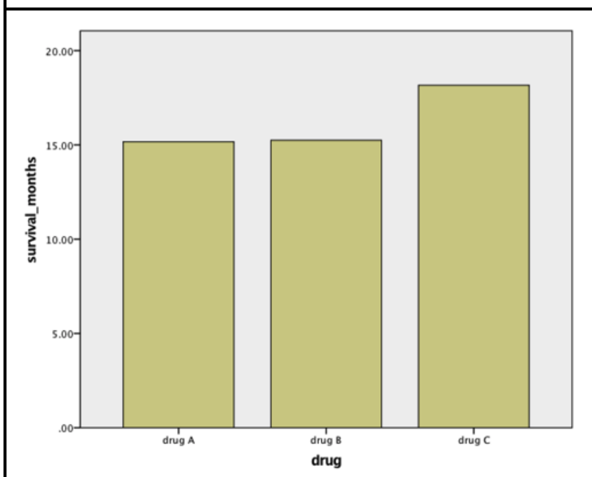
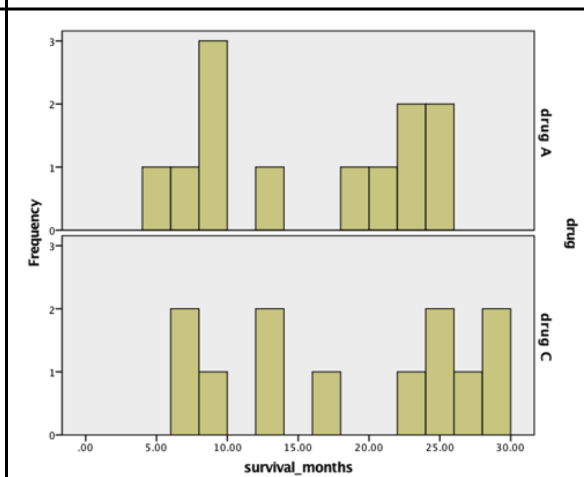


Figure 4.



8. A new drug comes out on the market. Let's call it drug C. Take a look at the bar charts and histograms above. How is it similar to drug A? How is it different from drug A?
9. Without knowing anything else, which seems like the better drug to take (assuming you want to survive a longer period of time)? Why?
10. When it is enough to look at the bar chart that summarizes the distributions? When is it more

helpful to look at distributions with histograms?

Pros and Cons of Means

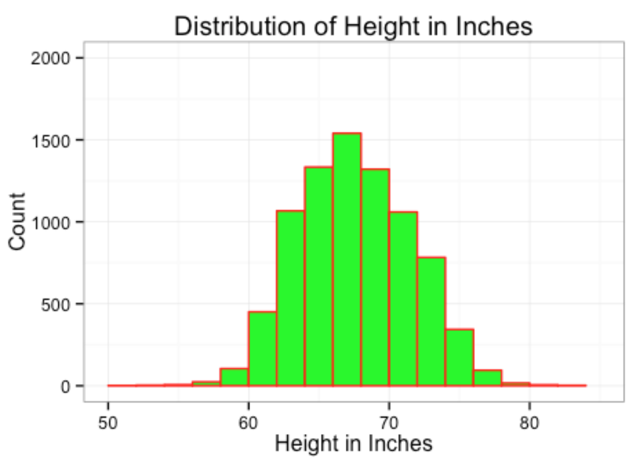
11. You have probably heard that medians are better description of central tendency than means for some types of data. When/why are medians better than means?
12. Even so, we are going to be using means a lot in inferential statistics! You should wonder – why? If medians are sometimes better than means, why are means used so much? What is a mean good for? We are going to cover this in the rest of the lesson today.

What does middle mean?

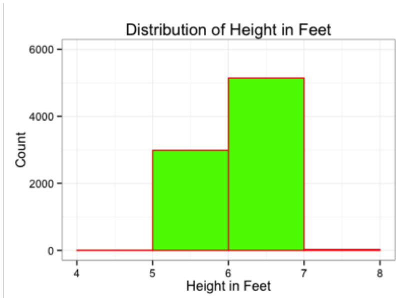
13. Here is a small distribution: 5, 5, 5, 10, 20. The median of this distribution is 5 but the mean of this distribution is 9. Both are trying to describe the "middle" or "central tendency" of this distribution.
    a. How is the median (5) the "middle" of this distribution?
    b. How is the mean (9) the "middle" of this distribution? (Hint: Consider all the *differences* from the data points to the mean…)

Measuring Heights Badly

14. This is some data from the National Longitudinal Study of Youth. They collected data from 8000+ youths (15-20 years old). Here is a histogram of their heights. Mark on the histogram about where you think the mean might be.

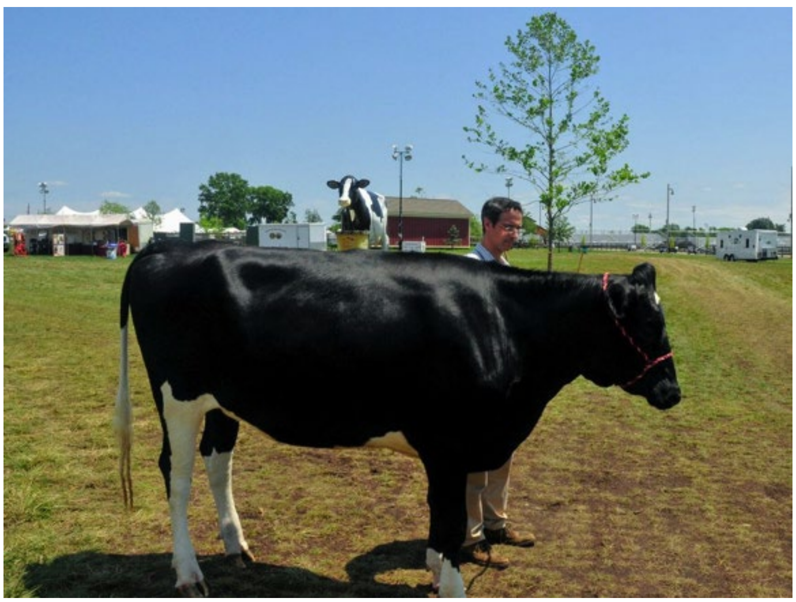


15. The average is 67.05 in (or 5.59 ft). Was your mark near that value? Adjust if necessary.
16. What if we only had measures to the nearest foot? Are these measurements just as good as measures to inches? Why or why not?
17. If we took the *average* of measures to the nearest foot, how accurate would the average be? (To be clear, that's like if we took someone who is 5' 3" and rounded them to 5' and took someone 5' 11" and rounded them to 6' and then took the average of the 5' and 6'… the average of these two people would be 5.5 ft.)
18. Here is the distribution of heights measured in nearest foot. Do you think the average is similar or different from the average height measured in inches?



19. The mean (based on the rounded heights) is _______. Why do you think the two means are ______________, even though the measurement of each individual's height is so much less exact?

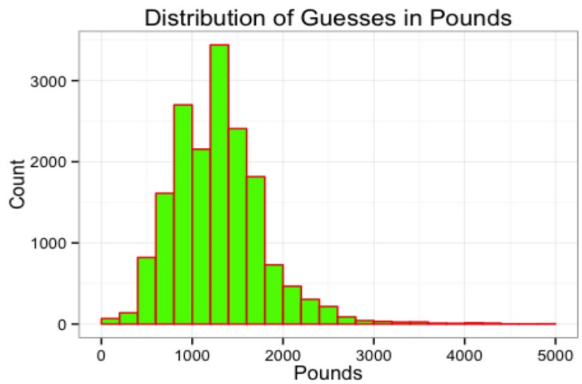
People who know nothing about cows

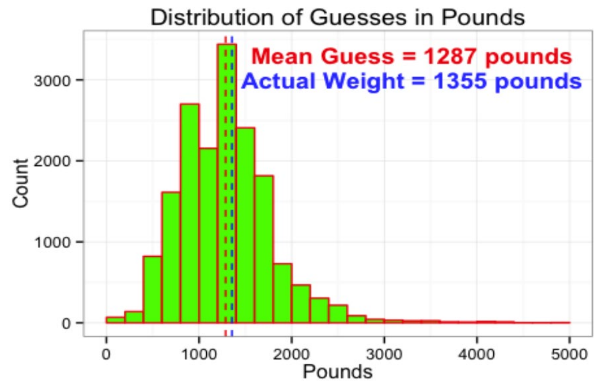


Photograph Credit: David Kestenbaum/NPR

Let's take a new situation. NPR (Planet Money) posted this picture of Penelope the cow, and asked people to guess how much the cow weighed. (For scale, the guy weighs 165 lbs.)

14. How much do you think the cow weighs? Write down your guess.
15. How far off do you think your guess is? Do you think it's too high? Too low? Not sure?
16. How accurate could we get by guessing? What would you consider accurate?
17. Here are the 17000 guesses collected on the internet. Why do we see all this variation? (This next figures were revealed to the class on the projector.)





18. Here's the crazy part. It turns out the average of all these guesses, is pretty close to the actual weight of the cow! Why might this be surprising (whether you personally are surprised or not)? Why is the average so close to the actual weight of the cow?
19. Why are the measurements "bad" in this situation?
20. Why is the average "good" in this situation?
21. What is the advantage of the mean (over the median) based on our discussion today?

Appendix D. Example Final Exam Questions

How are the concepts of mean and standard deviation similar? How are they different?

What is the conceptual difference between s and σ? (The formulas are not necessary for this answer.)

Why is the sampling distribution important to inferential statistics?

What is the Central Limit Theorem (CLT) and what is its purpose?

What are the similarities and differences between hypothesis testing and confidence intervals?

Appendix E. Mindset Questions

The statements that are italicized were reverse coded because high responses indicated a fixed (or entitative) mindset regarding intelligence/math ability. Strongly disagree was coded as a 1 and Strongly agree was coded as a 6. Students' responses for intelligence and math ability were separately averaged. High scores indicate more growth (or incremental) mindset.

Intelligence Statements
You can always substantially change how intelligent you are.
You can learn new things, but you can't really change your basic intelligence.
You have a certain amount of intelligence, and you can't really do much to change it.

Math Ability Statements
To be honest, you can't really change how much math ability you have.
You can change even your basic math ability level considerably.
No matter how much math ability you have, you can always change it quite a bit.

References

Austin, J. L., Lee, M. G., Thibeault, M. D., Carr, J., E., & Bailey, J. S. (2002). Effects of guided notes on university students' responding and recall of information. *Journal of Behavioral Education, 11*(4), 243-254. https://doi.org/10.1023/A:1021110922552

Baepler, P., Walker, J. D., & Driessen, M. (2014). It's not about seat time: Blending, flipping, and efficiency in active learning classrooms. *Computers and Education, 78*, 227-236. https://doi.org/10.1016/j.compedu.2014.06.006

Barsalou, L. W. (1999). Perceptions of perceptual symbols. *Behavioral and Brain Science, 22*(4), 637-660. https://doi.org/10.1017/S0140525X99532147

Barsalou, L. (2003). Situated simulation in the human conceptual system. *Language and Cognitive Processes, 18*(5-6), 513-562. https://doi.org/10.1080/01690960344000026

Beilock, S. L., & Carr, T. H. (2005). When high-powered people fail: Working memory and "choking under pressure" in math. *Psychological science, 16*(2), 101-105. https://doi.org/10.1111/j.0956-7976.2005.00789.x

Bergmann, J. & Sams, A. (2012). *Flip your classroom: How to reach every student in every class every day.* Washington, DC: International Society for Technology in Education.

Blackwell, L. S., Trzesniewksi, K. H. & Dweck, C. S. (2007). Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and intervention. *Child Development*, *78*(1), 246-263. https://doi.org/10.1111/j.1467-8624.2007.00995.x

Burke, A. S. & Fedorek, B. (2017). Does "flipping" promote engagement?: A comparison of a traditional, online, and flipped class. *Active Learning in Higher Education, 18*(1), 11-24. https://doi.org/10.1177/1469787417693487

Chase, W. G., & Ericsson, K. A., (1981). Skilled memory. In J. R. Anderson (Ed.), *Cognitive Skills and Their Acquisition*. Hillsdale, NJ: Erlbaum.

Chase, W. G., & Simon, H. A. (1973). Perception in Chess. *Cognitive Psychology*, *4*, 55-81. https://doi.org/10.1016/0010-0285(73)90004-2

Chi, M. T., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, *49*(4), 219-243. https://doi.org/10.1080/00461520.2014.965823

Cohen, P. A., Ebeling, B. J., & Kulik, J. A. (1981). A meta-analysis of outcome studies of visual-based instruction. *Educational Technology Research and Development, 29*(1), 26-36.

Damasio, A. R. (1989). 3. Concepts in the brain. *Mind & Language, 4*(1), 24-28. https://doi.org/10.1111/j.1468-0017.1989.tb00236.x

Deslauriers, L., Schelew, E., Wieman, C. (2011). Improved learning in a large-enrollment physics class. *Science, 332*(6031). https://doi.org/10.1126/science.1201783

Dewey, J. (1929). *The sources of a science of education*. New York: Liveright.

Dweck, C. S. (2007). *Mindset: The new psychology of success.* New York, NY: Ballantine Books.

Ericsson, K. A., Krampe, R. T., Tesch-Romer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100*(3), 363-406. https://doi.org/10.1037/0033-295x.100.3.363

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, *111*(23), 8410-8415. https://doi.org/10.1073/pnas.1319030111

Galway, L. P., Corbett, K. K., Takaro, T. K., Tairyan, K., & Frank, E. (2014). A novel integration of online and flipped classroom instructional models in public health higher education. *BMC*

Medical Education 14, 181. https://doi.org/10.1186/1472-6920-14-181

Garofalo, J. (1989). Beliefs and their influence on mathematical performance. *The Mathematics Teacher, 82*(7), 502-505.

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

Gick, M. L. & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology, 12*, 306-355.

Gick, M. L. & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology, 15,* 1-38. https://doi.org/10.1016/0010-0285(83)90002-6

Gilboy, M. B., Heinerichs, S., Pazzaglia, G. (2015). Enhancing student engagement using the flipped classroom. *Journal of Nutrition Education and Behavior, 47*(1), 109-114. https://doi.org/10.1016/j.jneb.2014.08.008

Glenberg, A. M. (1997). Mental models, space, and embodied cognition. In T. B. Ward, S. M. Smith, & J. Vaid (Eds.), *Creative Thought: An Investigation of Conceptual Structures and Processes* (pp. 495-522). Washington, DC, US: American Psychological Association.

Griffith, J. D., Adams, L. T., Gu, L. L., Hart, C. L., & Nichols-Whitehead, P. (2012). Students' attitudes toward statistics across the disciplines: A mixed-methods approach. *Statistics Education Research Journal, 11*(2), 45-56.

He, Y., Swenson, S., & Lents, N. H. (2012). Online video tutorials increase learning of difficult concepts in an undergraduate analytical chemistry course. *Journal of Chemical Education, 89*, 1128-1132. https://doi.org/10.1021/ed200685p

Hill, J. L., & Nelson, A. (2011). New technology, new pedagogy? Employing video podcasts in learning and teaching about exotic ecosystems. *Environmental Education Research, 17*, 393-408. https://doi.org/10.1080/13504622.2010.545873

Jensen, J. L., Kummer, T. A., Godoy, P. D. d. M. (2015). Improvements from a flipped classroom may simply be the fruits of active learning. *CBE-Life Sciences Education, 14*, 1-12. https://doi.org/10.1187/cbe.14-08-0129

Kapur, M. (2008). Productive failure. *Cognition and Instruction, 26*(3), 379-424. https://doi.org/10.1080/07370000802212669

Kapur, M. (2012). Productive failure in learning the concept of variance. *Instructional Science, 40*, 651-672. https://doi.org/10.1007/s11251-012-9209-6

Kapur, M. (2014). Productive failure in learning math. *Cognitive Science: A Multidisciplinary Journal, 38*, 1008-1022. https://doi.org/10.1111/cogs.12107

Koedinger, K. R. (2002). Toward evidence for instructional design principle: Examples from Cognitive Tutor Math 6. In D. S. Mewborn, P. Sztajn, D. Y. White, H. G., Wiegel, R. L. Bryant, & K. Nooney (Eds.), *Proceedings of Twenty-Fourth Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education, Vol. 1* (21-49). Columbus, OH: ERIC Clearinghouse for Science, Mathematics, and Environmental Education.

Koedinger, K. R., Corbett, A. T., & Perfetti, C. (2012). The Knowledge-Learning- Instruction framework: Bridging the science-practice chasm to enhance robust student learning. *Cognitive science*, *36*(5), 757-798. https://doi.org/10.1111/j.1551-6709.2012.01245.x

Lage, M. J., Platt, G. J., & Treglia, M. (2000). Inverting the classroom: A gateway to creating an inclusive learning environment. *The Journal of Economic Education*, *31*(1), 30-43. https://doi.org/10.2307/1183338

Ma, X. (1999). A meta-analysis of the relationship between anxiety toward mathematics and achievement in mathematics. *Journal for Research in Mathematics Education, 30*(5), 520-540. https://doi.org/10.2307/749772

Mani, A., Mullainathan, S., Shafir, E., & Zhao, J. (2013). Poverty impedes cognitive function. *Science, 341*(6149), 976-980. https://doi.org/10.1126/science.1238041

Mayer, R. E. (2001). *Multimedia learning.* Cambridge, UK: Cambridge University Press.

Means, B., Toyama, Y., Murphy, R., Bakia, M., & Jones, K. (2009). Evaluation of evidence-based practices in online learning: A meta-analysis and review of online learning studies. *US Department of Education.*

Merkt, M., Weigand, S., Heier, A., Schwan, S. (2011). Learning with videos vs. learning with print: The role of interactive features. *Learning and Instruction, 21*(6), 687-704. https://doi.org/10.1016/j.learninstruc.2011.03.004

Merrill, M. D. & Twitchell, D. G. (1994). *Instructional Design Theory.* Englewood Cliffs, NJ: Educational Publications, Inc.

Neef, N. A., McCord, B. E., & Ferreri, S. J. (2006). Effects of guided notes versus completed notes during lectures on college students' quiz performance. *Journal of Applied Behavior Analysis, 39*(1), 123-130. https://doi.org/10.1901/jaba.2006.94-04

Nouri, J. (2016). The flipped classroom: For active, effective and increase learning - especially for low achievers. *International Journal of Educational Technology in Higher Education, 13*(33), 1-10. https://doi.org/10.1186/s41239-016-0032-z

Prain, V., Tytler, R., & Peterson, S. (2009). Multiple representations in learning about evaporation. *International Journal of Science Education, 31*(6), 787-808. https://doi.org/10.1080/09500690701824249

Schoenfeld, A. H. (1989). Explorations of students' mathematical belief and behavior. *Journal for Research in Mathematics Education,* 338-355. https://doi.org/10.2307/749440

Schwan, S., & Riempp, R. (2004). The cognitive benefits of interactive videos: Learning to tie nautical knots. *Learning and Instruction, 14*, 293-305. http://dx.doi.org/10.1016/j.learninstruc.2004.06.005

Stigler, J. W., Givvin, K. B., & Thompson, B. J. (2010). What community college developmental mathematics students understand about mathematics. *MathAMATYC Educator*, *1*(3), 4-16.

Stigler, J. W. & Hiebert, J. (1999). *The teaching gap: Best ideas from the world's teachers for improving education in the classroom*. New York, NY: Free Press.

Son, J. Y., Ramos, P., DeWolf, M., Loftus, W., & Stigler, J. W. (2018). Exploring the practicing-connections hypothesis: Using gesture to support coordination of ideas in understanding a complex statistical concept. *Cognitive Research: Principles and Implications, 3*(1), 1-13. https://doi.org/10.1186/s41235-017-0085-0

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. C*ognitive Science*, *12*(2), 257-285. https://doi.org/10.1207/s15516709cog1202_4

Treagust, D., Chittleborough, G., Mamiala, T. (2003). The role of submicroscopic and symbolic representations in chemical explanations. *International Journal of Science Education, 25*(11), 1353-1368. https://doi.org/10.1080/0950069032000070306

VanLehn, K. (1989). Rule acquisition events in the discovery of problem solving strategies. *Cognitive Science, 15*(1), 1-48. https://doi.org/10.21236/ada225579

Watkins, A. E., Scheaffer, R. L., & Cobb, G. W. (2011). *Statistics: From data to decision* (2nd ed.) John Wiley & Sons, Inc.

Wilson, S. G. (2013). The flipped class: A method of address the challenges of an undergraduate statistics course. *Society for the Teaching of Psychology, 40*(3), 193-199. https://doi.org/10.1177/0098628313487461

Yeager, D. S., & Dweck, C. S. (2012). Mindsets that promote resilience: When students believe that personal characteristics can be developed. *Educational Psychologist, 47*(4), 302-314. https://doi.org/10.1080/00461520.2012.722805

Yong, D. Levy, R., Lape, N. (2015). Why no difference? A controlled flipped classroom study for an introductory differential equations course. *Problems, Resources, and Issues in Mathematics Undergraduate Studies, 25*, 907-921. https://doi.org/10.1080/10511970.2015.1031307