# TEACHING MODELING IN INTRODUCTORY STATISTICS USING THE BETTER BOOK APPROACH

Laura Fries, Caylor R. Davis, Icy (Yunyi) Zhang, and Ji Y. Son
University of California, Los Angeles
California State University, Los Angeles
json2@calstatela.edu

*Statistical modeling is typically seen as an advanced skill and rarely introduced to students at the introductory level. But in data science and its applications (e.g., public health, politics), modeling is a necessary component of data literacy. To address this need, we should teach introductory students modeling from the beginning, connecting the content to the modern world of data science. This may have the added benefit of bringing coherence to statistics. We teach this content using research-based pedagogy – the practicing connections framework – and utilize new technology (CourseKata.org) capable of conducting experiments to continuously improve instruction. We are thus able to test theories of how students learn difficult, time-consuming concepts, such as the concept of a statistical model. Together, the curriculum, pedagogical theory, and technology provide a process to make the book incrementally better at producing a modern, coherent, and flexible understanding of statistics.*

## INTRODUCTION

Introductory statistics teachers commonly report that many students leave the class with shallow knowledge that does not translate well to future classes or real world situations (Fries, Son, Givvin, & Stigler, 2020). They have learned the formulas and, at best, can mechanically apply them in a testing situation, but are often unable to build upon this knowledge in more advanced classes or use it to reason about situations that deviate slightly from the original context (e.g., Bassok, Wu, & Olseth, 1995; Son, Ramos, DeWolf, Loftus, & Stigler, 2018).

We hypothesize that students are ill-prepared to generalize their knowledge because they have not developed *coherent* statistics knowledge. The organization of most introductory statistics content has focused on differentiating statistical approaches rather than showing their inherent relatedness to one another. This leads to fragmented knowledge made up of discrete concepts and "rituals" (Gigerenzer, 2018) that are unmoored from a coherent organization of a domain. Typically, after introductory students have focused on figuring out whether a situation calls for a t-test or simple linear regression, a more advanced course tells them that both the t-test and simple regression are examples of the General Linear Model (GLM). Perhaps if we teach students this inherent similarity from the beginning, they can build up a coherent understanding of the domain. We are not alone in this thinking, as several other research and statistics education groups have also made the effort to put a heavier focus on teaching modeling in the introductory statistics curricula (e.g., Hutcheson & Sofroniou, 1999; Kaplan, 2009; Pfannkuch, Ben-Zvi, & Budgett, 2018).

However, substantial research has shown that simply telling someone something does not in itself guarantee that learning will take place (e.g., Koriat & Bjork, 2005); repeatedly telling students that the common thread through introductory statistics is modeling is not sufficient. How do we make the structure of the domain explicit, not so much in the minds of instructors and statisticians, but in the minds of our introductory students so that they can begin to see it for themselves? Based on the latest theories and findings in studies of teaching and learning, we propose that students must practice making connections to the core concept of the statistical model (Fries et al., 2020; Son, Blake, Fries, & Stigler, 2021). This instructional design framework is called the *practicing connections hypothesis*.

Despite the careful reliance on the latest in learning science, we still call it a hypothesis because most of the research done on teaching and learning occurs in one-hour lab settings focused on learning independent "bits" of knowledge or on helping students make a single connection between concepts. The idea of a statistical model is different -- it cannot be taught in an hour and the connections that students can make about a model in an hour are unusable. The benefit of learning a concept like the statistical model may not show up for many days, weeks, or even months. So how would we conduct research to test hypotheses about learning over longer time scales in realistic settings? Even if the practicing connections hypothesis sounds promising, how would we engage in research on such a theory?

In: R Helenius, E Falck (Eds.), *Statistics Education in the Era of Data Science*
*Proceedings of the Satellite conference of the International Association for Statistical Education (IASE),*
Aug-Sept 2021, Online conference.  

To answer these questions, we believe that focusing on modeling in the introductory statistics curriculum is not enough. Even when armed with a learning theory to help students appreciate the connections to modeling, it is still not enough. We need a novel approach to educational research, one where we can test and improve learning in complex domains and more ecologically valid settings. We need a way for teachers to partner with researchers and developers to build and implement solutions and test them rigorously. Currently, many of our most committed teachers are worn thin trying to fulfill all three roles -- teacher, researcher, and developer. We also need the humility to assume that materials designed with the best intentions and latest pedagogical ideas must be field tested and incrementally improved over time. In short, we need a better way to use data science to improve the teaching and learning of data science.

To address this need, we have developed the *better book* approach to research and development, housed in a novel technology platform (CourseKata.org), which has several important design features that allow us to examine complex learning in realistic settings: a focus on improving a free set of learning materials with embedded assessments that can be thought of as an interactive textbook; a way to conduct random assignment experiments, such that individual students can be randomly assigned to receive different versions of textbook content (e.g., different pages or videos); and a hub where researchers and developers can work together with instructors to incrementally improve the textbook. These three features allow innovations to be housed, not in a scholarly manuscript, but in the instructional materials themselves.

We apply this better book approach to the teaching of statistics with a focus on making connections to core concepts, such as modeling, that underlie the domain. Our group has separate manuscripts each focused on one aspect of this endeavor: the statistics curriculum centered on modeling (Son et al., 2021), the pedagogical theory focused on making connections (Fries et al., 2020), and the research and development approach to continuously make a better book (Stigler et al., 2020). In this brief paper, we will demonstrate how those three facets fit together. First, we describe how the practicing connections hypothesis shapes the statistics content in the instructional materials. Then, we outline how the better book approach allows for the continuous improvement of both the pedagogy and materials. Finally, we discuss the implications of our work, both for statistics educators looking for high quality instructional materials and for researchers who want to understand the complex system of real-world teaching and learning.

## HOW PRACTICING CONNECTIONS IMPACTS INTRODUCTORY STATISTICS CURRICULUM

As our goal is to develop coherent and transferable knowledge of statistics, we set out to help students make, for themselves, the connections that lend coherence to the introductory statistics curriculum. We call this *practicing connections* to capture the idea that students must productively struggle to make connections (Kapur, 2008) and this struggle with connections should be deliberately practiced over time (Ericsson, Krampe, & Tesch-Romer, 1993). Furthermore, the practicing connections framework suggests that students must work to make connections between *core concepts*, *key representations*, and *the world*, including its *contexts* and *practices*, in order to coordinate their knowledge together. It is this coordination that leads to flexible transfer (Schwartz & Goldstone, 2015).

To illustrate how the practicing connections framework impacts the introductory statistics curriculum, we will focus on one core concept -- that of a statistical model -- and how it unfolds through the curriculum. In order to enhance the coherence, we have also built the entire interactive textbook roughly around a narrative structure.

The story of statistics actually starts with the problem of variation. When two different objects or events are considered examples of the same thing (e.g., students) but they vary on some dimension (e.g., test score), how do we make sense of that variation? We spend the first part of the book exploring variation with our students by making and interpreting visualizations. We develop students' awareness of variation and teach them to look at the sample distribution for clues about the DGP (Data Generating Process). We get them to notice that samples are a reflection, but a puny and sometimes misleading reflection, of the true DGP.

Students start to wonder if there is a way to use the patterns of variation they see to predict some future data. Thus, we introduce the model. We first start with the simplest model -- a mean. Although students can use this "empty model" to predict the outcome for a case (e.g., a student's test score), the model also reveals error (the residual). Thus, individual data points can be parsed into MODEL and

ERROR. This is represented mathematically as DATA = MODEL + ERROR. Error was always there but now we can see it; it's everywhere. But there is hope because we can build better models to reduce it.

We start to include explanatory variables to build more complex models, which are, in essence, linear functions that can generate a specific prediction for different cases. They first learn that we can build models from categorical explanatory variables (e.g., a student's college major) and make different predictions for one group versus another. This then expands to complex models that can also include quantitative explanatory variables (e.g., a student's prior exam score) and make different predictions for cases based on the particular value of the explanatory variable. The concept of a model begins to gather momentum. The now-familiar concepts of DATA, MODEL, and ERROR are leveraged as they apply to more complex models, unifying what would have been considered different "statistics" or parameter estimates (mean, mean difference, slope) into a world where our goal is to minimize or balance error (e.g., residuals, sum of squares, standard deviation) by creating better models of the DGP.

Having developed an understanding of models, students are then able to evaluate and compare them against each other. Students learn to quantify how much variation has been explained by a more complex model over the empty model (e.g., PRE, Proportion Reduction in Error, and F ratio) but are constantly prompted to consider, "Wouldn't even these statistics be different if we had collected a slightly different sample?" or "Would a random DGP that had nothing to do with the explanatory variable be able to generate such a statistic?" Students realize that the same DGP could produce a variety of samples and thus could produce a variety of model estimates. This idea is leveraged into the concept of a sampling distribution. Students use simulation techniques such as randomization and bootstrapping to emulate a random DGP and create sampling distributions of PRE and F ratio (Hesterberg, 2015; Makar & Rubin, 2018). Then they can compare whether the PRE or F as big as the one from our sample could have been easily generated by such a DGP.

In this "story," the idea of the statistical model is gradually built up in relation to other core concepts such as distributions (of which there are three: samples, DGP, sampling distributions) and randomness (used as a possible data generating process). The pay off of these core concepts emerges steadily over time. In fact, there are times when the concept of a model seems more onerous than traditional treatments of concepts. For example, it is much easier to think of the mean as a measure of central tendency than as an empty model. However, core concepts like modeling can help connect together all of the content of introductory statistics, making the effort worth it in the long run.

The emphasis on exploring variation, modeling variation, and evaluating models provides the higher-order explicit connections we want our students to productively struggle with. The curriculum is embodied in a free, interactive, online textbook (available for preview at CourseKata.org). This textbook interleaves text and video instruction with visualizations, R coding exercises, and formative assessment questions. As students work their way through the textbook, all data are captured and available for instructors. The interactivity is not simply for the sake of engagement -- the goal of these interactive elements is to help students practice making connections. However, despite some promising signs of flexible and transferable learning (see Son et al., 2021), we know that not all students learn effectively from these materials. How do we further improve these materials and the way that students engage with them?

## THE BETTER BOOK APPROACH: TOWARDS BETTER CURRICULA AND PEDAGOGY

Stigler and colleagues (2020) describe education as a complex system, and one that can be difficult to study and difficult to change. Lewis (2015) has shown that applying improvement science to the system of education can lead to fruitful results. We adopt this thinking as we aim for continual improvement of the course content and pedagogy.

Any curriculum, even one designed with the best of intentions, needs to improve based on student learning data. So the textbook relies on new technology (CourseKata.org) capable of housing and delivering different versions of instructional materials to common learning management systems, interleaving assessments in the textbook (e.g., questions and R coding exercises alongside text, figures, and video), and making student engagement and performance data available for continuous improvements to the curriculum.

This common platform allows for us to capture large amounts of data as students enrolled in statistics courses engage with the interactive textbook. Researchers have anonymized access to what videos are watched and for how long, what pages were visited and how many times, when assignments were submitted and what students say in their submissions. We can see how many attempts students made on R coding exercises, and compile these data at the level of the student, the class, and the school. We are also able to track students' affect as they move through the course, as students are asked for feedback about the content and their related affective responses. All these data inform future studies and future modifications to the curriculum.

We have integrated the ability to conduct random assignment experiments within the interactive textbook, allowing for the study of teaching and learning as it happens *in vivo*, across hundreds or thousands of students. We are currently testing whether a series of videos depicting a particular diagram will work to improve students' understanding of different distributions and how they are interrelated; or if an analogical intervention will help students make connections between representations such as graphs, R code, and GLM notation; or if personalized versus generic text reminders will help students overcome homework procrastination. Each of these manipulations incrementally improves the materials and gives us insight into how they are actually used by students and teachers. For example, even though students randomly assigned to pages with videos respond more accurately on subsequent questions than a no-video group, many students in need of this supplemental instruction skip right over embedded videos. We can now ask, why aren't more students watching these videos, and what can be done about it? In this spirit, we use the phrase "better book approach" not to imply that our book *is* better; instead, it is a process by which, over time, our book can *get better*.

## IMPLICATIONS FOR EDUCATORS AND RESEARCHERS

### *Implications for Statistics Education*

Projects such as ours can practically support statistics instructors. Instructors can try teaching modeling and computational approaches with a free interactive textbook, connect their students to more authentic practices in the broader field of data science, and find materials that directly support active pedagogy in synchronous online or face-to-face classes.

Efforts such as ours to provide free or low-cost materials that are continuously improved based on learning data directly support teachers and students with a number of user-facing features. Our free textbook collects students' responses to over 1200 formative assessment items and provides immediate feedback to students. These responses also provide a wealth of information to instructors. Simply reading through their students' responses can give instructors a better sense of what to focus on during face-to-face or synchronous online class time. Although there are many technically skilled instructors, our textbook allows students to submit tickets when they find errors. These tickets are triaged so that technical fixes are made by the development team while conceptual questions are forwarded to the instructor and textbook improvement questions are forwarded to the authoring team. This way, instructors can focus on what they do best.

As statistics education continues to evolve, many have begun to embrace a more interdisciplinary approach to data science. Students need to learn concepts and skills such as modeling, which still falls squarely within the domain of statistics, but they also need exposure to programming and modern tools. Open-source runnable documents such as Jupyter notebooks provide students with opportunities to communicate their analyses using a mixture of figures, equations, computations, and writing. Jupyter notebooks can also help students engage in emerging efforts to promote reproducibility and transparency in the field. Our *better book* project has recently begun integrating Jupyter notebooks (see Blake, Winjum, & Stigler, 2020) and associated workflows (e.g., pair programming) into our modeling curriculum to more authentically teach students the skills needed in the modern world.

The practicing connections framework fits well within a broader movement in education toward student-centered, active pedagogy. Instructors who have already begun to include inquiry- or project-based learning, flipped classroom, and other types of active learning into their face-to-face or synchronous instruction will find that our materials are designed with those pedagogies in mind. The practicing connections framework has taken the most promising findings from the learning sciences and incorporated them into actual instructional materials to support the pedagogies that teachers desire. New innovations in curriculum, such as projects developed in Jupyter notebooks, are co-developed, shared, and improved upon by a community of instructors, researchers, and developers. As instructors begin to use our materials, they are also imagining how the practicing connections framework can be helpful but may need to be modified for other courses (e.g., advanced statistics, calculus, physics). What we present here is an initial approach, and we welcome contributions to continue to make the theory better and more useful.

*Implications for Education Research*

Perhaps the most unique implication of this work is the ability to study learning, using A/B testing, in real students taking courses for credit and then storing those innovations in the learning resources themselves. Experimental work taking place in a lab, for the duration of an hour or two and covering a single stand-alone concept, is notoriously challenging to apply to the complex system of real world learning. Although we have lab findings suggesting that students learn more from this video or that video, when we insert experiments into an interactive textbook, we see that the benefit is restricted to a small group of students who actually watch the videos. We can carry on the important interdisciplinary work of figuring out what design features lead to more students watching the videos (e.g., when an open-response question follows? when it is alone on a page?). These insights may also extend to other contexts, videos, and domains -- but even if they don't, they will *at the very least* impact thousands of textbook users.

The use of our online book in a variety of settings (from high schools and community colleges up through graduate programs; in a variety of geographic locations; in different departments) allows researchers to examine the heterogeneity in treatment effects. Virtually all teaching interventions occur under some conditions for some students but we need better methods of studying that heterogeneity (Tipton, Bryan, & Yeager, in preparation). The CourseKata technology provides the data infrastructure so that individual researchers do not have to put together heterogeneous samples for individual projects. This way, we can go beyond a focus on whether an intervention "works" ($p < .05$) to understanding why variation in learning exists and being able to reduce that variation. This is the way to ensure that students from all backgrounds end up with a flexible understanding of statistics.

## CONCLUSION

In this paper, we have presented a sweeping but brief overview of our multifaceted work to continuously improve an introductory statistics curriculum focused on modeling. Our approach to education as a complex system is to treat it as such, where simultaneously curriculum, theory, and research methodology must be tested and refined in implementation. Each component of our project -- the curriculum, theory, and R&D technology -- allows us to include more of the complex system of education in our emergent solution. The practicing connections hypothesis helps make our modeling centered introductory statistics more coherent in the minds of our students. The better book approach helps us continuously improve our curriculum and theory. Although a free interactive introductory statistics textbook based on modeling is a good start, working it out with the practicing connections theory and student data from the CourseKata technology is part of a broader, hopefully more generalizable effort towards solving important educational problems.

## REFERENCES

Bassok, M., Wu, L.-L., & Olseth, K. L. (1995). Judging a book by its cover: Interpretative effects of content on problem-solving transfer. *Memory & Cognition*, *23*(3), 354–367.

Blake, A. B., Winjum, B. J., & Stigler, J. W. (2020). nbteach: Author, distribute, and grade Jupyter notebook assignments, simply. Retrieved from https://github.com/UCLATALL/nbteach

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.

Fries, L., Son, J. Y., Givvin, K. B., & Stigler, J. W. (2020). Practicing connections: A framework to guide instructional design for developing understanding in complex domains. *Educational Psychology Review*, 1-24.

Gigerenzer, G. (2018). Statistical rituals: The replication delusion and how we got there. *Advances in Methods and Practices in Psychological Science*, *1*(2), 198-218.

Hesterberg, T. C. (2015). What teachers should know about the bootstrap: Resampling in the undergraduate statistics curriculum. *The American Statistician*, *69*(4), 371-386.

Hutcheson, G. D., & Sofroniou, N. (1999). *The multivariate social scientist: Introductory statistics using generalized linear models*. Sage.

Kaplan, D. T. (2009). *Statistical modeling: A fresh approach.* CreateSpace Independent Publishing

Platform.
Kapur, M. (2008). Productive failure. *Cognition and Instruction*, *26*(3), 379-424.
Koriat, A., & Bjork, R. A. (2005). Illusions of competence in monitoring one's knowledge during study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(2), 187-194.
Lewis, C. (2015). What is improvement science? Do we need it in education? *Educational Researcher, 44*(1), 54-61.
Makar, K., & Rubin, A. (2018). Learning about statistical inference. *International Handbook of Research in Statistics Education*, 261-294.
Pfannkuch, M., Ben-Zvi, D., & Budgett, S. (2018). Innovations in statistical modeling to connect data, chance and context. *ZDM*, *50*(7), 1113-1123.
Schwartz, D. L., & Goldstone, R. (2015). Learning as coordination. *Handbook of Educational Psychology*, 61-75.
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*, 1-18.
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.
Stigler, J. W., Son, J. Y., Givvin, K. B., Blake, A., Fries, L., Shaw, S. T., & Tucker, M. C. (2020). The better book approach for education research and development. *Teachers College Record*, *123*(2).
Tipton, E., Bryan, C., & Yeager, D. *To change the world, behavioral intervention research will need to get serious about heterogeneity*. Manuscript in Preparation.