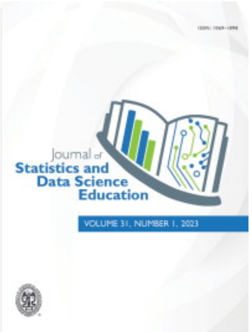

Journal of Statistics and Data Science Education

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ujse21

Profiles in the Teaching of Experimental Design and Analysis

Byran J. Smucker, Nathaniel T. Stevens, Jacqueline Asscher & Peter Goos

To cite this article: Byran J. Smucker, Nathaniel T. Stevens, Jacqueline Asscher & Peter Goos (2023): Profiles in the Teaching of Experimental Design and Analysis, Journal of Statistics and Data Science Education, DOI: 10.1080/26939169.2023.2205907

To link to this article: https://doi.org/10.1080/26939169.2023.2205907

JOURNAL OF STATISTICS AND DATA SCIENCE EDUCATION
2023, VOL. 00, NO. 0, 1–14
https://doi.org/10.1080/26939169.2023.2205907

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS

Check for updates

Profiles in the Teaching of Experimental Design and Analysis

Byran J. Smucker[a], Nathaniel T. Stevens[b], Jacqueline Asscher[c], and Peter Goos[d,e]

[a]Department of Statistics, Miami University, Oxford, OH; [b]Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, ON, Canada; [c]School of Engineering, Kinneret College on the Sea of Galilee, Tzemach, Israel; [d]Department of Biosystems, KU Leuven, Leuven, Belgium; [e]Department of Engineering Management, University of Antwerp, Antwerp, Belgium

ABSTRACT
The design and analysis of experiments (DOE) has historically been an important part of an education in statistics, and with the increasing complexity of modern production processes and the advent of large-scale online experiments, it continues to be highly relevant. In this article, we provide an extensive review of the literature on DOE pedagogy, and provide five perspectives on the subject: one from each of the authors as well as a composite profile derived from a survey of DOE instructors. Our work provides a snapshot of current DOE pedagogy that showcases both the similarities and variety in how the subject is taught, as well as a look ahead at how its instruction may evolve. Supplementary materials for this article are available online.

ARTICLE HISTORY
Received May 2022
Accepted April 2023

KEYWORDS
Active learning; Data science education; Experiment; Online experimentation; Optimal design; Statistics education

1. Introduction

At the heart of statistical design of experiments (DOE) is the systematic, efficient construction of a set of combinations of controlled levels of input variables (factors) and an assignment of these combinations to the experimental units, which together enable a straightforward, predetermined analysis of how the factors affect output variables (responses). DOE can be seen as the experimental approach to establishing causal inferences, by identifying and quantifying effects through controlled experiments.[1] Designed experiments have been widely used in the physical sciences (e.g., Deming and Morgan 1993; Gunter and Matey 1993; Hanrahan and Lu 2006; Kreutz and Timmer 2009; Leardi 2009), agriculture (e.g., Fisher 1971, Mead, Curnow, and Hasted 2017), engineering (Lazic 2006; Ilzarbe et al. 2008; Antony 2014), industry (Davies 1954; Goh 2001; Tanco et al. 2009), and in social sciences such as psychology and education (e.g., Lindquist 1953; Campbell and Stanley 2015). More recently, there has been a renewed and broadened interest in the field since it has found wide-spread application by technology companies who can experimentally generate enormous quantities of online data (Luca and Bazerman 2021; Thomke 2020; Kohavi, Tang, and Xu 2020). Anderson-Cook and Lu (2023) discuss several other modern applications of DOE and emphasize its sustained importance in the era of big data. As such, DOE is typically a core topic taught in both undergraduate and graduate statistics programs (American Statistical Association 2014; Chance and Peck 2014; Blades, Schaalje, and Christensen 2015; Horton 2015; Woodard 2023). In this article, we examine how DOE is currently taught in academia, as well as where its pedagogy may be moving in the future. More specifically, we investigate who is teaching and being taught DOE, what topics are being included in the courses, how software is being used, and what teaching methods are being employed, as well as the potential evolution of DOE instruction in the future. We accomplish this by describing five profiles of DOE instruction: the first taken from a survey of 50 DOE instructors, while the last four are contributed individually by the authors. Together, this provides readers with a sense of the core ideas common to many DOE courses as well as the variety of methods, topics, and paradigms that are used.

To limit the scope of our work, we focus largely on mainstream experimental design instruction within statistics departments in the United States and Canada, though we do present an international perspective as well, given the affiliations of the authors of this article. For the same reason, and because of our own training and background in industrial and engineering design applications, we do not consider experimental design courses emphasizing experiments within the social sciences or clinical trials. The body of material that we focus on is that which has evolved from R.A. Fisher's early work in agricultural experimentation (e.g., Fisher 1971), and the contributions of George Box and associates (e.g., Box and Draper 1987; Box, Hunter, and Hunter 2005). We refer later to Fisherian versus Boxian DOE paradigms. For the former, we assume an emphasis on replicated experiments with few factors, blocking structures, and ANOVA. For the latter, the focus is on unreplicated, multifactor experiments, often either two-level screening designs or designs in the service of response optimization, with regression as the basic analysis procedure.

CONTACT Byran J. Smucker smuckebj@miamioh.edu Department of Statistics, Miami University, Oxford, OH.

[1]This is to be contrasted with *observational* causal inference methods that also seek to identify and quantify causal effects, but without a controlled experiment. While such methods (because they lack an experiment) are not the focus of this paper, we address them briefly in Section 5.

Supplementary materials for this article are available online. Please go to *www.tandfonline.com/ujse*.

Table 1. Categorized list of literature on DOE pedagogy.

Category	References
General DOE Pedagogy	Hunter (1977); Easterling (2004); Goos and Leemans (2004)
DOE Projects and Case Studies	Hunter (1977); Kenett and Steinberg (1987); Box (1992); Mackisack (1994); Anderson-Cook (1998); Nolan and Speed (1999); Anderson-Cook and Dorai-Raj (2001); Howley (2003); Binnie (2004); Steiner et al. (2007); Lawson et al. (2011); Dunn (2013); Kuiper (2016); Kuhnt and Coleman (2020, 2021); Pyott (2021); Woodard (2023)
DOE Simulators	Bulmer (2003, 2012); Schrevens et al. (2004); Darius, Portier, and Schrevens (2007); Muske and Myers (2007); Steiner and MacKay (2009); Bulmer and Haladyn (2011); De Ketelaere et al. (2014); Kuiper (2016); Reis and Kenett (2017); Gramacy (2020)
Teaching DOE to non-Statistics Students	Pollock, Ross-Parker, and Mead (1979); Deming and Morgan (1983); Fillebrown (1994); Antony and Capon (1998); Zolman (1999); Lin Ho, Lyn Nge, and Hong Chua (2004); Lye (2005); Hiebert (2007); Hane (2007); Stafford, Goodenough, and Davies (2010)

Note that our work is descriptive. We do not attempt to formally study or empirically compare any of the teaching methods or approaches described in this article. Rather, we provide an account of the current "state of the DOE course" in North America and beyond, based on both the survey as well as our own courses. Future research could expand on this in many ways, including using designed experiments to understand how effective various approaches to DOE instruction are at increasing student understanding.

The remainder of the article is organized as follows. In the next section we provide an extensive review of the DOE pedagogy literature. To our knowledge, it has not been previously catalogued and synthesized to this extent. Section 3 contains the composite profile derived from our survey. Section 4 follows with the four individual DOE teaching profiles. Together, Sections 3 and 4 illustrate the commonalities but also the variation in DOE pedagogy with respect to topics and teaching methods. The article concludes with a summary of our findings and a discussion of the future of the DOE classroom.

2. Literature Review

To set the stage for our work, we review the literature on experimental design education. Table 1 provides a summary of the work we reviewed, while the Supplementary Material (SMA) gives a more extensive narrative regarding this literature.

As shown in Table 1, quite a bit has been written regarding specific projects and case studies that could be used in a DOE class, and the consensus is that student-performed real experiments are beneficial. Overall, the part of the DOE education literature focusing on projects aligns with the Guidelines for Assessment and Instruction in Statistics Education College Report (GAISE, Carver et al. 2016), which encourage educators to foster active learning, and to use and contextualize real data. Somewhat different from projects are case studies, which are examples and applications that include both a story and data. They are widely used to teach statistics in general and DOE in particular, though they may not be recognizable as "case studies" to those who use them more formally (Garvin 2003; Andrews 2021). Nolan and Speed (1999) are an exception, though their context is undergraduate mathematical statistics courses. A related category is the use of simulators to teach DOE, and again in Table 1 we see that a number of authors have reported on their use. (We note that the term "simulator" is commonly used in this context to denote an active learning tool that includes an underlying simulation engine as well as a story/theme.) An additional important class of papers in the literature is DOE instruction to non-statistics students. The relevance of this is witnessed by our survey results and by the fact that two of the authors of this article regularly teach DOE to students who are not majoring in statistics.

The literature thus contains many useful ideas about project- or class-based assignments that can be used effectively in the teaching of DOE, but, apart from Hunter (1977), Table 1 suggests that literature describing the teaching of experimental design more broadly is scarce. With this article, we provide a picture of the current status of DOE pedagogy, and answer questions such as "What are people teaching in their DOE classes?" and "How are they going about it?"

3. Teaching Experimental Design and Analysis: A Composite Profile

To gain an overview of DOE pedagogy as it is currently practiced, we conducted an anonymous survey[2] of DOE instructors at American and Canadian universities, which serves as a composite perspective on the teaching of DOE. The survey was sent to the chairs of the 219 departments identified by the American Statistical Association and the Statistical Society of Canada as ones that offer statistics programs. A list of these departments is included in the Supplementary Material (see SMB). Each chair was asked to share the survey with the instructor(s) that teach the DOE course(s) in their department. From this target population, 50 instructors responded to the survey, providing information about the DOE courses they teach and the manner in which they teach them. The full set of survey questions, and a summary of the responses to them, can also be found in the Supplementary Material (see SMC), though because of privacy concerns the full data are not available. It should be acknowledged that our sample is not random and there may exist a self-selection bias whereby those who chose to respond to the survey are those who care strongly about DOE pedagogy. We therefore emphasize that our findings do not necessarily generalize to the broader population of DOE instructors and courses, but merely represent a glimpse into DOE classrooms in Canada and the United States. What follows is a brief overview of the results of this survey.

3.1. Course and Instructor Demographics

We find that many of the DOE classes are reasonably small (75% have 50 or fewer students and only 10% have more than 100 students) and most of these courses (68%) are offered at most once

[2]This study was reviewed and received ethics clearance through the University of Waterloo Research Ethics Board (ORE#42536). Due to this protocol, we are unable to share the data.

per year. Although the audience is primarily statistics students, these courses also include students from business, engineering, computer science, data science, the life sciences, and social sciences, with most students taking the course because it is a degree requirement. We find that 42% of the classes are exclusively undergraduate, 24% are exclusively graduate, and 34% have a mixture of graduate and undergraduate students.

In terms of instructor demographics, almost all respondents (94%) have Ph.D.s, and among those who disclosed the field of their Ph.D., almost all (95%) received their Ph.D. in statistics or a closely related discipline. All respondents with Ph.D.s reported being engaged in research, and 51% of these research-active faculty publish about DOE in academic journals. We found no evidence from our survey that the topics or other elements of the courses differed according to instructor research status, though we include the standard caveat acknowledging that this was not a random sample. We also found that 68% of respondents engage in consulting and/or collaborative DOE work outside of the classroom, and 16% teach DOE in industry as well as in academia. This speaks to the applied nature of the topic.

3.2. Topics

We find that 84% of respondents use learning outcomes to guide their teaching. The survey presented four particular learning outcomes, and each respondent was asked to rank them from most important (first) to least important (fourth). The learning outcomes and their respective ranking frequencies are shown in Table 2. Interestingly, each outcome received each rank, indicating heterogeneity in the instructors' perceptions of importance. However, we do find some consensus in the rankings; the learning outcome most often ranked first is *Understand basic design and analysis principles*; the learning outcome most often ranked second is *Be able to use software to design and analyze experiments*; the learning outcome most often ranked third is *Be able to perform experiments by actually collecting data*; and the learning outcome most often ranked fourth is *Understand the theoretical underpinnings of DOE*. The course for which *Understand the theoretical underpinnings of DOE* is the most important learning outcome is a Ph.D.-level course.

Table 2. Learning outcome ranking frequencies ($n = 45$)[a].

	Importance			
	1st	2nd	3rd	4th
Understand basic design and analysis principles	39	3	2	1
Be able to use software to design and analyze experiments	1	29	13	2
Be able to perform experiments by actually collecting data	4	6	17	18
Understand the theoretical underpinnings of DOE	1	7	13	24

[a]Note that 45 of the 50 survey participants responded to this question.

In terms of content, most courses (82%) focus solely on DOE, though some (18%) teach DOE in combination with other topics such as linear models, sampling design, statistical process control, and introductory applied statistics. The motivating paradigm for most of these instructors (56%) is a combination of Box and Fisher, though 5% adhere exclusively to the Boxian paradigm and 36% exclusively to the Fisherian.[3] The specific DOE topics covered in these courses are summarized in Figure 1, and a tabulation of the number of classes that include each topic is provided in the Supplementary Material (see SMC). Some results are not surprising; for instance, we see that nearly all classes cover factorial experiments and blocking designs, but very few cover definitive screening designs or robust parameter designs. Topics such as fractional factorial designs and response surface methodology are more common in Boxian classrooms (classes 1–5 in Figure 1) than Fisherian ones (classes 6–23). However, some results are surprising; for instance, useful topics

[3]See our definition of Boxian vs. Fisherian paradigms given in the Introduction. We must acknowledge, however, that the survey respondents self-identified as Boxian, Fisherian, or a combination of both, without being given clear definitions of these paradigms, and so they may not have been consistently interpreted.

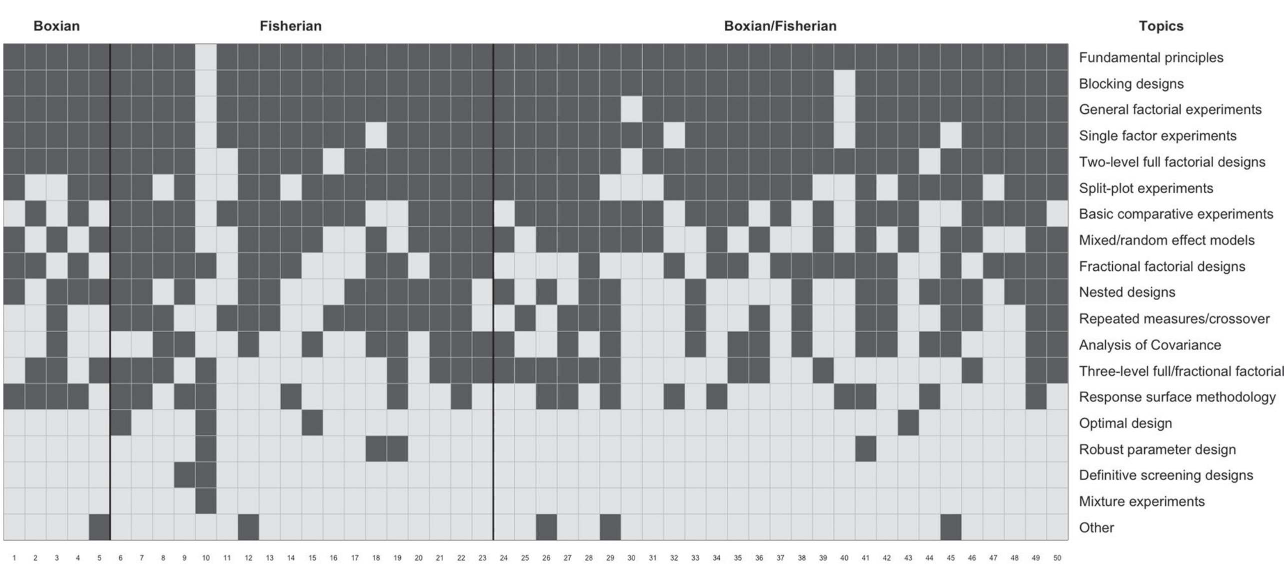

Figure 1. Topics by course. Visualization of which topics covered (dark gray) versus not covered (light gray) in each course. Respondents indicated that courses 1–5 are Boxian, 6–23 are Fisherian, and 24–50 are Boxian/Fisherian. Rows are sorted from most to least prevalent.

such as mixture experiments and optimal design appear much less frequently than we would have anticipated. "Other" topics include generalized linear mixed models and model selection, power analysis and sample size calculation, permutation tests, unbalanced designs and multiple imputation, REML, and Bayesian methods. Adjacent areas prevalent in the social and medical sciences, such as observational causal inference and clinical trials did not appear here. We emphasize again that our survey may be biased; though we aimed to distribute the survey broadly, it is possible that the respondents were disproportionately similar to the authors, as our focus is on industrial- and engineering-related experiments.

The median time-split between design and analysis is 40% of the course devoted to design, and 60% to analysis. The analysis method of choice is ANOVA, with 68% of courses featuring this more prominently than regression, 30% featuring it and regression equally, and only a single respondent featuring regression more prominently than ANOVA.

We find that 90% of courses incorporate a textbook (52% require, 38% recommend), while 10% do not. Among those that do, 29% incorporate multiple books. The five books most often used (in descending order of popularity) are (a) *Design and Analysis of Experiments* by Montgomery (2019); (b) *A First Course in Design and Analysis of Experiments* by Oehlert (2010); (c) *Design and Analysis of Experiments* by Dean, Voss, and Draguljić (2017); (d) *Design of Experiments: Statistical Principles of Research Design and Analysis* by Kuehl (2000); (e) *Experiments: Planning, Analysis, and Optimization* by Wu and Hamada (2011). The complete list of books used by respondents is provided in the Supplementary Material (see SMC).

3.3. Role of Software

Unsurprisingly, we find that 100% of respondents use software to some degree in their courses. Moreover, all respondents indicated that software is used by both the students for homework assignments and projects, as well as by themselves for instruction, with the median amount of instruction time devoted to software being 30%. The particular programs used (in descending order of frequency) are R (in 70% of classes), SAS (in 50% of classes), JMP (in 18% of classes), Minitab (three classes), and Matlab (a single class). Note that roughly a third of the respondents report using a combination of these software options in their courses, and that none of the respondents reported using any other software (such as, for instance, Design-Expert or Python). In the Supplementary Material (see SMC), we include a list of the R packages and SAS procedures the respondents reported commonly using. We remark that 44% of respondents also report using self-programmed functionality beyond available "off the shelf" software.

3.4. Teaching Methods

We find that respondents use a variety of teaching methods in DOE classes ranging from more traditional ones such as lectures and case studies to more modern ones such as flipped, blended or hybrid classrooms. However, traditional methods appear to be employed more often than modern ones. Figure 2 visualizes the distribution of teaching methods throughout the respondents' classrooms. Note that "Other" teaching methods included field trips to real-life ongoing experiments.

One relatively common component of the courses is the use of a simulator; 28% of respondents report featuring a simulator for instruction and/or assessment. The three simulators most often used (in descending order of popularity) are (a) homemade simulators; (b) The Garden Sprinkler (De Ketelaere et al. 2014); and (c) The Islands (Bulmer and Haladyn 2011).

Other assessment methods include exams, in-class and take-home assignments, labs, projects, and case studies. The distribution of assessment methods throughout the respondents' classrooms is visualized in Figure 3. Although we saw a mixture of traditional and modern methods of instruction, we see that the methods of assessment used are predominantly traditional: take-home assignments, a project, and one or more exams. Note that "Other" assessment methods included quizzes and reviewing published papers.

3.5. Future Evolution: Recent and Planned Changes to Courses

The survey posed an additional six open-ended questions probing the participants about the changes they have made to their courses in the last two years and the ones they plan to make in the following two years. These questions addressed three aspects of teaching methods separately: the use of software, topics, and course content. These questions were labeled as optional, and 20 of the 50 respondents did not reply to any of them. We observe wide variation in these free-form responses,

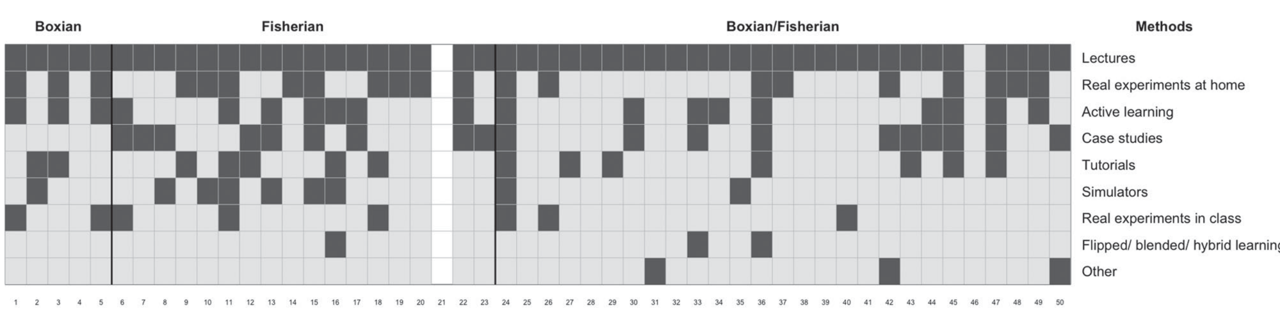

Figure 2. Teaching methods by course. Visualization of which teaching methods are used (dark gray) versus not used (light gray) in each course. Respondent 21's nonresponse is indicated by white squares. Respondents indicated that courses 1–5 are Boxian, 6–23 are Fisherian, and 24–50 are Boxian/Fisherian. Rows are sorted from most to least prevalent.

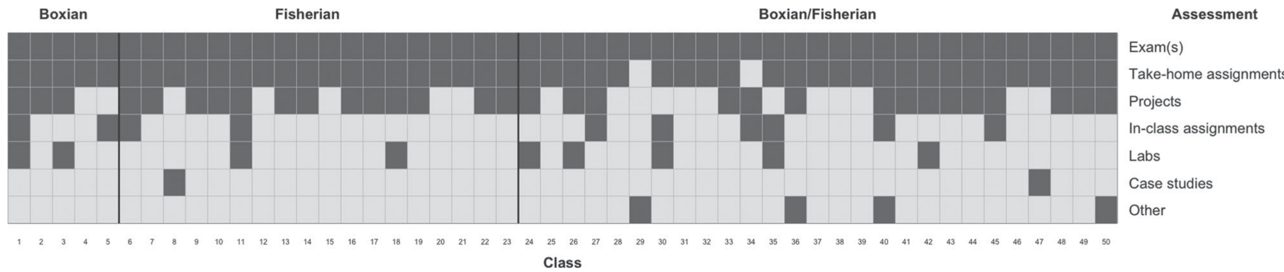

Figure 3. Assessment method by course. Visualization of which assessment methods are used (dark gray) versus not used (light gray) in each course. Respondents indicated that courses 1–5 are Boxian, 6–23 are Fisherian, and 24–50 are Boxian/Fisherian. Rows are sorted from most to least prevalent.

Table 3. Basic comparison of the four courses described by the authors.

Category	Smucker	Stevens	Asscher	Goos
Theme	"Designed so that Statistics Students Will Not be a Part of the Replication Crisis"	"Designed using Online Experimentation as a Source for Examples, Assignment Problems, and Project"	"Designed so that Engineering Students can use DOE in Industry and Research"	"Designed so Students are Equipped to Design Experiments in Both Standard and Unusual Settings"
Institution and Department	Miami University, Department of Statistics	University of Waterloo, Department of Statistics and Actuarial Science	Kinneret College on the Sea of Galilee, Quality and Reliability Engineering; also Technion - Israel Institute of Technology	KU Leuven, Faculty of Bioscience Engineering & Leuven Statistics Research Center
Audience	Advanced statistics undergraduates; first-year statistics master's students	Advanced undergraduates, masters and PhD; statistics, data/ computer/ actuarial science, engineering	Quality and reliability engineering undergraduates; undergraduate and graduate engineering students	Bioscience engineering undergraduates; M.Sc. students in statistics and data science
Assumed Prerequisites	Calculus, Probability; Linear Regression with matrix algebra	Calculus, Linear Algebra, Statistics and Probability, Linear Regression	Introductory Probability and Statistics, Linear Regression and ANOVA	Linear Regression with matrix algebra
Course frequency	1–2 sections per year	2 sections per year	1 section per year	1 section per year
Typical class size	30	150	20	120
Software	SAS; JMP	R	JMP	JMP
Textbook	Montgomery; Oehlert; Dean, Voss, and Draguljić (all optional)	Montgomery; Wu and Hamada (both optional)	Montgomery (optional)	Goos & Jones (compulsory)

but one overall conclusion is that there is a tendency to tweak rather than overhaul DOE courses. We note, however, that while the DOE courses themselves may be evolving slowly and in many different directions, the way that they are deployed by the institutions, and the decisions regarding which students take the DOE courses, are not captured by these responses.

After informally summarizing/categorizing the free-form responses, we note the following. Regarding teaching methods, though active learning and project-based pedagogical approaches have become more common, there doesn't appear to be a wholesale move toward such methods among the respondents in our survey. Five respondents reported a move to online teaching, three specifically noting the pandemic as the reason, and an additional two noted other changes made due to the pandemic. It is likely that additional participants made similar pandemic-related changes but did not report them, as the survey introductions specifically asked instructors to consider pre-pandemic versions of their courses. Regarding the use of software, we see in our survey that the DOE courses at universities are retaining and strengthening their focus on R and SAS. We also observe that newer DOE topics from academic DOE research (e.g., definitive screening designs, Jones and Nachtsheim 2011) are not yet finding their way into these courses. Also, while industry interest in online experimentation is expanding, we see minimal movement to include the issues specific to this context in DOE courses.

4. Teaching Experimental Design and Analysis: Individual Profiles

We now turn to more personal reflections regarding DOE pedagogy, in the form of course profiles from each of the authors. We subdivide each essay into descriptions of demographics, topics, teaching methods, the use of software, and future changes. In order to emphasize the unique points of view of each author, we break with convention and use the first person in these four individual pieces. Note that it is not our intention to present any of the personal profiles as exemplary or better than the rest. As such, we do not make specific recommendations about topics or teaching and assessment methods. Rather, we include the profiles to highlight certain commonalities but also to illustrate the rich variety in perspectives and approaches. We hope these profiles serve to seed ideas and inspire DOE educators.

We first provide a comparison of the logistical information from each of the classes (Table 3), and then each author gives additional narrative detail about their course.

4.1. The Traditional DOE Course, Designed so That Statistics Students Will Not Be a Part of the Replication Crisis (Smucker)

I (Smucker) inherited a course that was quite traditional, with a focus on designs with categorical factors, analyzed using ANOVA. Though my course is still rather conventional, it has developed in at least three important respects: (a) I've added discussion of designs for regression analyses; (b) I require the students to perform and analyze real experiments; and (c) I emphasize that without careful control and prespecification, p-values will be unreliable. I cover elements of both the Fisherian and Boxian paradigms in my course. Overall, my course development has been strongly influenced by its status as a required course for our statistics majors and M.S. students, my belief in the benefit of students performing experiments on their own, and my observation of contemporary discussions regarding best practices in exploratory and confirmatory experiments.

4.1.1. Topics

Classically, experimental design classes have emphasized categorical designs such as one-way and two-way full factorial designs and blocking designs such as randomized complete block and balanced incomplete block designs. My course includes these designs, along with categorical designs that include random effects, whether full factorial designs, split-plot designs, or designs with nested structures. Even though it may not be prioritized in some design classes, the random effects modeling has been retained because it is important for the student audience to be introduced to these models. For the categorical designs, I use traditional ANOVA as the paradigm of analysis. However, due to my interest in industrial and engineering statistics, I also emphasize experiments with continuous factors, including two-level factorial, fractional factorial, and response surface designs. I want my students to understand that ANOVA and regression are the same model, a point that I failed to appreciate as I was learning these methods. Since my students are typically statistics majors, the course not only introduces how to design and analyze experiments, but also includes modeling techniques that are more generally applicable. The most important example of this is the introduction of random and mixed effects models, estimated using restricted maximum likelihood. The basic topics I cover follow Montgomery's text. The other recommended textbooks offer additional perspective and examples regarding this material.

Stimulated by the controversy regarding the reproducibility of experimental results in the social and biomedical sciences (e.g., Ioannidis 2005; Open Science Collaboration 2015) and the role that p-values have played in the problem (Wasserstein and Lazar 2016; Wasserstein, Schirm, and Lazar 2019), I have begun to strongly emphasize the difference between confirmatory and exploratory experiments (Jaeger and Halliday 1998; Wagenmakers et al. 2012). Roughly speaking, confirmatory experiments have been analyzed according to a prespecified plan using a prespecified model, while exploratory analyses are unencumbered by such preplanning. Confirmatory analyses severely restrict the analyst in what they can learn from the experiment, but reduce the chance that they will make spurious conclusions based on unreliable p-values obtained from some version of the forking paths fallacy (Gelman and Loken 2013). An exploratory experiment allows much more freedom to explore the data and generate hypotheses for future experiments, but formal inference cannot reasonably be performed in such cases. I believe it is critical for students to consider these issues so they will not perpetuate bad statistical practice to which even statisticians may fall prey.

4.1.2. Role of Software

In our statistics program, students learn both R and SAS. In my experimental design class, we focus on SAS/PROC MIXED to handle both basic ANOVA and more complicated random effects models. This is mostly an artifact of the course I inherited, but reflects a desire for students to see a variety of software tools during their training. We use JMP for experiments such as fractional factorial designs and response surface designs, because it has been created to easily design, randomize, and analyze these sorts of experiments.

4.1.3. Teaching Methods

The class is delivered in a fairly traditional manner, with two lecture sessions and one lab session each week. During the lectures, students take guided notes using a provided skeleton outline. Active learning is promoted via periodic in-class assignments during the lecture periods which require the students to think through material recently introduced, as well as the weekly labs in which the students perform activities, most often using SAS or JMP, under the guidance of the instructor. The class also typically includes a midterm and final exam as well as two projects.

Along with my desire that my students not contribute to the ongoing replication crisis, I also want them to have experience conducting real experiments. That this is a good idea is well-established in the literature (see the literature review in Section 2). I have found that important aspects of performing experiments are not easily appreciated in a classroom setting. For example, randomization can be expounded and even demonstrated in lecture examples, but this is quite different from recognizing experimental units in a real experiment and randomly applying the treatments to those units as specified by your design. Another example is the determination of factor levels in two-level experiments. In class, these levels simply appear and we discuss how to code them into a $[-1,+1]$ interval, but in a real experiment these levels require careful thought and perhaps some initial experimentation. Also, teaching the steepest ascent method, a commonly used early step in response surface methodology, invariably results in student confusion, though students may not even realize that they don't understand the difference between the predicted and true response along the path of steepest ascent until they have to implement this during a real experiment.

This belief in the necessity of real experimentation is influenced by my curious experience of completing (a) an undergraduate degree in industrial engineering, (b) a master's degree in statistics and operations research, (c) a Ph.D. in statistics and operations research, and (d) a dissertation on experimental design, all without ever having conducted a real experiment that I had designed. In order that this not be true of my own students, they perform at least two real experiments in the class. Most recently, I have asked students to design and execute a multi-factor exploratory experiment and follow it up at the end of the semester with a simple, one-factor confirmatory experiment. This combines two of the most important things I wish for the students to take from the class: the experience of doing real experiments and an understanding of reproducibility.

4.1.4. Future Evolution

There is a growing understanding that experimental design remains important in the data science era (e.g., Settles 2009; Jones-Farmer 2019; Stevens 2020). I would like to incorporate data science applications within the course, and possibly a module that addresses some of the particular challenges associated with online experiments. Second, pedagogically I would like to redesign and streamline the course into a modular structure that would include fundamental modules (one-factor and multi-factor completely randomized designs; block designs; two-level factorial designs; designs with random effects), but with some additional modules (examples: fractional factorial designs; response surface designs; designs for data science; optimal design; screening designs) that are optional and could be swapped in and out depending on the semester. Third, it is worth considering whether R would be a more useful software to use, given its ubiquity in industry and data science.

4.2. The Data Science DOE Course, Designed Using Online Experimentation as a Source for Examples, Assignment Problems, and Final Project (Stevens)

Prior to my current appointment, I (Stevens) taught in the M.S. in Data Science (MSDS) program at the University of San Francisco. While in that environment I became aware of the prevalence of designed experiments in the realm of data science, and the utility of DOE in a data scientist's tool belt. For that program, I developed a DOE course aimed at preparing data scientists for online controlled experiments (OCEs). Such a course necessarily contains a host of topics beyond the scope of a traditional DOE course, but I was nonetheless inspired to incorporate some of those ideas in the more traditional course that I now teach at the University of Waterloo and that I discuss below. My course development has been strongly motivated by the need to teach a core, rather traditional, set of experimental design topics to a wide spectrum of students (see Table 3), mixed with my desire to pique their interest and expose them to the interesting and modern application area of OCEs.

4.2.1. Topics

All of the experimental designs and analyses in the course are scaffolded by the QPDAC (*Question*, *Plan*, *Data*, *Analyze*, *Conclusion*) framework for statistical investigations (MacKay and Oldford 2000; Steiner and MacKay 2005). I review hypothesis testing and two-group comparisons, multi-group comparisons and the multiple comparison problem. We cover randomized complete block designs, Latin square designs, full factorial designs, 2^k factorial and 2^{k-p} fractional factorial designs, central composite designs, and response surface methodology. Throughout all of this content I augment treatment of continuous responses with discussion of binary responses, and I teach these topics primarily from a regression standpoint, emphasizing linear and logistic regression equally. We also discuss computational approaches to inference, such as the randomization test. Montgomery (2019) serves as a broad, approachable text for the course material, particularly for the undergraduates. However, for the sake of the graduate students in the course, I supplement this with Wu and Hamada (2011) for an alternative perspective and added statistical rigor.

The applications considered in this course, whether in the form of lecture examples, assignment questions, or final project problems, are drawn from the world of online controlled experiments (Kohavi, Tang, and Xu 2020). Although DOE has traditionally been applied in the realms of agriculture, manufacturing, pharmaceutical development, and the physical and social sciences, in recent years, designed experiments have become commonplace within internet and technology companies for product development/improvement, customer acquisition/retention, and just about anything that impacts a business's bottom line. In fact, it has been reported that companies such as Google, Amazon, Facebook, and Microsoft each run in excess of 10,000 experiments per year (Kohavi and Thomke 2017). And these experiments can be quite lucrative. For instance, Google's infamous 41 shades of blue experiment reportedly increased annual revenue by $200 M;[4] Bing generated an additional $100 M in annual revenue by changing the way the search engine displayed ad headlines (Kohavi and Thomke 2017); and Barack Obama raised $60 M in donations during his 2008 U.S. presidential campaign by optimizing the campaign website with a factorial experiment (Siroker 2010; Siroker and Koomen 2013). Anecdotally, I have found that weaving this tangible, relatable, and exciting application area through the more traditional content increases student interest, stimulates their motivation, increases engagement, and improves retention and understanding. It also serves as a modern source of open research problems that may be of interest to the graduate students in the course.

4.2.2. Role of Software

I exclusively use R in my course, and I heavily emphasize its use for the automation of analyses. The students engage with R passively during lecture and tutorial examples in-class, and actively when working on assignments and the final project outside of class.

4.2.3. Teaching Methods

The course is delivered by lectures (160 min per week) and tutorials (50 min per week), and the students are assessed using assignments, quizzes, a project, and a final exam. The lectures and tutorials are the same for both undergraduate and graduate students, but their assessments differ, typically with graduate students receiving additional or alternative problems with an increased level of difficulty. These assessments tend to be atypical relative to a traditional course. To emphasize the real-world relevance of the material, I manufacture new assessments each term using inspiration from recent tech blogs (e.g., posts about

[4] *https://www.theguardian.com/technology/2014/feb/05/why-google-engineers-designers*

A/B testing from Netflix,[5] Airbnb,[6] Spotify[7], or Lyft[8]) and current data science job ads for roles that explicitly require expertise in the design and analysis of experiments. I then embed hyperlinks to the blog posts and job ads in the assessment for students to browse. These contexts provide the motivation for definition questions, analysis questions, interview-style communication questions, and even derivation questions. This tactic has been met with very positive feedback, with some students going so far as to call the assessments *fun*. Several examples are included in the Supplementary Material (see SMD).

The culmination of course material and this pedagogical approach is the final project in which the students embark on a Netflix-inspired experimental investigation with a hypothetical problem and a web-based response surface simulator. The problem is motivated by a job ad[9] from 2016 for a Senior Data Scientist in Streaming Experimentation and Modeling who would "design, run, and analyze A/B and multivariate tests," "analyze experimental data with statistical rigor," and "adapt existing methods such as Response Surface Methodology (RSM) to online A/B testing." In particular, the students seek to determine which factors and which factor levels minimize, on average, the length of time it takes a user to decide what to watch. Unlike my assignment problems, in which the students analyze experimental data I provide, a key feature of the project is the need for the students to design their own experiments and collect their own data. They do so using a simulator I have constructed[10] that is akin to the garden sprinkler simulator (De Ketelaere et al. 2014), Watfactory (Steiner and MacKay 2009) or Gramacy's Experiment Game (Gramacy 2020). The simulator requires students to consider the practical problem of balancing accuracy, precision, and efficiency (by way of budget) when exploring an unknown response surface. The students document the design, analysis, and decisions associated with their experiments in a final report. They are evaluated on their communication, the sensibility of the choices they make, the efficiency of their investigation, and the accuracy of their predicted optimum. Students tell me the realistic and comprehensive nature of the project provides a rich talking point in interviews for full-time employment after graduation. The project description for a recent offering of the course is available on the simulator homepage.

4.2.4. Future Evolution

As noted above, I currently teach this course using R. While R is the software of choice for statisticians, and although it has some representation in the world of data science, Python is widely used by the data scientists typically running OCEs (Luna 2022, Anaconda 2021). Because this course is offered to fulfill a *statistics* requirement for many of the students that take it, I'm unlikely to replace R with Python entirely, but I do hope to develop a repository that contains all of the course's worked computational examples in Python.

4.3. The Case-Based Active Learning DOE Course, Designed so That Engineering Students Can Use DOE in Industry and Research (Asscher)

My (Asscher's) course focuses on the principles and application of DOE. I use a wide variety of case-based active learning teaching methods, and rely heavily on the DOE tools available in JMP. All of my choices are motivated by the need to bridge the gap between theory and practice, and are strongly influenced by my experience working as a consultant in industry and academia and teaching DOE courses in industry.

4.3.1. Topics

My DOE course covers a fairly standard menu of experimental designs popular in industry: two-level full and fractional factorial designs with and without center points, blocking, response surface designs, random and fixed effects, nesting, split plot designs, robust parameter design and Gage Repeatability and Reproducibility (GRR) studies.

The topics of Definitive Screening Designs (Jones and Nachtsheim 2011) and Optimal Design (Goos and Jones 2011) are covered briefly. The focus is first on the choice of design based on the problem at hand, followed by the choices made in constructing a particular design, for example which factors to include, their levels and the number of replicates. Theory is kept to a minimum. Analysis is also taught, with an emphasis on how properties of the analysis are determined by the design: how do the standard errors of the effects depend on the number of runs and the variation in the response; or, how does the full initial model that can be fitted depend on the choice of design. In the analysis, attention is also devoted to communicating results in addition to reaching conclusions.

The scope of the issues considered in my course regarding the application of statistical DOE in real situations is ambitious. For example, when we run a sequence of experiments on an experimental system that is simpler than the true, full scale process (e.g., a pilot plant or a reduced scale process with small batches), we must both check the stability of the experimental system over time and compare it to the true process. Additional application topics include: the collection and precursory analysis of existing data; documentation; addition of reference runs (e.g., runs at conditions either known from an existing process, or included in previous experimentation, or currently conjectured to be optimal); sequential designs; choice of measurement; dealing with variation in raw materials; choice of strategy (e.g., explore good but expensive conditions to show process feasibility vs. minimizing cost and maximizing production); identification of covariates; identification and examination of assumptions; extension of standard GRR experiments to more realistic measurement systems. These types of application topics have been discussed in Box, Hunter and Hunter (2005) and Goos and Jones (2011), the latter of which integrates application and theory using a case study format. An aspect of DOE that gets special attention in my course is its lexicon, since I have identified this as a major source of confusion for students. For example, a simple

[5] https://netflixtechblog.com/decision-making-at-netflix-33065fa06481
[6] https://medium.com/airbnb-engineering/experiments-at-airbnb-e2db3abf39e7
[7] https://engineering.atspotify.com/2020/10/29/spotifys-new-experimentation-platform-part-1/
[8] https://eng.lyft.com/a-b-tests-for-lyft-hardware-570330b488d4
[9] https://www.linkedin.com/jobs/view/senior-data-scientist-streaming-experimentation-and-modeling-at-netflix-139384997/
[10] https://nathaniel-t-stevens.shinyapps.io/Netflix_Simulator_v2/

exercise early in the course where students describe designs to each other using terms such as runs, replicates, treatments, effects, and confounding helps them to follow lectures. The menu of designs in my course and the list of application issues are relevant to a wide variety of industries, including pharmaceutical, semiconductor, testing, plastic, medical devices, agriculture, food and defence. A focus on a particular industry would dictate additional topics, for example DOE for irrigation systems requires topographical covariates; DOE for paint requires mixture designs. I recommend Montgomery's text but do not follow it closely or rely on it, as English is a second or third language for my students.

4.3.2. Role of Software

I rely on JMP in my DOE course, using it in three different ways as illustrated by the following example regarding teaching split plot designs. The first approach is to teach the topic of split plot designs without any use of software, then show how to build them using JMP. This separation approach is software neutral. The second approach is to open JMP's split plot design tool, and then teach the meaning of the inputs in the order required by the software, for example, which factors are "hard to change" and which are "easy to change"? Should we add interactions to the default model? Here JMP provides the motivation for learning each element of the topic. The third is more innovative: the students use the JMP tools for designing, evaluating and comparing experiments to explore alternative strategies for choosing a split plot design. Here the topic and the JMP tools are learned simultaneously. As an example, all three approaches were described for the topic of split plot designs; in practice I use only one for each topic.

Beyond my use of JMP in the classroom for demonstration and simulation, I teach JMP by providing the students with short video clips. The teaching material on the JMP site[11] is also used. Note that the transcripts available for this material are useful for students for whom English is a second language.

4.3.3. Teaching Methods

The key features of my teaching are variety and active learning. Variety is essential: it adds interest and enables different aspects of the course to be taught using different teaching methods. The teaching methods I use include real projects, simulators, and in-class case-based active learning workshops, in addition to lectures and tutorials with student participation. I have invested considerable time and energy into learning new teaching methods, participating in local and international conferences on active learning in engineering and problem- and project-based methods. When I developed my own in-class case-based active learning workshops, I invited teaching experts to visit my classes and provide feedback. These workshops are given in stages, with students working in pairs or small groups. Each workshop focuses on a particular aspect of experimental design. Some include the use of JMP.

One of the principles I use in preparing these workshops is to show parallel scenarios (i.e., alternative possible versions of the problem and of the results). This addresses a problem that I have experienced when teaching with both simulators and regular case studies: when students discover that one experimental design is superior in a particular situation, they often infer that it is always superior. One example of an in-class case-based active learning workshop is included in the Supplementary Material (see SME), with notes elaborating the teaching objectives and how to use the workshop in the classroom.

The first of three real experiments that the students conduct at home is a small two factor experiment investigating how freezing candles affects their burn time. This assignment is given at the beginning of the course, and the problems that arise due to a lack of randomization, poor choice of materials etc. provide motivation for topics such as blocking. The second is a five-factor experiment, with each student choosing their own process. The third experiment is a GRR experiment.

4.3.4. Recent and Future Evolution

In the last two years, I have changed the way I use JMP to teach, exploring different approaches as discussed above. This change is in response to both improvements in the software and a change in the skills of the students. Students now quickly embrace new software and are motivated to learn how to use it.

Future changes are a challenge. I frequently identify an important topic that I would like to add to my DOE course, but I never identify a topic to remove. I would like to devote more time to definitive screening designs, and to include online experimentation. A major change that I have considered making is limiting the topic of full and fractional two-level factorial designs while extending the use of optimal design, allowing treatment of design space constraints, and sequential designs. The problem with this plan is that an understanding of principles such as confounding and power is needed to use optimal design, and these principles are easier to teach using classical two-level designs.

4.4. The Optimal Experimental Design Course, Designed so Students Are Equipped to Design Experiments in Both Standard and Unusual Settings (Goos)

When I (Goos) took over the regression class as well as the experimental design class, I made two major changes to their content: First, I moved ANOVA from the experimental design class (a third year undergraduate course) to the regression class (a second year undergraduate course), since it can be viewed as a special case of regression and since regression with one or more categorical explanatory variables is included in that class. In the regression class, students use matrix algebra and learn about quantifying the uncertainty concerning parameter estimates and predictions, and about the importance of avoiding multicollinearity (referred to as aliasing and confounding in DOE classes). Second, I decided to use the theory and philosophy of optimal experimental design as the foundation of the experimental design class. All of my choices were motivated by the fact that the course is a compulsory one for engineering students, who need flexible tools to tailor experimental designs to engineering problems in their professional lives, rather than a limited set of standard experimental designs (which may inspire them to redefine their future problems to fit the standard designs they happened to encounter during their DOE course).

[11] https://www.jmp.com/en_us/online-statistics-course/design-of-experiments.html

4.4.1. Topics

The course starts with examples of successful multi-factor experiments from various industries, illustrating the usefulness of experimental design for product and process optimization and for tackling engineering problems, refreshing the main concepts from regression analysis, and highlighting the fact that experimental factors interact and may have nonlinear effects on the response(s). There is substantial emphasis on interactive graphical representations of regression models (e.g., contour plots and prediction profilers).

The attention then shifts to full and fractional factorial designs, orthogonality and aliasing, the weaknesses of one-factor-at-a-time experimentation, and a discussion of practical constraints that often prevent us from using standard experimental designs. At this point, I introduce the flexible optimal experimental design approach, and demonstrate its usefulness by performing a virtual experiment in class, where the students define the active effects, and I use the optimal experimental design to detect their specified effects.

Starting from the fifth lecture, I follow the textbook *Optimal Design of Experiments: A Case Study Approach* (Goos and Jones 2011). It turns out that engineering students like the book a lot because it uses case studies from various industries (food industry, metal industry, chemical industry, etc.) and because the case studies are introduced through a dialog between two consultants and a consultee.

Roughly speaking, the remaining lectures of the course each deal with one chapter from the textbook. Topics covered include D- and I-optimal experimental designs for completely randomized experiments, follow-up experiments, mixture experiments, blocked experiments, split-plot and strip-plot experiments.

When discussing the analysis of data from blocked experiments, both random block effects and fixed block effects are treated. For my students, this is their first acquaintance with random effects and generalized least squares. Once familiar with the concept of random block effects, the analysis of data from split-plot and strip-plot experiments is logical. Due to the fact that I teach students in bio-science engineering, I also discuss the agricultural origin of split-plot and strip-plot designs as well as traditional designs for field trials. I show that these designs fall under the umbrella of optimal experimental designs and can be generated with standard software for optimal experimental design.

4.4.2. Role of Software

In consultation with the program director and representatives from all fields of study involved, we decided to use JMP for all statistics courses in the B.Sc. in Bioscience Engineering. The use of JMP offers the advantage that it is easy to discuss all of the topics in the DOE textbook and to use interactive tools for analysis and product and process optimization. This is extremely appealing to engineering students.

In the M.Sc. program in statistics, I teach roughly the same course on DOE as in the B.Sc. program in bioscience engineering and I also use JMP, while the vast majority of teachers in the program use R and a few use SAS. Quite a few M.Sc. students in statistics are hostile to my use of JMP, and ask me why I do not use R. My reply is that advanced DOE is implemented in a better and more user-friendly way in JMP than it is in R; and that future statisticians should be aware of the existence of a broad range of statistical software packages, each of which has strengths and weaknesses. I then also explain that, in the event they become statistical consultants/trainers, they will be confronted with consultees/trainees who have their own favorite statistical packages.

4.4.3. Teaching Methods

My course consists of ordinary 2-hr lectures (16 in total) and 2-hr exercise sessions in PC labs, where JMP is available (10 in total). Students are also required to download JMP on their personal laptops because there is one major assignment during the semester.

Each lecture generally starts with the motivating case study from the textbook, explanations of any new concepts and/or theory needed to tackle the case study, and a demonstration of how to generate alternative designs for the problem at hand, how to evaluate the quality of the designs, how to analyze the data resulting from the experiment, and how to translate the results of this analysis into optimal settings for the product or process under investigation. During each lecture, students thus see how a practical problem is tackled and solved using DOE and regression analysis.

The exercise sessions involve DOE problems, statistical analyses and interpretations. For the exercises, the students are partitioned into groups of 30–40 students, and expected to work out the exercises themselves, but they can ask tips and tricks from the teaching assistant present. Some of the exercises involve modern experimental designs such as definitive screening designs (Jones and Nachtsheim 2011) and orthogonal minimally aliased response surface designs (Núñez Ares and Goos 2020), as soon as the students have sufficient technical knowledge to judge the pros and cons of these designs.

An important part of the assessment in the course is the project assignment. In the assignment, the students have to conduct a (virtual) screening experiment to study the impact of eight factors on the performance of a garden sprinkler (De Ketelaere et al. 2014). After identifying the active factors, the students have to conduct a follow-up, response surface experiment, to study these active factors in more detail. Finally, the students have to identify the settings of the factors that optimize three different responses, while taking into account the cost of the solution chosen. I opt for a virtual experiment for logistic and efficiency reasons.

My course is concluded with a written open book exam, which tests whether students know how to tackle a new case study, whether they are able to interpret JMP output concerning designs and the corresponding data analyses correctly, and whether they are able to perform certain technical calculations by hand.

4.4.4. Future Evolution

Because I teach optimal experimental design and this framework is flexible enough to cope with the complex nature of 21st century products and processes in industry, I do not plan major changes in the way I teach DOE to bioscience engineering students. For the M.Sc. students in statistics, who generally are interested in applications outside industry, it would be good to include some of the less traditional application areas of DOE, such as online controlled experiments, marketing experiments and discrete choice experiments.

5. Discussion and Conclusions

The design and analysis of experiments is practically useful in a wide variety of fields. DOE courses are therefore compulsory in many undergraduate and graduate-level statistics degree programs. In this article, we explored and showcased the details of these courses: who takes them, who teaches them, how they are taught, how they are assessed, and how they are evolving. Through the composite profile based on 50 survey respondents, as well as essays from the four authors, we have seen considerable similarities among DOE courses, but also notable differences.

One common aspect of DOE courses is that they are populated by students from a variety of backgrounds, including statistics, business, engineering, computer science, data science, the life sciences, and social sciences. We find that roughly half of DOE instructors are not DOE researchers. This does not appear to lead to a material difference in topics covered, teaching and assessment methods, or plans for future evolution of the course.

With respect to topics covered, we find that nearly all surveyed DOE courses cover randomization, replication, and blocking design principles, as well as single-factor experiments and multifactor experiments, with blocking designs and factorial designs. We also find that the majority of these courses cover fractional factorial designs, split-plot designs, repeated measures designs and analyses based on fixed, random, and mixed effect models. Aside from these core topics, extensive variation exists in the additional topics covered. We note that optimal design, robust parameter design, definitive screening designs, and mixture experiments are taught very rarely. Figure 1 provides further insights regarding these general statements. Despite widespread interest in observational causal inference (e.g., Holland 1986; Pearl and Mackenzie 2018, Gelman and Vehtari 2021), and the conceptual overlap with DOE insofar as identifying and quantifying causal relationships are concerned, we found such methods absent from the surveyed DOE courses. Though these causal inference approaches are strongly related to missing data, most of the work in missing data for traditional, small-sample DOE (e.g., Akhtar and Prescott 1986; Imhof, Song, and Wong 2002; Ahmad and Gilmour 2010; Wongoutong 2022) has not used the causal inference framework. Notable exceptions include the consideration of causal inference methods in the context of two-level factorial designs (e.g., Dasgupta, Pillai, and Rubin 2015; Espinosa, Dasgupta, and Rubin 2016; Pashley and Bind 2022). To date, it appears that observational causal inference tends to be taught in dedicated courses, often within biostatistics or econometrics curricula, rather than as part of a traditional experimental design class.

With respect to teaching and evaluation methods, the traditional lecture—assignment—test—exam framework appears still to be quite common; relatively few DOE classrooms incorporate more recent instruction and assessment methods, such as flipped, blended, or hybrid learning. See Figures 2 and 3 for deeper insights beyond these general remarks. We do find that in many DOE courses, the students conduct their own experiments, either physically or virtually; there is consensus that such exercises are helpful, if not necessary, to provide students with exposure to the more practical aspects of experimentation, such as choosing factors and levels, measurement procedures, implementing randomization, doing data collection, etc.

We find that the use of software for both instruction and assessment is ubiquitous. Although there is some variation in the statistical software of choice in DOE classes, DOE practitioners in business and industry use a much broader range of specialized DOE packages. Nearly all classrooms incorporate a textbook in some manner, though there is no standard textbook that is uniformly adopted, and there is disagreement as to whether there exists a single textbook that adequately treats all of the material one may wish to cover in a DOE course. The number of courses that adopt multiple textbooks suggests that such a text does not exist. Several survey respondents lamented the lack of modern motivations and examples in existing DOE texts, but they did not indicate why they use some books over others, so we cannot provide guidance on textbook selection beyond the justification of our own choices.

With respect to our own approaches to DOE pedagogy, like the survey respondents, we find both commonalities and differences. We all agree that real or virtual experiments are a necessary component of a DOE course; these force students to grapple with the application and not just the theory of DOE. They also require that students be able to communicate about both the design and analysis of experiments. We each use active learning methods in our teaching as well. What we do specifically differs depending on our students, the context of our course, our experiences, and the broader theme we have set for the course, so we do not aim to make *specific* recommendations about what to include in a DOE course and how to teach it. As we have seen, there are many effective approaches. That said, we do make the following *general* recommendations to a reader who wants to either design a new DOE course or overhaul or tweak an existing one: (i) consider the needs of the population of students taking the course, (ii) experiment with multiple teaching methods, and (iii) get students designing, running, and analyzing experiments themselves. As demonstrated by the four essays, there is room for the unique interests and influences that each instructor brings to the classroom.

Where is the pedagogy of DOE going in the future? We do not have a clear answer to that question. There is little indication from the survey that, at the time of responding, instructors were consciously moving to more active learning, hybrid, or flipped models of teaching. However, the survey was administered at the beginning of the pandemic, so we expect many DOE instructors *have* successfully adopted these modes of teaching since then. The applied nature of DOE makes these courses well-suited for these instruction methods. We certainly expect the connections between data science and experimentation to continue to develop, and as more resources become available, we would expect more instructors to dedicate examples and even parts of their courses to online applications. If this is the case, we may not see a large-scale adoption of newer design methods, such as optimal designs or definitive screening designs, in mainstream DOE instruction. Instead, we could see these more specialized methods being taught to audiences—for example, engineering students—who are more likely to use them, while more general DOE courses would include additional topics related to data science. Another possibility is that in the future, online experimentation will be taught mainly by non-statisticians, in courses separate from the traditional DOE courses. This seems to be what happened, for instance, to DOE instruction in the social sciences, which in our observation is largely conducted by non-statisticians in separate programs. Finally, we note that only two of the survey respondents taught courses specifically for doctoral students in statistics. We conjecture that this reflects an impression among statisticians outside of DOE that this area is not ripe for innovative research. However, this ignores the active, though small, community of researchers who are working in areas such as computer experiments, screening experiments, optimal design, etc, along with the emerging and impactful research opportunities in the realm of online controlled experiments (Larsen et al. 2022). We hope that this article and these exciting research avenues will help to rejuvenate interest in DOE training at all levels of instruction, including the Ph.D. level.

Supplementary Materials

Supplementary Material A (SMA) is a narrative version of the literature review. Supplementary Material B (SMB) is comprised of the list of the departments to which our survey was sent. Supplementary Material C (SMC) gives the survey questions and a summary of responses to them. Supplementary Material D (SMD) provides examples of assessments related to Online Controlled Experiments. Supplementary Material E is an example of a case-based active learning assignment.

Acknowledgments

The first author would like to thank Ellen Yezierski and the Discipline-Based Education Research Associates program in the Center for Teaching Excellence at Miami University, which originally inspired this work. We are grateful to Christine Anderson-Cook for providing helpful feedback on an early draft of the survey. We would also like to thank Changbao Wu for guidance on administering the survey. Jacqueline Asscher is a member of the JMP Europe Steering Committee and is active in the Israel JMP Users Group.

Disclosure Statement

Jacqueline Asscher is a member of the JMP Europe Steering Committee and is active in the Israel JMP Users Group. Besides this, the authors have no financial or non-financial competing interests to report.

References

Ahmad, T., and Gilmour, S. G. (2010), "Robustness of Subset Response Surface Designs to Missing Observations," *Journal of Statistical Planning and Inference*, 140, 92–103. https://www.sciencedirect.com/science/article/abs/pii/S0378375809001955.

Akhtar, M., and Prescott, P. (1986), "Response Surface Designs Robust to Missing Observations," *Communications in Statistics - Simulation and Computation*, 15, 345–363. https://www.tandfonline.com/doi/abs/10.1080/03610918608812512.

American Statistical Association. (2014), "Curriculum Guidelines for Undergraduate Programs in Statistical Science," available at https://www.amstat.org/asa/files/pdfs/EDU-guidelines2014-11-15.pdf .

Anaconda. (2021), "On the Path to Impact," 2021 *State of Data Science Report*, Web Article (last accessed April 14, 2022), https://www.anaconda.com/state-of-data-science-2021.

Anderson-Cook, C. M. (1998), "Designing a First Experiment: A Project for Design of Experiment Courses," *The American Statistician*, 52, 338–342. https://www.tandfonline.com/doi/abs/10.1080/00031305.1998.10480592.

Anderson-Cook, C. M., and Dorai-Raj, S. (2001), "An Active Learning in-Class Demonstration of Good Experimental Design," *Journal of Statistics Education*, 9. https://www.tandfonline.com/doi/full/10.1080/10691898.2001.11910645.

Anderson-Cook, C. M., and Lu, L. (2023), "Is Designed Data Collection Still Relevant in the Big Data Era?" *Quality and Reliability Engineering International*, 39, 1085–1101. DOI:10.1002/qre.3326.

Andrews, S. (2021), *The Case Study Companion: Teaching, Learning and Writing Business Case Studies*, Abingdon: Routledge.

Antony, J. (2014), *Design of Experiments for Engineers and Scientists*, London: Elsevier.

Antony, J., and Capon, N. (1998), "Teaching Experimental Design Techniques to Industrial Engineers," *International Journal of Engineering Education*, 14, 335–343. https://enbis.org/wp-content/uploads/2018/09/3299_4928249026.pdf .

Binnie, N. (2004), "Using EDA, ANOVA and Regression to Optimise Some Microbiology Data," *Journal of Statistics Education*, 12, 2. https://www.tandfonline.com/doi/full/10.1080/10691898.2004.11910738.

Blades, N. J., Schaalje, G. B., and Christensen, W. F. (2015), "The Second Course in Statistics: Design and Analysis of Experiments?," *The American Statistician*, 69, 326–333. https://www.tandfonline.com/doi/abs/10.1080/00031305.2015.1086437.

Box, G. E. (1992), "Teaching Engineers Experimental Design with a Paper Helicopter," *Quality Engineering*, 4, 453–459. https://www.tandfonline.com/doi/abs/10.1080/08982119208918925.

Box, G. E., and Draper, N. R. (1987), *Empirical Model-Building and Response Surfaces*, New York: Wiley.

Box, G. E., Hunter, J. S., and Hunter, W. G. (2005), "Statistics for Experimenters," in *Wiley Series in Probability and Statistics*. Hoboken, NJ: Wiley.

Bulmer, M. (2003), "Growing Virtual Plants for Teaching and Learning Statistics," in *Proceedings of the Apple University Consortium Academic and Developers Conference*, pp. 3-1–3-7. Available at https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=f09b284b8cf5f5e9e73fad981425a9e5c769b77e.

Bulmer, M. (2012), "Virtual Worlds for Teaching Statistics," *International Journal of Innovation in Science and Mathematics Education*, 11. https://openjournals.library.sydney.edu.au/CAL/article/view/6061.

Bulmer, M., and Haladyn, J. K. (2011), "Life on an Island: A Simulated Population to Support Student Projects in Statistics," *Technology Innovations in Statistics Education*, 5. https://escholarship.org/uc/item/2q0740hv.

Campbell, D. T., and Stanley, J. C. (2015), *Experimental and Quasi-Experimental Designs for Research*, Ravenio Books.

Carver, R., Everson, M., Gabrosek, J., Horton, N., Lock, R., Mocko, M., Rossman, A., Roswell, G. H., Velleman, P., Witmer, J., and Wood, B. (2016), "Guidelines for Assessment and Instruction in Statistics Education (GAISE) College Report 2016," available at https://commons.erau.edu/publication/1083.

Chance, B., and Peck, R. (2014), "From Curriculum Guidelines to Learning Objectives: A Survey of Five Statistics Programs," arXiv preprint, https://arxiv.org/abs/1412.7261.

Darius, P. L., Portier, K. M., and Schrevens, E. (2007), "Virtual Experiments and Their Use in Teaching Experimental Design," *International Statistical Review*, 75, 281–294. https://onlinelibrary.wiley.com/doi/10.1111/j.1751-5823.2007.00028.x.

Dasgupta, T., Pillai, N. S., and Rubin, D. B. (2015), "Causal Inference from 2^K Factorial Designs by Using Potential Outcomes," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 77, 727–753. https://www.jstor.org/stable/24775307.

Davies, O. L. (1954), *The Design and Analysis of Industrial Experiments*, Oliver & Boyd.

De Ketelaere, B., Siebertz, K., van Bebber, D., and Rutten, K. (2014), "The Garden Sprinkler: An Interactive Web-Based Application for Teaching Design of Experiments," *ICOTS9 Proceedings*, https://iase-web.org/icots/9/proceedings/pdfs/ICOTS9_C266_RUTTEN.pdf .

Dean, A., Voss, D., and Draguljić, D. (2017), *Design and Analysis of Experiments* (2nd ed.), Cham: Springer.

Deming, S. N., and Morgan, S. L. (1983), "Teaching the Fundamentals of Experimental Design," *Analytica Chimica Acta*, 150, 183–198. https://www.sciencedirect.com/science/article/abs/pii/S0003267000854707.

Deming, S. N., and Morgan, S. L. (1993), *Experimental Design: A Chemometric Approach*, Amsterdam: Elsevier.

Dunn, P. K. (2013), "Comparing the Lifetimes of Two Brands of Batteries," *Journal of Statistics Education*, 21. https://www.tandfonline.com/doi/abs/10.1080/10691898.2013.11889666.

Easterling, R. G. (2004), "Teaching Experimental Design," *The American Statistician*, 58, 244–252. https://www.tandfonline.com/doi/abs/10.1198/000313004X1477.

Espinosa, V., Dasgupta, T., and Rubin, D. B. (2016), "A Bayesian Perspective on the Analysis of Unreplicated Factorial Experiments Using Potential Outcomes," *Technometrics*, 58, 62–73. https://www.tandfonline.com/doi/abs/10.1080/00401706.2015.1006337.

Fillebrown, S. (1994), "Using Projects in an Elementary Statistics Course for Non-Science Majors," *Journal of Statistics Education*, 2. https://www.tandfonline.com/doi/full/10.1080/10691898.1994.11910470.

Fisher, R. A. (1971), *The Design of Experiments* (9th ed.), New York: Hafner Press, MacMillan Publishing Co.

Garvin, D. A. (2003), "Making the Case," *Harvard Magazine*, 106, 56–65. https://sitios.itesm.mx/va/dide2/tecnicas_didacticas/casos/articulod.garvin.pdf.

Gelman, A., and Vehtari, A. (2021), "What Are the Most Important Statistical Ideas of the past 50 Years?" *Journal of the American Statistical Association*, 116, 2087–2097. https://www.tandfonline.com/doi/full/10.1080/01621459.2021.1938081.

Gelman, A., and Loken, E. (2013), "The Garden of Forking Paths: Why Multiple Comparisons Can Be a Problem, Even When There is no 'Fishing Expedition' or 'p-hacking' and the Research Hypothesis Was Posited Ahead of Time," working paper, Columbia University, New York, available at http://stat.columbia.edu/~gelman/research/unpublished/forking.pdf.

Goh, T. N. (2001), "A Pragmatic Approach to Experimental Design in Industry," *Journal of Applied Statistics*, 28, 391–398. https://www.tandfonline.com/doi/abs/10.1080/02664760120034126.

Goos, P., and Jones, B. (2011), *Optimal Design of Experiments: A Case Study Approach*, Chichester: Wiley.

Goos, P., and Leemans, H. (2004), "Teaching Optimal Design of Experiments Using a Spreadsheet," *Journal of Statistics Education*, 12(3). https://www.tandfonline.com/doi/full/10.1080/10691898.2004.11910631

Gramacy, R. B. (2020), "A Shiny Update to an Old Experiment Game," *The American Statistician*, 74, 87–92. https://www.tandfonline.com/doi/abs/10.1080/00031305.2018.1505659.

Gunter, B. H., and Matey, J. R. (1993), "How Statistical Design Concepts Can Improve Experimentation in the Physical Sciences," *Computers in Physics*, 7, 262–272. https://aip.scitation.org/doi/pdf/10.1063/1.4823173.

Hane, E. N. (2007), "Use of an Inquiry-Based Approach to Teaching Experimental Design Concepts in a General Ecology Course," *Teaching Issues and Experiments in Ecology*, 5, 1–19. https://www.esa.org/tiee/vol/v5/research/hane/pdf/Hane_2007.pdf.

Hanrahan, G., and Lu, K. (2006), "Application of Factorial and Response Surface Methodology in Modern Experimental Design and Optimization," *Critical Reviews in Analytical Chemistry*, 36, 141–151. https://www.tandfonline.com/doi/abs/10.1080/10408340600969478.

Hiebert, S. M. (2007), "Teaching Simple Experimental Design to Undergraduates: Do Your Students Understand the Basics?" *Advances in Physiology Education*, 31, 82–92. https://journals.physiology.org/doi/full/10.1152/advan.00033.2006.

Holland, P. W. (1986), "Statistics and Causal Inference," *Journal of the American Statistical Association*, 81, 945–960. https://www.tandfonline.com/doi/abs/10.1080/01621459.1986.10478354.

Horton, N. J. (2015), "Challenges and Opportunities for Statistics and Statistical Education: Looking Back, Looking Forward," *The American Statistician*, 69, 138–145. https://www.tandfonline.com/doi/abs/10.1080/00031305.2015.1032435.

Howley, P. P. (2003), "Teaching How to Calibrate a Process Using Experimental Design and Analysis: The Ballistat," *Journal of Statistics Education*, 11. https://www.tandfonline.com/doi/full/10.1080/10691898.2003.11910709.

Hunter, W. G. (1977), "Some Ideas about Teaching Design of Experiments, with 2^5 Examples of Experiments Conducted by Students," *The American Statistician*, 31, 12–17. https://www.tandfonline.com/doi/abs/10.1080/00031305.1977.10479185.

Ilzarbe, L., Álvarez, M. J., Viles, E., and Tanco, M. (2008), "Practical Applications of Design of Experiments in the Field of Engineering: A Bibliographical Review," *Quality and Reliability Engineering International*, 24, 417–428. https://onlinelibrary.wiley.com/doi/10.1002/qre.909.

Imhof, L. A., Song, D., and Wong, W. K. (2002), "Optimal Design of Experiments with Possibly Failing Trials," *Statistica Sinica*, *12*, 1145–1155. https://www.jstor.org/stable/24307020.

Ioannidis, J. P. (2005), "Why Most Published Research Findings Are False," *PLoS Medicine*, 2, e124. https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.0020124.

Jaeger, R. G., and Halliday, T. R. (1998), "On Confirmatory versus Exploratory Research," *Herpetologica*, 54, S64–S66. https://www.jstor.org/stable/3893289.

Jones, B., and Nachtsheim, C. J. (2011), "A Class of Three-Level Designs for Definitive Screening in the Presence of Second-Order Effects," *Journal of Quality Technology*, 43, 1–15. https://www.tandfonline.com/doi/abs/10.1080/00224065.2011.11917841.

Jones-Farmer, L. A. (2019), "Leveraging Industrial Statistics in the Data Revolution: The Youden Memorial Address at the 63rd Annual Fall Technical Conference," *Quality Engineering*, 31, 205–211. https://www.tandfonline.com/doi/abs/10.1080/08982112.2019.1572187.

Kenett, R. S., and Steinberg, D. M. (1987), "Some Experiences Teaching Factorial Design in Introductory Statistics Courses," *Journal of Applied Statistics*, 14, 219–227. https://www.tandfonline.com/doi/abs/10.1080/02664768700000027.

Kohavi, R., Tang, D., and Xu, Y. (2020), *Trustworthy Online Controlled Experiments: A Practical Guide to a/B Testing*, Cambridge: Cambridge University Press.

Kohavi, R., and Thomke, S. (2017), "The Surprising Power of Online Experiments," *Harvard Business Review*, 95, 74–82. http://davidfrico.com/kohavi17.pdf.

Kreutz, C., and Timmer, J. (2009), "Systems Biology: Experimental Design," *The FEBS Journal*, 276, 923–942. https://febs.onlinelibrary.wiley.com/doi/full/10.1111/j.1742-4658.2008.06843.x.

Kuehl, R. O. (2000), *Design of Experiments: Statistical Principles of Research Design and Analysis*, Pacific Grove, CA: Duxbury Press.

Kuhnt, S., and Coleman, S. (2020), "Hands-On Projects for Teaching DoE" [Conference Presentation], ENBIS-20 Online, available for ENBIS Members at https://enbis.org/media-centre/sonja-kuhnt-shirley-coleman-hands-on-projects-for-teaching-doe/.

Kuhnt, S., and Coleman, S. (2021), "Hands-On Projects for Teaching DoE" [Conference presentation], ENBIS-21 Online, available for ENBIS Members at https://enbis.org/media-centre/sonja-kuhnt-shirley-coleman-hands-on-projects-for-teaching-doe-2/.

Kuiper, S. (2016), "Stat2Labs Front Page," *Stat2Labs*, available at https://stat2labs.sites.grinnell.edu/.

Larsen, N., Stallrich, J. W., Sengupta, S., Deng, A., Kohavi, R., and Stevens, N. T. (2022), "Statistical Challenges in Online Controlled Experiments: A Review of A/B Testing Methodology," arXiv preprint arXiv:2212.11366.

Lawson, J., Aggarwal, P., Leininger, T., and Fairchild, K. (2011), "Characterizing Variability in Smestad and Grätzel's Nanocrystalline Solar Cells: A Collaborative Learning Experience in Experimental Design," *Journal of Statistics Education*, 19, https://www.tandfonline.com/doi/abs/10.1080/10691898.2011.11889598.

Lazic, Z. R. (2006), *Design of Experiments in Chemical Engineering: A Practical Guide*, Weinheim: Wiley.

Leardi, R. (2009), "Experimental Design in Chemistry: A Tutorial," *Analytica Chimica Acta*, 652, 161–172. https://www.sciencedirect.com/science/article/abs/pii/S0003267009008058.

Lin Ho, S., Lyn Nge, W., and Hong Chua, K. (2004), "The Catapult Project: An Innovative Approach for Learning Statistical Design of Experiments," in *2004 IEEE International Engineering Management Conference (IEEE Cat. No. 04CH37574)*, 3, 1056–1060, https://ieeexplore.ieee.org/document/1408853.

Lindquist, E. F. (1953), *Design and Analysis of Experiments in Psychology and Education*, Boston: Houghton Mifflin.

Luca, M., and Bazerman, M. H. (2021), *The Power of Experiments: Decision Making in a Data-Driven World*, Cambridge, MA: MIT Press.

Luna, J. C. (2022), "Top Programming Languages for Data Scientists in 2022," *Datacamp Data Science Blog*, Web Article (last accessed April 14, (2022), *https://www.datacamp.com/blog/top-programming-languages-for-data-scientists-in-2022*.

Lye, L. M. (2005), "Tools and Toys for Teaching Design of Experiments Methodology," in *33rd Annual General Conference of the Canadian Society for Civil Engineering*.

MacKay, R. J., and Oldford, R. W. (2000), "Scientific Method, Statistical Method and the Speed of Light," *Statistical Science*, 15, 254–278. *https://www.jstor.org/stable/2676665*.

Mackisack, M. (1994), "What is the Use of Experiments Conducted by Statistics Students?" *Journal of Statistics Education*, 2, 1–15. *https://www.tandfonline.com/doi/full/10.1080/10691898.1994.11910461*.

Mead, R., Curnow, R. N., and Hasted, A. M. (2017), *Statistical Methods in Agriculture and Experimental Biology* (3rd ed.), Boca Raton, FL: Chapman and Hall/CRC.

Montgomery, D. C. (2019), *Design and Analysis of Experiments* (10th ed.), New York: Wiley.

Muske, K. R., and Myers, J. A. (2007), "A Realistic Experimental Design and Statistical Analysis Project," *Chemical Engineering Education*, 41, 31–38. *https://journals.flvc.org/cee/article/view/122472*.

Nolan, D., and Speed, T. P. (1999), "Teaching Statistics Theory through Applications," *The American Statistician*, 53, 370–375. *https://www.tandfonline.com/doi/abs/10.1080/00031305.1999.10474492*.

Núñez Ares, J., and Goos, P. (2020), "Enumeration and Multicriteria Selection of Orthogonal Minimally Aliased Response Surface Designs," *Technometrics*, 62, 21–36. *https://www.tandfonline.com/doi/abs/10.1080/00401706.2018.1549103*.

Oehlert, G. W. (2010), *A First Course in Design and Analysis of Experiments*, Creative Commons License, available at *http://users.stat.umn.edu/~gary/book/fcdae.pdf*.

Open Science Collaboration. (2015), "Estimating the Reproducibility of Psychological Science," *Science (New York, N.Y.)*, 349, aac4716. *https://www.science.org/doi/10.1126/science.aac4716*.

Pashley, N. E., and Bind, M. A. C. (2022), "Causal Inference for Multiple Treatments Using Fractional Factorial Designs," *Canadian Journal of Statistics*, 51, 444–468. *https://onlinelibrary.wiley.com/doi/10.1002/cjs.11734*.

Pearl, J., and Mackenzie, D. (2018), *The Book of Why: The New Science of Cause and Effect*, New York: Basic Books.

Pollock, K. H., Ross-Parker, H. M., and Mead, R. (1979), "A Sequence of Games Useful in Teaching Experimental Design to Agriculture Students," *The American Statistician*, 33, 70–76. *https://www.tandfonline.com/doi/abs/10.1080/00031305.1979.10482663*.

Pyott, L. (2021), "Tennis Anyone? Teaching Experimental Design by Designing and Executing a Tennis Ball Experiment," *Journal of Statistics and Data Science Education*, 29, 22–26. *https://www.tandfonline.com/doi/full/10.1080/10691898.2020.1854638*.

Reis, M., and Kenett, R. S. (2017), "A Structured Overview on the Use of Computational Simulators for Teaching Statistical Methods," *Quality Engineering*, 29, 730–744. *https://www.tandfonline.com/doi/abs/10.1080/08982112.2016.1272122*.

Schrevens, E., Portier, K., Darius, P., Cooman, A., Tenorio, J., and Medina, A. (2004), "Simulation-Based Environments for Practicing Data-Collection Skills in Greenhouse Experimentation," in International Conference on Sustainable Greenhouse," Acta Horticulturae, pp. 871–876. *https://www.actahort.org/books/691/691_108.htm*.

Settles, B. (2009), "Active Learning Literature Survey," Computer Sciences Technical Report 1648, University of Wisconsin–Madison, available at *https://minds.wisconsin.edu/handle/1793/60660*.

Siroker, D. (2010), "How Obama Raised $60 Million by Running a Simple Experiment," Web Article (last accessed April 14, 2022), available at *https://www.optimizely.com/insights/blog/how-obama-raised-60-million-by-running-a-simple-experiment/*.

Siroker, D., and Koomen, P. (2013), *A/B Testing: The Most Powerful Way to Turn Clicks into Customers*, Hoboken: Wiley.

Stafford, R., Goodenough, A. E., and Davies, M. S. (2010), "Assessing the Effectiveness of a Computer Simulation for Teaching Ecological Experimental Design," *Bioscience Education*, 15, 1–9. *https://www.tandfonline.com/doi/full/10.3108/beej.15.1*.

Steiner, S. H., and MacKay, R. J. (2005), *Statistical Engineering: An Algorithm for Reducing Variation in Manufacturing Processes*, Milwaukee: ASQ Quality Press.

Steiner, S. H., and MacKay, R. J. (2009), "Teaching Variation Reduction Using a Virtual Manufacturing Environment," *The American Statistician*, 63, 361–365. *https://www.tandfonline.com/doi/abs/10.1198/tast.2009.08042*.

Steiner, S. H., Hamada, M., Giddings White, B. J., Kutsyy, V., Mosesova, S., and Salloum, G. (2007), "A Bubble Mixture Experiment Project for Use in an Advanced Design of Experiments Class," *Journal of Statistics Education*, 15. *https://www.tandfonline.com/doi/full/10.1080/10691898.2007.11889458*.

Stevens, N. T. (2020), "Discussion of 'Statistics = Analytics?,' *Quality Engineering*, 32, 145–148. *https://www.tandfonline.com/doi/full/10.1080/08982112.2019.1675172*.

Tanco, M., Viles, E., Ilzarbe, L., and Alvarez, M. J. (2009), "Implementation of Design of Experiments Projects in Industry," *Applied Stochastic Models in Business and Industry*, 25, 478–505. *https://onlinelibrary.wiley.com/doi/10.1002/asmb.779*.

Thomke, S. H. (2020), *Experimentation Works: The Surprising Power of Business Experiments*, Boston: Harvard Business Press.

Wagenmakers, E. J., Wetzels, R., Borsboom, D., van der Maas, H. L., and Kievit, R. A. (2012), "An Agenda for Purely Confirmatory Research," *Perspectives on Psychological Science : a Journal of the Association for Psychological Science*, 7, 632–638. *https://journals.sagepub.com/doi/pdf/10.1177/1745691612463078*.

Wasserstein, R. L., and Lazar, N. A. (2016), "The ASA Statement on p-Values: context, Process, and Purpose," *The American Statistician*, 70, 129–133. *https://www.tandfonline.com/doi/full/10.1080/00031305.2016.1154108*.

Wasserstein, R. L., Schirm, A. L., and Lazar, N. A. (2019), "Moving to a World beyond '$p < 0$'" *The American Statistician*, 73, 1–19. *https://www.tandfonline.com/doi/full/10.1080/00031305.2019.1583913.*

Wongoutong, C. (2022), "Imputation Methods for Missing Response Values in the Three Parts of a Central Composite Design with Two Factors," *Journal of Statistical Computation and Simulation*, 92, 2273–2289. *https://www.tandfonline.com/doi/full/10.1080/00949655.2022.2027424*.

Woodard, V. (2023), "Five Hands-on Experiments for a Design of Experiments Course," *Journal of Statistics and Data Science Education*. *https://www.tandfonline.com/doi/full/10.1080/26939169.2023.2195889*

Wu, C. J., and Hamada, M. S. (2011), *Experiments: Planning, Analysis, and Optimization* (2nd ed.), Hoboken: Wiley.

Zolman, J. F. (1999), "Teaching Experimental Design to Biologists," *The American Journal of Physiology*, 277, S111–118. *https://journals.physiology.org/doi/abs/10.1152/advances.1999.277.6.s111*.