

## Common Questions in IRB Applications

### **Provide a brief lay summary describing this study.**

Secondary and postsecondary learning is a field ripe for change. Huge advances in the learning sciences, offering new theories to guide the creation of instructional experiences. We have seen also order-of-magnitude changes in technology and innovation, providing us the tools to implement new forms of learning, and to distribute them at large scale. And, we have witnessed the recent application of improvement methodologies to education--methodologies that can lead to rapid and sustained improvements in implementation.

At the same time, the core of instruction experienced by most students has changed little. Instructors still teach the same courses, and mostly in the same way. When individual instructors do innovate, they mostly do it on their own. They lack a means of collaboration with researchers, who could bring relevant theories of learning science to the table, and with designers and developers who could build their innovations in a scalable way. They lack the means to test whether the innovations they design actually improve student outcomes, both academic and otherwise. And they lack a means of sharing their approach with other practitioners. Thus, these teaching innovations are essentially inert and local in their (sometimes unknown) impact.

Our goal in this project is to break through the barriers that have thus far slowed the impact of learning sciences, improvement methodologies, and technology on the improvement of effective, personalized learning opportunities for secondary and postsecondary learners.

We will study statistics learning among students enrolled in an introductory course. The course pedagogy and design were based on research in cognitive psychology and statistics learning and align with recommendations endorsed by the American Statistical Association. We want to use the data from students' interactions with course material to understand statistics learning and improve student outcomes. Information will come from several sources. The sources include optional pre- and post-course surveys, students' responses to questions in our online textbook, their engagement with online course materials, and their performance on quizzes and exams.

The UCLA PI will have access to student identifiers, which will be maintained for two purposes: (1) to enable the addition of materials provided by course instructors (e.g., course grades), and (2) to remove participants if, at any point after they've completed the course, they wish for their data not to be included in research.

All data will be de-identified before researchers (other than the PI) gain access to it.

## **Indicate the purpose of the research, specifying the problems and/or hypotheses to be addressed**

Over the past several decades, statistical reasoning has emerged as a critical skill. Despite this, students often complete statistics courses unprepared to apply what they have learned. This has led statistics educators to call for change in the way we think about and teach statistics.

The purpose of the proposed research is to better understand factors that help or hinder student learning in an introductory statistics course. Results from this research will be used to (1) enhance our knowledge of how students learn statistics, and (2) improve course design for future students.

Samples of research questions include:

- How does successful completion of a previous statistics course relate to performance, attitudes, and engagement in a subsequent statistics course?
- How does time spent engaging with the materials impact student learning?
- Can we develop an "early warning system"? That is, are the early struggles that predict later difficulty in the course?
- Are there course variations that are more effective in producing student learning? For instance, does the provision of additional, embedded videos enhance student performance?

## **Describe the study design and proposed data analyses**

Data are in the form of (1) students' responses to content questions embedded in the course, and (2) pre-/mid-/post surveys.

As students read the textbook, they complete R exercises and respond to multiple choice and open response questions that assess their understanding of the content they're reading.

Course pre-/mid-/post surveys assess their perceptions of the course (level of challenge, overall learning, focus on concepts vs. memorization), value of different course features (online textbook, lectures, etc.), and self-reported study habits, attitudes and beliefs about statistics and programming, and interest in taking additional stats/programming courses.

We will use statistical analyses such as ANOVA, Chi-Square tests, and statistical modeling, as appropriate.

## **Explain in detail how or why the data/specimens are available to UCLA investigators for research and identify the source and the provider of the data/specimens:**

Through our project, we have created the content of an online "textbook" used to teach Introductory Statistics. We have also created the platform on which it is housed (i.e., CourseKata). The platform stores students' answers to questions embedded in the book, that can be used by students and their teachers to check for understanding. That data will be used by our research team to understand what students have learned. We will use that understanding to improve the contents of the book.

The CourseKata platform allows for instructors to upload additional student data. For instance, they can provide course grades as an additional measure of student learning. Once

linked to students' records, identifying information will be stripped from data files, with only the UCLA PI having access to the code key.

### **Study Design**

No direct subject contact – None of the research activities involve direct contact with study participants and include only analyses of data.

### **Locations**

- School site
- Internet

### **Do the data and/or specimens include personally identifiable information (PII) or protected health information (PHI)?**

Yes. Coded - The data/specimens are coded and a key to decipher the code exists and could be used to link PII to the data/specimens. The UCLA PI and the provider of the data/specimens enter into an agreement prohibiting the release of the key under any circumstances.

### **Indicate all identifiers that may be accessed or included in the research records for the study:**

Dates

Email address

### **Will coded or personally identifiable data be collected, transmitted or stored via the internet?**

Yes

Data will be collected from the online course and stored within CourseKata, but the data available for download by the research team will include no personal identifying information. Only the UCLA PI will be able to download identifying information. It will not be transmitted via the Internet beyond that point.

A firewall will be used to protect the research computer from unauthorized access.

Controlled access privileges will be used on the hardware storing the data.

### **During the study indicate how data will be stored and secured including paper records, electronic files, audio/video tapes, specimens. Specify how the code key will be securely maintained, as applicable.**

Encryption or password protection software will be used

Secure network server will be used to store data

### **Distribution Rules: Describe the criteria used to determine the adequacy of requests to obtain data and/or specimens (e.g., the type of researchers that will be eligible to receive data):**

The PI at UCLA will have access to de-identified data for study. He will collaborate with additional educational researchers with expertise that might be brought to bear to better understand student learning (e.g., researchers with expertise in motivation). All researchers will

sign an agreement indicating that they will notify the UCLA PI immediately if any identifying information is found in the data file. That information will be deleted immediately.

**Choose the description that is applicable to this study:**

The research does not involve greater than minimal risk (45 CFR 46.404/21 CFR 50.51)

**Indicate the potential risks/discomforts, if any, associated with each intervention or research procedure. Additionally discuss any measures that will be taken to minimize risks. If data are available, estimate (a) the probability that a given harm may occur, (b) its severity, and (c) its potential reversibility. The information provided should be reflected in risks section of the informed consent documents.**

Our study will place no demands on students beyond the work they would have otherwise completed for their statistics class. As such, the research procedures present no risks or discomforts.

The only remotely possible risk is that a researcher views a student's survey or homework answers in connection to the student's name. Our use of a key code should make that impossible, but if information is seen by a researcher, they will immediately bring it to the attention of the UCLA PI and it will be deleted.

And, to be clear, no sensitive information is collected via the course surveys.

**Are there any potential direct benefits (physical, psychological, social or other) to study participants?**

No

**Describe the potential benefits to society including the importance of the knowledge to be gained.**

An understanding of statistics is of growing importance in society, in which data plays an ever-enlarging role. We hope that by improving this online statistics and data science textbook, we will improve students' understanding.

**RISKS/BENEFIT ANALYSIS: Indicate how the *risks to the participants are reasonable in relation to anticipated benefits, if any, to participants and the importance of the knowledge that may reasonably be expected to result from the study:***

There is no more than a minimal risk for participating, but there are potentially substantial benefits to improving future students' learning.

**Indicate your plans for obtaining informed consent for this study.**

A waiver of consent is being requested. This study is essentially an evaluation of course materials (not students). All data used by researchers will be de-identified.

**Would it be appropriate to provide participants with information about the study after their participation?**

No, the data will not be stored with identifiers with which to contact the participants

**Explain why the research could not practicably be carried out without the waiver of consent.**

The study involves analyzing the learning data that students will produce as a normal part of their statistics class. Though the survey data has been added to the statistical content, the questions pose no more than minimal risk and researchers will not have access to identifying information.