Practical Aspects of Running Experiments with Human Participants

Frank E. Ritter (frank.ritter@psu.edu)
Jonathan H. Morgan (jhm5001@psu.edu)
College of Information Sciences and Technology
The Pennsylvania State University

Jong W. Kim (jong.kim@ucf.edu)
Department of Psychology
University of Central Florida

Keywords: Psychology experiments, human participants, ethics for experiments, experimental design

ABSTRACT: *There can often be a gap between theory and its implications for practice when gathering data on human behavior. This gap can be particularly significant outside psychology departments. While most students at the undergraduate or early graduate levels in psychology are taught how to design experiments and analyze data in courses in psychology and statistics, there is, unfortunately, a dearth of materials providing practical guidance for running experiments. In this tutorial we provide a summary of a practical guide for running experiments with human participants (Ritter, Kim, Morgan, & Carlson, in press).*

1. Introduction

There are many skills required for simulation. In addition to creating the simulations, you sometimes have to gather new data to help build the simulation. The knowledge about how to gather data through experimentation is partially covered in experimental design. How to analyze the data is covered in statistics courses and books.

On the other hand, the lack of materials on the details of running human experiments can lead to a gap between theory and practice, which is particularly acute outside of psychology departments where less studies are run. Consequently, labs developing simulations frequently must teach these practical skills to students informally because it is not included in their formal or informal education. Researchers in psychology often end up appalled by the lack of this common but undocumented sense when these studies are reported by researchers applying experimental psychology methods outside of psychology.

The details about how to run the studies themselves, how to interact with subjects, and the details are often learned solely through apprenticeship in a psychology or HCI lab. However, many researchers who are running or want to run studies do not have access to this tacit knowledge.

This tutorial provides a practical guide on the practical aspects of how to prepare and run experiments with human participants.

1.1 Why do we need a practical guide?

In general, scientific inquiries in the areas of human-computer interaction (HCI), human factors, cognitive psychology, and cognitive science involve human participants. One distinguishing factor of these disciplines, and thus experiments in these areas, has been the centrality of the human participant. Consequently, working in these areas requires not only understanding the theoretical and ethical issues for running human participants but also the practical aspects of the process itself. To start to frame this discussion, we are working to provide an overview of this process and related issues.

1.2 Purpose of this tutorial

In this tutorial we will present a summary of a practical guide (Ritter, Kim, & Morgan, 2009; Ritter, Kim, Morgan, & Carlson, in preparation) that can help people run experiments more effectively, more safely, and more comfortably for both the subjects and experimenters. Our purpose is to provide hands-on knowledge about experimental procedure.

We are generally speaking here from our background running cognitive psychology, cognitive ergonomics, and human-computer interaction (HCI) studies.

Because it is practical advice, we do not cover experimental design or data analyses and it may be less applicable in more distant areas.

1.3 Who is this tutorial useful for?

We believe that the tutorial will be useful to anyone who is starting to run research studies, training people to run studies, or studying the experimental process. Attendees will be provided copies of the written materials. A lecture and discussions will be used to present the material. If the conference allows it, the materials may be partly provided ahead of time for reading ahead. Particularly, the tutorial will be useful for students, teachers, lab managers, and researchers in industry. It will also be useful to computer scientists and other technologists who need to run empirical studies to gather data used to test and develop models.

2. Contents

This tutorial focuses on topics that are important for running repeatable studies. The tutorial will make the case for the importance of repeatable and valid experiments and some of the potential ethical issues that can arise. Table 1 notes several of the major components of studies explained in the larger report.

Table 1. Several important study components.

Component	Explanation
Scripting	What will be done with participants, writing it down in a script.
Missing subjects	How do you deal with subjects who do not show up?
Decorum	How do you dress and how do you address the participants?
Recruiting	How do you recruit a diverse yet representative set of participants without unwanted bias?
Literature	What literature should you read as background prep. for running a study?
Debriefing	How to debrief after a study session.
Payments	How to arrange payment for the participants, and the importance of getting this correct.
Piloting	The need to run pilot subjects to practice the method and also to find where the method (e.g., the script) needs to be modified.
Simulator studies	The role for simulated studies and how to treat model results as data.
Chances for insights	The need to keep your eyes and ears open for further insights while running studies.

In the tutorial we will address these topics, which we will briefly explain here to give a flavor of the material.

2.1 Repeatability and Validity

When running an experiment, insuring its repeatability and validity are of greatest importance, assuming the experiment is conducted ethically. Repeatability arises from running an experiment the same way for each participant. In addition, reducing unwanted variance in the participants' behavior is important as well. Ensuring this repeatability is partly the job of the research assistants (RAs), who often are not informed about these concepts and their practical application. Thus, RAs should strive to provide each participant with a consistent and comfortable but neutral testing experience.

Understanding how subjects will complete the task and working towards uniformity across all iterations of the procedure for each subject are important. The repeatability of the experiment is a necessary condition for scientific validity. There are, however, several well-known effects that can affect the experimental process. Chief among these is the experimenter's effect, or the influence of the experimenter's presence on the participants and how this effect can vary across experimenters. Depending upon the experimental context, the experimenter effect can lead to either increased or decreased performance. The magnitude and type of effect that can be attributed to this effect generally depends upon the type and extent of personal interaction between the participant and experimenter. Thus, you should strive to provide each participant with a comfortable but neutral testing experience.

Besides the experimenter effect, there are other risks to the experimental process. We highlight some here and illustrate how to avoid them, either directly or through proper randomization. Randomization is particularly important because you will most likely be responsible for implementing treatments, while understanding the other risks will help you take steps to minimize them. Finally, there are other experimental effects that are outside of your control—we do not cover these here. Even though you cannot eliminate all contingent events, you can note idiosyncrasies and with the principle investigator either correct them or report them as a potential problem.

Another common source of variation across trials is the effect of the experimental equipment. For instance, if you are having subjects interact with a computer or other fixed display, you should take modest steps to make sure that the participant's distance to the display is the same for each subject—

this does not mean, necessarily, putting up a tape measure, but in some cases, it does. It is necessary to be aware that the viewing distance can influence performance and in extreme cases can lead to blurred vision, irritated eyes, headache, and movement of torso and head (e.g., Rempel, Willms, Anshel, Jaschinski, & Sheedy, 2007). Furthermore, if subjects are picking up blocks or cards or other objects, the objects should either always be in the same positions, or they should be always randomly placed because some layouts of puzzles can make the puzzles much easier to solve. There will be other effects where variation in the apparatus can lead to unintended differences, and you should take advice locally to learn how to reduce them.

2.2 Ethics

There are several topics that you need to keep in mind when running subjects that we will cover. Chief among these are the ethics pertaining to the running of participants, and the gathering and reporting of data including published and unpublished documents. If you have any questions, you should contact the lead researcher (or principal investigator), or other resources at your university.

We would like to generalize the results to a wide population, indeed, the whole population. It is useful to recruit a representative population of subjects to accomplish this. It has been noted by some observers that experimenters do not always recruit from the whole population. In some studies, this is a justifiable approach to ensure reliability (for example, using a single sex in a hormonal study) or to protect subjects who are at greater risk because of the study (for example, non-caffeine users in a caffeine study).

Where there are not threats to validity, experimenters should take some care to include a representative population. This may mean putting up posters in a broad range of locations, and it may include paying attention to sex balance and even age balance in a study, and then correcting the balance by recruiting more subjects with these features. As a research assistant, you can be the first to notice this, and to bring it to the attention of the investigator, and help to address this.

It is necessary to avoid any procedures in a study that restrict participants' freedom of consent, that coerce their participation in a study. Some participants, including minors, patients, prisoners, and individuals who are cognitively impaired are more vulnerable to coercion. For example, enticed by the possibility of payments, minors might ask to participate in a study. If, however, they do so without parental consent, this is unethical because they are not old enough to give their consent—agreements by a minor are not legally binding.

Students are also vulnerable to exploitation. The grade economy presents difficulties, particularly for course where a lab component is integrated into the curriculum. In these cases, professors must not only offer an experiment relevant to the students' course-work but also offer alternatives to participating in the experiment.

To address these problems, it is necessary to identify potential conditions that would compromise the participants' freedom of choice. For instance, in the second example, recall that it was necessary for the professor to provide an alternative way to obtain credit. In addition, this means ensuring that no other form of social coercion has influenced the participants' choice to engage in the study. Teasing, taunts, jokes, inappropriate comments, or implicit quid pro quo arrangements are all inappropriate. These interactions can lead to hard feelings (that's why they are ethical problems!), and loss of good will towards experiments in general and you and your lab in particular.

When preparing to run the study, you should prepare how to deal with sensitive data as well. There are at least two issues here—data that you anticipate is sensitive and unanticipated data that arises that is sensitive. Data that is intrinsically sensitive should be handled carefully. Personal data is the most common. Information on an individual, such as related to race, creed, gender, gender preference, religion, friendships, and so on, must be protected. This data should not be lost or mislaid. It should not be shared with people not working on the project, either formally if you have an IRB that requires notice, or informally, if your IRB does not have this provision (this may occur more often outside of the US). You should seek advice from your colleagues about what practices are appropriate in your specific context. In some situations, you are not allowed to take data from the building, and in most cases, you are encouraged to back it up and keep the backed-up copy in another safe and secure location. In nearly all cases, anonymising data, that is, removing names and other ways data can be associated with a particular individual, removes most or all of the potential problems.

The second type of sensitive data is data that can arise where the subject's responses have implications outside of the scope of the study. This can include subjects implicating themselves in illegal activity, or unintentionally disclosing an otherwise hidden medical condition. For example, if you are administering caffeine, and you ask the subject what drugs they

take (to avoid known caffeine agonists or antagonists), you may find information about illegal drug use. If you take a subject's heart rate or blood pressure measurements, you may discover symptoms of underlying disease.

You should know what to do in these cases before they arise. Generally, preparation for a study should involve discussions about how to handle sensitive data, and if there is a chance that the study may reveal sensitive data about the participants. You should fully understand how your institutions policies regarding sensitive data, and how to work with the subjects when sensitive information becomes an issue. If you have questions, you should ask the principle investigator.

2.3 Recruiting

Recruiting participants for your experiment can be a time consuming and potentially difficult task, but it is a very important procedure to produce meaningful data. An experimenter should thus carefully plan out with the lead researcher (or the principal investigator) to conduct successful participant recruitment for the research study. Ask yourself, "What are the important characteristics that my participants need to have?" Having a coherent reason for which participants are allowed or disallowed into your study is important.

First, it is necessary to decide a population of interest from which you would recruit participants. For example, if an experimenter wants to measure the learning effect of foreign language vocabulary, it is necessary to exclude participants who have prior knowledge of that language. On the other hand, if you are studying bi-lingualism you will need to recruit people who speak two languages. In addition, it may be necessary to consider age, educational background, gender, etc., to correctly choose the target population.

Second, it is necessary to decide how many participants you will recruit. The number of participants can affect your final results. The more participants you can recruit, the more reliable your results will be. However, limited resources (e.g., time, money) often force an experimenter to use the minimum number of participants. You may need to refer to previous studies to get some ideas of the number of participants, or may need to calculate the power of the sample size for the research study, if possible (most modern statistical books have a discussion on this, and teach you how to do this, e.g., Howell, 2008). Finally, you will upon occasion have to consider how many Ss are too many. It is believed to be the case, that running large number of subjects is both wasteful of time and effort, and also that the types of statistics that are typically used become less appropriate with large sample sizes. With large sample sizes effects that are either trivial or meaningless in a theoretical sense become significant (reliable) in a statistical sense. This is not a normal problem, but if you arrange to test a large class you might get close to this problem.

There are several ways that participants can be recruited. The simplest way is to use the experimenters, themselves. In simple vision studies, this is often done because the performance differences between people in these types of tasks is negligible and knowing the hypothesis to be tested does not influence performance. Thus, the results remain generalizable even with a small number of participants.

The next way that subjects can be recruited that we will consider is a sample of convenience. Samples of convenience consist of people who are accessible to the researcher. Many studies use this approach, so much so that this is not often mentioned. Generally for these studies, only the sampling size and some salient characteristics are noted that might possibly influence the participants' performance on the task. These factors might include age, major, sex, education level, and factors related to the study, such as nicotine use in a smoking study, or number of math courses in a tutoring study. There are often restrictions on how to recruit appropriately, so stay in touch with your advisor and/or IRB.

In studies using samples of convenience, try distributing an invitation email to a group mailing list (e.g., students in the psychology department or an engineering department) done with approval of the list manager and your advisor. These posts should be reasonably likely to appeal to the list members and they should be likely candidates (Cheyne & Ritter, 2001). Also, you can post recruitment flyers in a student board, or an advertisement in a student newspaper. Use efficiently all resources and channels that are available to you.

There are disadvantages to using a sample of convenience. Perhaps the largest is that the resulting sample is less likely to lead to generalizable results. The subjects you recruit are less likely to represent a sample from a larger population. Students who are subjects are different from students who are not subjects. To name just one feature, they are more likely to take a psychology class and end up in a subject pool. And, the sample itself might have hidden variability in it. The subjects you recruit from one method (an email to them) or from another method (poster) may be different. We also know that they differ over time—those that come early to fulfill

a course requirement are more conscientious than those that come late. So, for sure, randomly assign subjects to the conditions in your study.

The largest and most carefully organized sampling group is a random sample. In this case, researchers randomly sample a given population by carefully applying sampling methodologies meant to ensure statistical validity and equal likelihood of selecting each potential subject. Asking students questions at a football game as they go in does not constitute a random sample—some students do not go (selection bias). Other methods such as selecting every 10th student based on a telephone number or ID introduce their own biases. For example, some students do not have a publicly available phone number, and some subpopulations register early to get their ID numbers. Truly choosing a random sample is difficult, and you should discuss how best to do this with your lead researcher.

One approach for recruiting participants is a subject pool. Subject pools are generally groups of undergraduates who are interested in learning about psychology through participation. Most Psychology departments organize and sponsor subject pools.

Subject pools offer a potential source of participants. You should discuss this as an option with your lead researcher, and where appropriate, learn how to fill out the requisite forms. If the students in the study are participating for credit, you need to be particularly careful with recording who participated because the students' participation and the proof of that participation represent part of their grade.

The theory is that participating in a study provides additional knowledge about how studies are run, and provides the participant with additional knowledge about a particular study. The researchers, in turn, receive access to a pool of potential subjects.

2.4 The Related Literature

This short document does not assume that you have a background in statistics or have studied experimental design. To help run a study you often do not need to know these areas (but they do help!). If you need help in these areas, there are other materials that will prepare you to design experiments and analyze experimental data. In addition, most graduate programs with concentrations in HCI, cognitive science, or human factors engineering feature coursework that will help you become proficient in these topics.

Many introductory courses in statistics, however, focus primarily on introducing the basics of regression and ANOVA. These tools are unsuitable for many studies analyzing human data where the data is qualitative or sequential. Care, therefore, must be taken to design an experiment that collects the proper kinds of data. If ANOVA and regression are the only tools at your disposal, we recommend that you find a course focusing on the design of experiments featuring human participants, as well as the analysis of human data. We also recommend that you gather data that can be used in a regression because it can be used to make stronger predictions, not just that a factor influences a measure, but in what direction (!) and by how much.

So, it is generally useful to have read in the area in which you are running experiments. This reading will provide you with further context for your work, including discussions about methods, types of subjects, and pitfalls you may encounter. For example, the authors of one of our favorite studies, an analysis of animal movements, note that data collection had to be suspended after having been chased by elephants! If there are elephants in your domain, it is useful to know about them. There are, of course, less dramatic problems such as common mistakes subjects make, correlations in stimuli, self-selection biases in a subject population, power outages, printing problems, or fewer participants than expected. While there are reasons to be blind to the hypothesis being tested by the experiment (that is, you do not know what treatment or group the subject is in that you are interacting with, so that you do not implicitly or inadvertently coach the subjects to perform in the expected way), if there are elephants, good experimenters know about them, and prepared research assistants particularly want to know about them!

As a result, the reading list for any particular experiment is both important and varies. You should talk to other experimenters, as well as the lead researcher about what you should read as preparation for running or helping run a study.

2.5 Piloting

Conducting a pilot study based on the script of the research study is important. Piloting can help you determine whether your experimental design will successfully produce answers to your inquiries. If any revision to the study is necessary, it is far better to find it and correct it before running multiple subjects, particularly when access to subjects is limited. It is, therefore, helpful to think of designing experiments as an iterative process characterized by a cycle of design, testing, and redesign. In addition, you are likely to find that this process works in parallel with

other experiments, and may be informed by them (e.g., lessons learned from ongoing related lab work).

Thus, we highly recommend that you use pilot studies to test your written protocols (e.g., instructions for experimenters). The pilot phase provides experimenters with the opportunity to test the written protocols with practice participants, and is important for ironing out misunderstandings, discovering problematic features of the testing equipment, and identifying other conditions that might influence the participants. Revisions are a normal part of the process; you should not hesitate to revise your protocols. This will save time later. There is also an art to knowing when not to change the protocol. Your principle investigator can help judge this!

It is also useful at this stage to write the method section of your paper. Not only is your memory much fresher but also you can show other researchers your method section and receive suggestions from them before you run the study, a good time to get suggestions. These suggestions can save you a lot of time, in that these reviews essentially constitute another way of piloting the study.

2.6 Chance for insights

Gathering data directly can be tedious, but it can also be very useful and inspiring. Gathering data gives you a chance to obtain insights about aspects of behavior that are not usually recorded, such as the user's questions, their postures, and their emotional responses to the task.

Obtaining these kinds of insights and the intuition that follows from these experiences is important for everyone, but gathering data is particularly important for young scientists. It gives them a chance to see how previous data has been collected and how studies work. Reading will not provide you with this background or the insights associated with it; rather this knowledge only comes from observing the similarities and differences that arise across multiple subjects in an experiment.

So, be engaged as you run your study and then perform the analysis. These experiences can be a source for later ideas, even if you are doing what appears to be a mundane task. In addition, being vigilant can reduce the number and severity of problems that you and the lead investigator will encounter. Often, these problems may be due to changes in the instrument, or changes due to external events. For example, current events may change word frequencies for a study on reading. Currently, words such as bank, stocks, and mortgagees are very common, whereas these words were less prevalent a few years ago.

3. Conclusions

Once a science is mature enough practitioners will know the methods, while a science is growing, the method will have to be more explicitly taught. While a method is moving between areas, such as behavioral studies moving from psychology to computer science and engineering, the method will need to be made more explicit, and it can be useful for a method to become more explicit.

In our tutorial we will provide practical advice regarding the important and basic inquiry of how to run an experiment with human participants. We are working on extending and polishing a written guide that will be useful to anyone who is starting to run research studies, training people to run studies, or studying the experimental process (Ritter, Kim, Morgan, & Carlson, in press, 2012). This tutorial and guide will particularly help researchers in industry and government in addition to students who are not in large departments or who are running participants in departments that do not have a large or long history of experimental studies of human behavior.

Currently, the report is in use at eight universities in the US, Canada, and England for graduate and advanced undergraduate courses in cognitive science, human factors, information science, and in human-computer interaction courses.

As a colleague noted, this contains just common sense. We have found that this common sense is not so common, and that new researchers, both students and those taking up a new methodology, need a good dose of common sense, and that researchers who have many fields to master as in simulation and training will need and can use help in learning the tacit knowledge in psychology about how to gather human data.

4. Acknowledgements

Developing the report of practical advice on running studies was sponsored by ONR (W911QY-07-01-0004, N00014-10-1-0401, and N00014-11-1-0275).

5. References

Cheyne, T., & Ritter, F. E. (2001). Targeting respondents on the Internet successfully and responsibly. *Communications of the ACM, 44*(4), 94-98.

Rempel, D., Willms, K., Anshel, J., Jaschinski, W., &

Sheedy, J. (2007). The effects of visual display distance on eye accommodation, head posture, and vision and neck symptoms. *Human Factors, 49*(5), 830-838.

Ritter, F. E., Kim, J. W., Morgan, J. H., & Carlson, R. A. (in press, 2012). *How to run experiments: A practical guide to research with human participants*. Currently 112 pages. Thousand Oaks, CA: Sage.

Author Biographies

FRANK RITTER is on the faculty of the College of IST, an interdisciplinary academic unit at Penn State to study how people process information using technology. He edits the *Oxford Series on Cognitive Models and Architectures* and is an editorial board member of *Human Factors, Cognitive Systems Research,* and *IEEE SMC IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans.*

JONATHAN H. MORGAN is a researcher at the College of Information Sciences and Technology's Applied Cognitive Science Lab with expertise in modeling social interactions and social moderators, including the moderating effect that relative spatial positions have on group performance.

JONG W. KIM is a researcher at the University of Central Florida. His research interests are in learning and forgetting of skill. He received a PhD in Industrial Engineering in 2008 from Penn State.