How to Run Experiments:  
A Practical Guide to Research with Human Participants

Frank E. Ritter, Jong W. Kim, Jonathan H. Morgan, and Richard A. Carlson
frank.ritter@psu.edu  jongkim@psu.edu  jhm5001@psu.edu
College of Information Sciences and Technology
The Pennsylvania State University
University Park, PA 16802

2 July 2011, interim draft

approximately: 24,600 words, 1 tables, 2 figures

Target size: 50-60k words, 2 Tables, 5-25 Figures, 0-10 photos

Copyright 2011, Ritter, Kim, and Morgan.
Abstract and Preamble

There are few practical guides on how to prepare and run experiments with human participants in a laboratory setting. In our experience, we have found that students are taught how to design experiments, and how to analyze data in courses such as Design of Experiments and Statistics. On the other hand, the dearth of materials available to students preparing and running experiments has led to a gap between theory and practice in this area, which is particularly acute outside of psychology departments. Consequently, labs frequently must not only impart these practical skills to students informally but must also address misunderstandings arising from this divorce of theory and practice in their formal education.

We present here advice that can help young experimenters and research assistants to run experiments effectively and more comfortably with human participants. In this book, our purpose is to provide hands-on knowledge and actual procedures of experiments. We hope this book will help students of psychology, engineering, and the sciences to run studies with human participants in a laboratory setting. This will particularly help students who are not in large departments, or are running participants in departments that do not have a large or long history of experimental studies of human behavior.

We are generally speaking here from our background running cognitive psychology, cognitive ergonomics, and human-computer interaction studies. Because it is practical advice, we do not cover experimental design or data analyses. This practical advice will be less applicable in more distant areas, or when working in more complex situations, but may be still of use. For example, we do not cover how to use complex machinery, such as fMRI or ERP. We also do not cover field studies or studies that in the US require a full IRB review. This means that we do not cover how to work with unusual populations, such as prisoners, animals, and children, or how to take and use measures that include risks to the subjects or to the experimenter (e.g., saliva, blood samples, or private information).

This book arose during a discussion at Jong Kim’s PhD graduation. Ritter asked what he thought he had not learned as well or what we had not taught him and how to teach the next students better, and the conversation turned to experimental methods and the tactics and details of running studies. During the graduation ceremony they outlined this book. A worthy genesis for a book we think.

We have addressed this book toward advanced undergraduates and early graduate students starting to run experiments without previous experience; but we believe this guide will be useful to anyone who is starting to run research studies, training people to run studies, or studying the experimental process.

When running an experiment, insuring its repeatability is of greatest importance—it is critical to address variations in either method or in participant behavior. Running an experiment in exactly the same way regardless of who is conducting it or where (e.g., different research teams or laboratories) is essential. In addition, reducing unanticipated variance in the participants’ behavior is key to an experiment’s repeatability. This book will help you achieve these
requirements, increasing both your comfort and that of the participants who participate in your experiments. We hope you find it relevant and useful.

This book consists of several sections with multiple appendices. We concisely describe below each section’s contents.

**Section 1, Overview of the Research Process**, describes briefly where experiments fit into the research process. If you have taken either an experimental methods course or a research design course, you can skip this chapter. If, on the other hand, you are either a new research assistant or are working on a project in which you are unclear of your role or how to proceed, this chapter may provide some helpful context.

**Section 2, Preparation for Running Experiments**, describes pertinent topics for preparing to run your experiment—such as supplemental reading materials, recruitment of participants, choosing experimental measures, and getting Institutional Review Board (IRB) approval for experiments involving participants.

**Section 3, Potential Ethical Problems**, describes ethical considerations necessary for safely running experiments with human participants—i.e., how to ethically recruit participants, how to handle data gathered from participants, how to use that data, and how to report that data. Being vigilant and aware of these topics is an important component to rigorous, as well as ethical, research.

**Section 4, Risks to Validity to Avoid While Running an Experiment**, describes risks that can invalidate your experimental data. If you fail to avoid these type of risks, you may obtain either false or uninterruptible results from your experiment. Thus, before starting your study, you should be aware of these risks and how to avoid them.

**Section 5, Running a Research Study**, describes practical information about what you have to do when you run experiments. This section will give an example procedure that you can follow.

**Section 6, Concluding a Research Session and Study**, describes practical information about what to do at the conclusion of each experimental session and at the end of a study.

**Section 7, Example Research Studies**, describes example experimental studies that give you a brief synopsis of procedural steps. It provides details of example experiments. [this is to be cut up and moved into the chapters]

**Section 8, Conclusion**, summarizes the book and describes several examples, forms, and checklists included as appendices. These forms will vary by lab and IRB committee, but provide examples of the style and tone.

**Acknowledgements**

Preparation of this manuscript was partially sponsored by a grant from the Division of Human Performance Training, and Education at the Office of Naval Research, under Contract # W911QY-07-01-0004 and N00014-11-1-0275. The views and conclusions contained in this report are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Government or the Pennsylvania State University.
How to run experiments: A practical guide

Ellen Bass, Gordon Baxter, Stephen Broomell, Richard Carlson, Karen Feigh (several times), Alex Kirlik, Kuo-Chuan (Martin) Yeh, Razvan Orendovici, and Erika Poole, have provided useful comments, and Joesph Sanford has helped prepare this mss, but any incompleteness and inadequacies remain the fault of the authors.
How to run experiments: A practical guide

Table of Contents

RUNNING BEHAVIORAL EXPERIMENTS WITH HUMAN PARTICIPANTS: A PRACTICAL GUIDE (FOR STUDENTS OF HCI, IE, AND COGNITIVE PSYCHOLOGY) .............................................1
1 OVERVIEW OF THE RESEARCH PROCESS ..................................................................7
  1.1 OVERVIEW .............................................................................................................7
  1.2 DEFINITION OF TERMS .......................................................................................11
  1.3 FURTHER READINGS ..........................................................................................13
2 PREPARATION FOR RUNNING EXPERIMENTS ..........................................................15
  2.1 LITERATURE IN THE AREA .................................................................................15
  2.2 CHOICE OF A TERM: PARTICIPANTS OR SUBJECTS .........................................16
  2.3 RECRUITING PARTICIPANTS ..............................................................................17
  2.4 SUBJECT POOLS ..................................................................................................18
  2.5 CARE, CONTROL, USE, AND MAINTENANCE OF APPARATUS ......................19
  2.6 TESTING FACILITY ..............................................................................................21
  2.7 CHOICE OF MEASURES: PERFORMANCE, TIME, ACTIONS, ERRORS, VERBAL PROTOCOL ANALYSIS, AND OTHER MEASURES ............................................21
    2.7.1 Types of measures .........................................................................................21
    2.7.2 Levels of measurement ..................................................................................22
    2.7.3 Scales of measurement ..................................................................................23
  2.8 ERROR DATA .........................................................................................................24
  2.9 RUN ANALYSIS WITH PILOT DATA ....................................................................24
  2.10 INSTITUTIONAL REVIEW BOARD (IRB) .............................................................25
  2.11 FURTHER READINGS ..........................................................................................26
3 POTENTIAL ETHICAL PROBLEMS .............................................................................26
  3.1 RECRUITMENT OF A BROAD SELECTION OF SUBJECTS .....................................27
  3.2 COERCION OF PARTICIPANTS ............................................................................27
  3.3 SENSITIVE DATA ..................................................................................................28
  3.4 PLAGIARISM ..........................................................................................................28
  3.5 FRAUD ...................................................................................................................29
  3.6 SUMMARY .............................................................................................................29
  3.7 FURTHER READINGS ............................................................................................29
4 RISKS TO VALIDITY TO AVOID WHILE RUNNING AN EXPERIMENT .......................30
  4.1 VALIDITY DEFINED: SURFACE, INTERNAL, AND EXTERNAL .........................30
  4.2 RISKS TO VALIDITY .............................................................................................32
    4.2.1 Power: How many participants? .................................................................32
    4.2.2 Experimenter effects .....................................................................................33
    4.2.3 Participant effects .........................................................................................34
    4.2.4 Randomization ..............................................................................................34
  4.3 EXAMPLE PROBLEMS ..........................................................................................35
  4.4 FURTHER READINGS ............................................................................................35
5 RUNNING A RESEARCH STUDY .................................................................................37
  5.1 SCRIPT ..................................................................................................................37
  5.2 PILOTING ...............................................................................................................37
  5.3 DRESS CODE FOR EXPERIMENTERS ...............................................................38
  5.4 WELCOME ............................................................................................................38
  5.5 MISSING SUBJECTS ............................................................................................38
  5.6 DECORUM .............................................................................................................38
  5.7 TALKING WITH SUBJECTS ...................................................................................38
How to run experiments: A practical guide

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.8 Debriefing</td>
<td>39</td>
</tr>
<tr>
<td>5.9 Payments and Wrap-up</td>
<td>40</td>
</tr>
<tr>
<td>5.10 Simulator Studies</td>
<td>40</td>
</tr>
<tr>
<td>5.11 Problems and How to Deal with Them</td>
<td>40</td>
</tr>
<tr>
<td>5.12 Example Problems</td>
<td>41</td>
</tr>
<tr>
<td>6 Concluding a Research Session and a Study</td>
<td>42</td>
</tr>
<tr>
<td>6.1 Data Care, Security, and Privacy</td>
<td>42</td>
</tr>
<tr>
<td>6.2 Data Backup</td>
<td>42</td>
</tr>
<tr>
<td>6.3 Chance for Insights</td>
<td>42</td>
</tr>
<tr>
<td>7 Example Research Studies</td>
<td>44</td>
</tr>
<tr>
<td>7.1 Skill Retention Study</td>
<td>44</td>
</tr>
<tr>
<td>8 Afterword</td>
<td>47</td>
</tr>
</tbody>
</table>

INDEX

APPENDIX A: Example Script to Run an Experiment                         | 49   |
APPENDIX B: Example Consent Form                                       | 52   |
APPENDIX C: Example Debrief Form                                       | 54   |
APPENDIX D: Example IRB Application                                     | 55   |
REFERENCES                                                               | 64   |

1.1 Bits to add
How to run experiments: A practical guide

1  Overview of the Research Process

Broadly speaking, scientists and engineers investigate scientific questions, and experiments constitute one powerful form of investigation. For many studies, human participation is necessary to adequately explore the question. For instance, evaluating the usability of a haptic interface (e.g., a Wii remote) before its introduction to the market would be an example. The question, then, is what considerations should inform the investigator when conducting these kinds of research studies?

In general, scientific inquiries in the areas of human-computer interaction (HCI), human factors, cognitive psychology, and cognitive science require the involvement of human participants. One distinguishing factor of these disciplines, and thus experiments in these areas, has been the centrality of the human participant. Unlike inert materials, humans learn, and bring with them far more aspects of their history.

Consequently, working in these areas requires not only understanding the theoretical and ethical issues incumbent to running human participants but also the practical aspects of the process itself. To start to frame this discussion, we provide an overview of this process, and issues related to it.

The rest of this chapter describes briefly where experiments fit into the research process. If you have had an experimental methods course or a research design course, you can skip this chapter. If you are a new research assistant, or are working on a project in which you are unclear about how you fit in, this chapter may provide some helpful context. We also define some common terms so that you can better communicate with the principal investigator and other members of your research team.

Questions answered by this chapter [put on all chapters]

1.1  Overview

Here is a list of important tasks that form our overview of the research process. They are presented as if they are performed in order. In practice, these tasks are most often performed in order in classes and the first time you are working in an area. After you have worked in an area for a while the steps become less ordered.

Or the more common and worse scenario, you are an engineer who has been asked to make a widget, you realize somewhere along the way that someone will need to use said widget and that you might need to verify that it works, or to make a decision between widget interfaces. You now need to run an experiment to understand how novices will use it. Where do you start, what do you do? Before you do anything else you need to decide why you are running the experiment in the first place. You need to make a list of high level questions to answer and then a series of actual testable questions for each. These questions can then be translated into what data you need to collect, which will in turn specify what types of statistical analysis can and should be performed (if any) and what you will ultimately be able to say.
It is important to prepare your questions and plan before actually collecting data, before choosing your measures, and before recruiting participants, because each of these activities will be better informed if you have a clear plan of how to go from high-level questions through precise technical questions to data collection and analysis. In some cases you cannot answer these questions very clearly, but answering them as clearly as you can will improve the results because you will make the implicit assumptions more explicit.

### 1.2 Overview of the Research Process

Figure 1 provides an overview of the research process. We use it to describe the process in more detail.

![Figure 1. An overview of the research process.](image)

1. Establish a hypothesis

   The most common hypothesis is that one factor in the world influences another factor that can be measured. This is the most common hypothesis tested. To test this hypothesis, the first factor, such as time exposed to a stimuli has to be varied, and the second factor, such as how long something is remembered, must be measured.
Sometimes, a hypothesis can simply be that the area is interesting, and that gathering data will provide insights. This is occasionally referred to as a fishing expedition, because you do not know what you will catch. This type of hypothesis is sometimes criticized for being too general, but for exploratory work, it can be very successful. For example, we suspected that subjects when confronted with a problem would choose problem strategies based upon certain features, but we did not know which features factored into the participants’ choices. So, we included multiple types of problems and multiple types of features. Then, with analysis, we were able to pull out which features participants based their decisions upon across a wide range of problem types (Reder & Ritter, 1992).

(2) Set up the experiment

To investigate what you want to know—to test your established hypothesis, it is necessary to draw a picture in your mind about what is needed to achieve your goal.

For example, suppose that you want to know how students can retain foreign vocabulary words in terms of the spacing of learning (i.e., massed or distributed practice). To set up this experiment, an experimenter would need materials for displaying the vocabulary words and for recording responses from participants (e.g., time or accuracy). So, in this step, you will prepare stimuli or a simulation, create an experiment in EPrime or another tool, and create an apparatus to gather data to test your hypothesis.

(3) Pilot study to understand the theory’s edges

When your experimental apparatus is setup, you should verify that it works, in that the stimuli are correctly presented and that the data you think will be gathered is gathered in the format you expect. This stage helps you to test to see if and how your experimental apparatus works and to identify what is working, what is not working, and what is missing for your successful investigation. Probably, a theory would explain a part of the phenomena. The theory’s edge indicates a feasible region where the phenomena are scientifically explained and understood. Understanding the theory’s edge is important because you can identify a concrete direction of your research based on the understanding of the theory’s edge.

For example, you might ask friends and colleagues to try a mirror tracing task. You might run people casually, in their offices, and record their times to learn how their response times differ by stimuli. You would not report these results, but use them to adjust your apparatus (perhaps adjusting the mirror) and your stimuli (you might find how long it takes to follow each pattern), and the procedure. (You might have to remind them not to look directly at the sheet of paper)

(4) Submit IRB forms

Around this time, and perhaps earlier and perhaps later, as your study protocol becomes clear enough you should submit your study for IRB approval. We cover later how to do this and when you do not have to do this.
Recruiting subjects

Sometimes, recruiting subjects is easy, and sometimes recruiting is hard. It depends on your local circumstances. If you have access to a subject pool, it is easier. If, on the other hand, your study requires particular subjects with particular expertise (such as airplane pilots), recruiting is harder.

Recruiting subjects can start while piloting and setting up the study if the study is easy to prepare or if subjects are hard to recruit. If subjects are easy to recruit and the study harder to prepare, it probably should be done in the opposite order.

Running subjects

With your study piloted and with IRB approval, you can start to run the study, that is, running subjects through the study and retaining the results for later analysis and publication.

Running subjects may give you different outcomes than those of your pilot study. The primary cause for these differences is generally due to individual variability—participants may think or react in unanticipated ways. Or, you may get different results because your study is more formal. In either case or even the case where there are fewer surprises, you are interested in seeing the truth about the world based on examining a sample of it. How to run the study is the focus of this book.

Analyze the results

If you have been careful, you have analyzed your pilot data to make sure that the output from the study can be analyzed. Sometimes, timestamps are in the wrong format to be read by an analysis program, or are not recorded, or the subject’s name has remained attached to a file (rather than their ID). Checking that the data can be put into the analysis software is time well spent before you run participants in a larger trial.

Doing the analysis and running the study can provide insights that will often lead to further ideas for studies.

Adjust the experiment and run a modified study

It may be necessary to adjust the experiment to address problems identified in a previous stage, or as a second study. This process might be repeated, however, and through this process of adjustment the experiment will stabilize and results will accumulate.

Run a modified study

After the iteration of adjusting a series of your pilot studies, you would reach an experimental design that gives you much more stable and interpretable outcomes. It often seems that the process takes more time than you think, but this process is necessary to produce interpretable and repeatable results.
(10) Write report(s)

In this step you take the results and prepare a manuscript, perhaps in the form of a technical report for a sponsor, as a conference paper, journal article, or maybe as a thesis. You will usually have help and guidance throughout this step, and there are useful books on this area and we do not address this topic further in this book. This step is worth keeping in mind during the previous steps because this is what you are working towards.

Note that these steps are normative; they are what should happen. In practice, this process often runs in parallel, can vary in order (insights do not always come before or between experiments), and is iterative. Furthermore, breakthroughs frequently result from interactions between multiple experiments and researchers in a lab, so it is usually not a solitary activity.

1.3 Definition of Terms

We start by providing some definitions of terms that are frequently used in the process of creating and performing research studies.

(1) Independent variable vs. dependent variable

Briefly, independent variables are variables varied (such as training schedule) and dependent variables are the measurements you take (e.g., response time). For more information, please refer to Section 2.7.1.

(2) Informed consent

Informed consent is a process to (a) provide specific information about the research study and its procedures to the participants (or subjects), (b) answer questions to ensure the participants understand the research, (c) provide the participants with adequate time to consider their decisions, and (d) obtain the voluntary agreement from the participants to take part in the research study.

(3) IRB

IRB stands for an Institutional Review Board, which is a review committee to help protect the rights and welfare of human participants in a research study.

Studies in US institutions for publication require IRB approval according to Federal law. Good review boards can do this quickly; great review boards can do this quickly and help you with your research, pointing out how to be more responsible to subjects and how to improve your study (gathering data that is interpretable is also a responsibility). They will also require your cooperation, and you should look forward to working with them.

An IRB can (a) approve/disapprove a research study, (b) modify a research study, (c) conduct continuing reviews, (d) observe/verify changes to a study, (e) suspend or terminate approval, and (f) observe the consent process and the research procedures. Many or most of these steps are not your responsibility, but you may be asked to help perform them.
(4) Lead researcher

A lead researcher generally indicates a person who is responsible for designing and running experiments. An experimenter actually administers experimental procedures to participants and collects data. A research assistant assists in this process.

(5) Null hypothesis vs. alternative hypothesis

The null hypothesis proposes that there is nothing (null) going on in the experimental investigation—factor 1 has no influence on factor 2, and any correlation you see is just noise, a random occurrence. That is, the inferential statistical significance testing leads you to make a decision that there is an influence of factor 1 on factor 2, and the experiment is designed to have only face 1 vary between conditions.

The hypothesis is competing with the null hypothesis and something that researchers want to prove to be true through the research process.

Here are some examples. One of the major findings by Ebbinghaus (1885/1913) is the principle of distributed practice. This principle holds it is desirable to spread out practice rather than massing it together in a session. That is, a 10-minute practice per day for 5 days can provide a better learning and retention rate than a 50-minute practice in one day. Based on this principle, let us construct a research hypothesis. In this case, the null hypothesis would be that there is no statistically significant difference in the learning and retention rates between distributed and massed practice groups. The hypothesis would be that a statistically significant difference does exist between the two groups, and that the distributed practice group displays higher learning and retention than the massed practice group.

There may be alternative hypotheses as well. If the study is not well designed or well run, it could be that something besides the hypothesis is causing the changes observed. For example, if the first 20 subjects to show up are put into one group, and the last 20 to show up are put into another group, then the differences between groups might not be caused by how they are treated, but by whatever caused them to be in the first or last group, such as conscientiousness.

(6) Principal Investigator

A principal investigator is an official point of contact and is responsible for a grant contract. A principal investigator also often takes the role of the lead researcher, responsible for conducting experiments involving human participants and gathering data. The terms for who is responsible for how the study is conducted, who is responsible to a sponsor, and who is in charge of you will vary across labs.

(7) Reaction time (RT)

Reaction time is an elapsed time between the presentation of a stimulus and the consequent response to that stimulus from the subject.

(8) Significance testing
Significance testing refers to the process used to determine whether an effect (or effects) observed in the experiment is a real effect, rather than just the result of the random variation. In this significance test, you would either reject the null hypothesis or fail to reject the null hypothesis. If you reject the null hypothesis you provide support for but fail to prove the hypothesis. The better your experiment technique and the larger the effect the more support you will obtain.

Stimuli

Stimuli are events that evoke or cause a reaction. That is, in this context, stimuli are events evoking reactions from a subject.

Figure 2 for moving: team size and relationships.

1.4 Further Readings

A course in experimental methods is probably the best way to learn about how to design, run, and analyze studies. In addition, we can provide a list of suggested reading materials that provide you with further knowledge about experimental design and methods. We list them in an alphabetical order of the first author.


  This is a relatively large book. It covers a wide range of methods, some in more depth than others. It includes some instructions for how to perform the methods.

How to run experiments: A practical guide


This is a book for the first course in experimental methods in psychology. It is a useful and gentle introduction to how to create and run studies and how to present the results. It does not focus on the practical details like this book does.
2 Preparation for Running Experiments

Let us again consider a usability study evaluating a haptic (touch-based input or output) interface. For this investigation, a lead research scientist or a lead researcher would establish a study hypothesis and design an experiment by first defining what to measure (dependent variables), what factors to manipulate (independent variables), and what environmental conditions to consider.

Within the lab that it is run, multiple experiments are going on at the same time. Joining the lab as a new research assistant, you have come to help out and to learn in this area, specifically with running research studies. What do you do? Where do you start? How do you avoid common and easily fixed problems? This chapter describes how to get started.

2.1 Literature in the Area

This book 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 feature coursework that will help you become proficient in these topics.

Many introductory courses in statistics, however, focus primarily on introducing the basics of ANOVA and regression. These tools are unsuitable for many studies analyzing human subject 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.

Returning to the topic of readings, it is generally useful to have read in the area in which you are running experiments. This reading will provide you 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, notes 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 very individualized. 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.

2.2 Choice of a Term: Participants or Subjects

Disciplines vary as to which term they prefer: subject or participant and how their role is not completely passive. Participant is the newer term, and was adopted by many research communities to emphasize the researcher’s ethical obligations to those participating in their experiment. Even more descriptive terms such as learner, student, or user can be used and are generally preferred. Nevertheless, subject is still commonly used, and appears in older research. For students in many psychology programs, the term, participants, is preferred to that of subjects. The Publication Manual of the American Psychological Association (APA), 5th ed. (American Psychological Association, 2001’, p. 70) suggests replacing the impersonal term, subjects, with the more descriptive term, participants. The APA goes on to define participants as individuals: college students, children, or respondents. The APA manual suggests this, but does not require it.

Page 73 of the 6th Edition of the APA Pub Manual says this about the use of the term "subjects"

Write about the people in your study in a way that acknowledges their participation but is also consistent with the traditions of the field in which you are working. Thus, although descriptive terms such as college students, children, or respondents provide precise information about the individuals taking part in a research project, the more general terms participants and subjects are also in common usage. Indeed, for more than 100 years the term subjects has been used within experimental psychology as a general starting point for describing a sample, and its use is appropriate.

Whether following the APA guideline or not, we should recognize that S, Ss, S’s, E, Es, E’s indicate Subject, Subjects, Subject’s, Experimenter, Experimenters, and Experimenter’s in earlier research—Fitts’s 1954 study is one example where these abbreviations are used. Furthermore even within the discipline of psychology, opinion can be split. Roediger (2004) argues against the change to participants suggested in the latest version of the APA’s Publication Manual. He argues that subjects is both more consistent and clearer, noting that the term has been in use since the 1800’s and that it better defines the relationships involved. He argues that the term participants fails to adequately capture the distinction between the experimenter and those in the study—strictly speaking experimenters are participants as well. We use these terms interchangeably in this document because we recognize other research communities may still prefer subjects, and because not all psychologists, and certainly not everyone running behavioral experiments, are members of the APA.

Another distinction¹ to draw in this area is what the purpose of the study is. If the subject of interest to you is a psychological phenomenon, an aspect of human behavior, they may appear more as subjects in the traditional use of the term. On the other hand, it may be that you are actually interested in how someone performs when given a certain interface or new tool and task. In this case you are actually interested in how well the widget works and your subjects are really

¹ We thank Karen Feigh for this suggested view.
How to run experiments: A practical guide

more like participants who are participating with you in your work, helping you to generalize results and to improve the product. In any case, take advice about what to call the people you work with.

2.3 Recruiting Participants

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, thus, should carefully plan out with the lead researcher (or the principal investigator) to conduct successful recruitment for the study. Ask yourself, “What are the important characteristics that my participants need to have?” Your choices will be under scrutiny, so 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 the ability to generalize from your final results. The more participants you can recruit, the more reliable your results will be. However, limited resources (e.g., time, money, etc.) force an experimenter to find the appropriate and reasonable number of participants. You may need to refer to previous studies to get some ideas of the number of participants, or you 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 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 useful 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 everyone in 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 is 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. 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 these types of 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.

2.4 Subject Pools

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 required 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.

A whole book could be written about subject pools. Subject pools are arrangements that psychology or other departments provide to assist researchers and students. The department sets up a way for experimenters to recruit subjects for studies. Students taking particular classes are either provided credit towards the class requirement or extra credit. When students do not wish to participate in a study, alternative approaches for obtaining course credit are provided.
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.

2.5 Care, Control, Use, and Maintenance of Apparatus

What materials do you need to run experiments? The experiments in a controlled environment (e.g., a laboratory) usually require participants to interact with a computer device, a prototype, or a mock-up. For example, it is possible to implement a task environment in a computer screen—such as an air traffic control task like Argus (Schoelles & Gray, 2001), a driving simulator like Distract-R (Salvucci, in press), experimental tasks with E-Prime (e.g., MacWhinney, St. James, Schunn, Li, & Schneider, 2001), or a spreadsheet task environment (Kim, Koubek, & Ritter, 2007).

Part of what you will have to do to set up and run a study is to understand the task environment so that you can prepare it for each session, save the data if it collects data, and shut it down after each session.

As you begin to work on your research task, you are likely to consider several approaches for improving your study. Finding, developing, or modifying the task environment to support your study is often an early consideration. The task environment provides the setting for investigating the questions of interest, and having the right task environment is a key element to a successful study. If designing and implementing a new task environment for your research study seems infeasible, try reusable and sharable environments. This is increasingly possible.

After choosing and setting up the task environment, the next step is to determine what method you will use to record the participant’s performance. Data collection deserves serious thought. Data can be qualitative (i.e., not in a numerical form) or quantitative (i.e., in a numerical form). Different hypothesis and theories require different types of data to test them, and thus methods to collect data. For example, you can use a camcorder in an interview to gather qualitative information or a keystroke logger like RUI (Kukreja, Stevenson, & Ritter, 2006) to measure numerical values of quantitative data in unobtrusive and automatic ways. We suggest avoiding manually recording data—it is hard, takes a significant amount of time, and is prone to error. Though, sometimes, manual data collection is unavoidable and for pilot studies it is quite often appropriate. Often with a little forethought ways can be found to automate the process.

An apparatus is often required to gather behavioral data. In cognitive science, recording user behavior by using experimental software, a video recorder, a voice recorder, or a keystroke/mouse logger, etc are all common practices. There are also tools for generating studies such as ePrime. Also, some studies require using an eye-tracker to gather eye-movement data.

Experimental software

Many studies are performed with custom built, or bespoke software. The research team conducting the study usually develops these custom applications. They can vary from a simple program to present stimuli and record reaction times to more complex programs (interactive simulations for instance). As a new research assistant, you will be instructed on how to start up
and run the software necessary for your work. On the other hand, as you run subjects with such programs, try moving from a passive to an active user. Make suggestions that you think might improve the program’s usability as they arise, note mistakes in the program, and observe how subjects interact with the program in novel or interesting ways. These insights can lead to further studies and to further hypotheses to test.

**E-Prime**

E-Prime\(^2\) was the first commercial tool designed to generate psychological experiments on a personal computer (MacWhinney, St. James, Schunn, Li, & Schneider, 2001). E-Prime is compatible with Microsoft Windows® XP/Vista. PsyScope\(^3\) is another experiment generation program, and a predecessor of E-Prime. You can download it free under a GNU General Public License\(^4\). PsyScope runs on the MacIntosh. You may be asked to use these tools in your current study or may find them to be great value in producing study stimuli more quickly.

**Keystroke loggers**

It is often useful to record the user’s behavior while they perform the task, not just the total task time. This can be done in several ways. Some researchers have used video recordings. This provides a very stable result that can include multiple details. It also can provide a rich context, particularly if both the subject and their surroundings are recorded. On the other hand, analyzing video recordings is time consuming and can be error prone. Analyzing video data requires examining the video frame-by-frame to find when the user performs each action, and then recording each action by hand into your dataset.

Another approach is to record just the keystrokes or mouse clicks. There are commercial versions available from companies like Noldus that will record keystrokes. We have also designed a keystroke logger, RUI (Recording User Input). RUI is a keystroke and mouse action logger for the Windows and Mac OS X platforms (Kukreja, Stevenson, & Ritter, 2006). It is a very useful tool for recording user behavior in human-computer interaction studies. RUI can be used to measure response times of participants interacting with a computer interface over time.

Using RUI, however, does raise issues regarding privacy in public clusters (e.g., a classroom). University policies almost universally prohibit installing any tool for experimentation that obtains or could obtain a user’s information on identity such as a login ID or a password (Kim & Ritter, 2007). Fortunately, Kim and Ritter (2007) describe one possible portable solution to this problem. They used a simple shell script to automatically run RUI on an external drive, a jump drive. When RUI is operated from an external drive it provides a way to efficiently use RUI on public cluster machines and then remove it when the study is over.

**Eye-trackers**

An eye tracker is a device to measure eye positions and movements. It can offer useful data of cognitive processes when a user interacts with an interface (e.g., a computer screen, a physical


\(^3\) [http://psy.ck.sissa.it](http://psy.ck.sissa.it)

product, etc). This device is sensitive, requiring special care to guarantee the measurement’s quality.

2.6 The Testing Facility

A testing facility can be called a psychological testing room, human factors lab, an ergonomics lab, a usability lab, or a HCI lab. Rosson and Carroll (2002) describe a usability lab as a specially constructed observation room. In this observation room, an investigator can simulate a task environment and record the behavior of users. Thus, the room should be insulated from outside influences, particularly noise. However, it is sometimes necessary to observe and record behaviors of a group of users interacting with each other. In these cases, it may be hard to capture this data in a lab setting. Ideally, the testing facility should be flexible enough to conduct various types of research.

Jacob Nielson (1994) edited a special issue of the journal about usability laboratories. This special issue provides several representative usability laboratories in computer, telecommunications, and consumer product companies (e.g., IBM, Symantec, SAP, Phillips, or Microsoft, etc.). While this special issue is somewhat dated, the underlying concerns and some of the technological details remain accurate. In addition, many of the social processes and uses for video have only become more important.

If you are designing your own study, you should try to arrange access to a room that allows participants to focus on the experimental task. Lead researchers will often have such rooms, or can arrange access to them.

2.7 Choice of Measures: Performance, Time, Actions, Errors, Verbal Protocol Analysis, and Other Measures

2.7.1 Types of measures

There are several types of measures. Questionnaires are one common and flexible type. By answering the questions, participants self-report about the question, thus providing researchers insights into their behavior. The quality and type of these responses, however, depend upon the quality and type of the questions asked—so carefully selected and carefully worded questions are important.

One example where questionnaires can be used effectively is studying self-judgment and its effects. Under certain conditions, our feelings about our knowledge and our actual knowledge may differ. In this case, our hypothetical researcher asks participants to make a judgment about what they know after memorizing vocabulary words. Using a Likert scale is a common way to measure self-judgment. Likert scales typically consist of five to seven points with ratings ranging from “Strongly disagree” to “Strongly agree”. Our hypothetical researcher would then test the participants and compares the participants’ responses about their knowledge with the results.

Other types of measures can include physiological measures. Cozby (2004) introduces a few popular physiological measures such as galvanic skin response (GSR), electromyogram (EMG), and electroencephalogram (EEG) that help us understand psychological variables. Also, fMRI
How to run experiments: A practical guide

(functional magnetic resonance image) is a popular method of measuring and examining brain activities. If you are interested in learning more about these techniques, refer to the section of Further Readings, specifically *Psychophysiological recording* (Stern, Ray, & Quigley, 2001).

2.7.2 Levels of measurement

Often within a single study, multiple measures with different characteristics are gathered. Let us discuss some common measures taken in an HCI or cognitive science experiment. For instance, you can measure the task completion time; or you can measure the number and the times of the keystrokes and mouse actions performed by the participants during the task. You can also measure what errors were made during the task, and so on.

It is necessary to decide what you are observing and measuring from the participants who are performing the experimental task. The decision is important because the choice of measures is directly related to what aspects of the participants’ behavior is being captured by the task. In general there are two types of variables: (a) independent variables, and (b) dependent variables.

Independent variables cause, or manipulate the changes in the participants’ behavior that the researchers seek to observe during the study. Thus, independent variables are sometimes called manipulated variables, treatment variables, or factors (Keppel & Wickens, 2004).

To cement our understanding of variables, let us presume that we want to measure how humans forget something they have learned. We will return to this example later, but for now, we will focus the study’s independent and dependent variables. Variables that can manipulate forgetting performance include training types, retention intervals (how long a participant will retain learned information), and input modalities (what types of skills a participant is to learn). Thus, we would consider these variables the study’s independent variables. They are deliberately varied to create the effects—they are independent.

Dependent variables indicate what we will observe. Their values are (presumed to be) dependent on the situation set up by the independent variables. Dependent variables can either be directly observed or may be derived. Response time and error rates are two typical dependant variables. The measures can also be more complex. Workload measure for example, allows researchers to measure how hard users have to work. The NASA TLX directly measures each of the six individual subscales, but sometimes a desired measure is used based on combining them. We can observe the time that is required to complete a task if the investigation is to understand human performance caused by forgetting. Also, we can observe errors produced by participants to measure forgetting. These variables are considered to be dependent variables. There can be one or more dependent variables. One dependent variable in an experiment uses univariate statistical methods, and more than two dependent variables are referred to multivariate methods.

To sum up, dependent variables are the responses being observed during the study while independent variables are those factors that researchers manipulate to either cause or change those responses.
2.7.3 Scales of measurement

Variables can be basically measured using four types of scales (Ray, 2003): (a) nominal measurements, (b) ordinal measurements, (c) interval measurements, and (d) ratio measurements. Knowing these scales of measurement is important because the data interpretation techniques available to you for interpreting the results are a function of the scales of measurement used, and the use of such data, perhaps even how it is stored and equipment is calibrated can depend on what kind of data it is.

Nominal (also referred to as categorical) measurements are used to classify or name variables. There is no numeric measure of values representing names or separate categories. For example, participants can be classified into two groups—a male group and a female group, to measure performance on using a GPS navigation system. In this case, the gender difference is an independent, categorical variable to compare performance. Or, if the numbers 1 to 10 are treated as words, such as how often they are said, then there is not necessarily even an order to them, they could be sorted alphabetically.

Ordinal measurements, in contrast, represent some degree of quantitative difference (or relative amount). For example, football rankings in the Big Ten conference are an ordinal measurement, they are in order, as are ratings on a scale of 1 to 10. Differences between the first and second team, between 9th and 10th, and between ratings of 4 and 5 and 6 and 7 are not necessarily equal, just ordered.

Interval measurements rely upon a scale values based on a single underlying quantitative dimension. The distance, therefore, between the consecutive scale values are meaningful. For example, the interval between 6 and 12 is the same as the interval between 12 and 18.

Ratio measurements determine values with respect to an absolute zero—there is no length shorter than 0 inches for instance. The most common ratio measurement can be found in a count measures (i.e., the number of hits or misses). For example, in a shooting game, the number of hits is used to determine the shooter’s accuracy.

In addition to these numeric measures we would like to note two other types of data. Frequently, sets of sequential data, or protocols, are gathered from human participants for a given task. Protocols may be multiple streams of data including verbal utterances, motor actions, environmental responses, or eye movements (Newell & Simon, 1972). As an example of a verbal protocol, consult the testing methodology developed by Ritter and Larkin (1994) for the principled analysis of user behavior. Sequential data cannot be reduced to a single number, but can be used in interesting ways (Sanderson & Fisher, 1994).

Verbal data often provides insights into understanding human behavior. Ericsson and Simon (1993b) published a summary of how and when to use verbal reports as data to observe humans’ internal cognitive processes. The basic assumption of their verbal protocol theory is that verbalization of a human’s memory contents (not their view of their thought processes) can be used to derive the sequence of thoughts to complete a task. Thus, verbalization can be a valid form of data representation that offers us unique insights into cognition (see chunking). This type of data requires audio recordings, and often comes with special apparatus for recoding and special
software and tools for analyzing the result. It is time consuming, but can be very helpful for understanding how the task is performed.

2.8  Error Data

Another type of data to gather is error data. Error data consists of trials or examples where subjects did not perform the experimental task or some aspects of the task correctly. This type of data can provide useful examples of where cognition breaks down. In addition, it helps describe the limits of performance and cognition.

Error data is generally more expensive to collect because in most cases participants perform the task correctly. Thus, more trials have to be run to gather a hundred errors than it takes to gather a hundred correct responses. If errors are not interesting theoretically for your research study, some pilot running of the experiments may be required to generate an experiment where errors do not occur too often.

2.9  Run Analyses with Pilot Data

We can highly recommend that you run pilot subjects, gather data from them, and analyze the data before launching a large experimental study. The number to run can be found with experience, or by talking with your PI. Analysis of pilot data can provide a baseline, or identify problems with the testing techniques or measures used. Your pilot subjects can be your friends, family, or subjects taken from your subject pool.

If the results from the pilot data are not what you expected, you can revise the design of the experiment (e.g., change which independent variables are recorded, change the target task, or add another treatments, etc.). If the results from the pilot data match your expectations, you can plan to launch your more formal experiments to gather data because an interesting new study topic may have emerged.

Keep in mind that with a small number of subjects you might only be able to see large effect sizes. A large effect size means that the difference of your treatment is large with respect to how much people generally vary. For example, freshman will vary in weight, as will seniors, for example, with a standard deviation of 30 pounds. If the seniors weigh more, like 30 pounds, the effect of going from freshman year to senior year is about the amount the population varies. This is thus an effect size of 30/30 or 1. If the these student vary in number of gray hairs on their heads by 10, and the seniors on average have 1 more gray hair, it will take a lot more students measured to show that number of gray hairs varies than it takes to show that weight varies between these two groups. If you are not finding an effect with a pilot study, you might just need to run more subjects or revise your expected effect size.
2.10 Institutional Review Board (IRB)\textsuperscript{5}

Investigators in psychology or human factors in many countries now must obtain approval from the appropriate host institution or organization prior to conducting research. The organization charged with approving research applications in a university setting in the United States is called the Institutional Review Board (IRB), which is specific to a university or government lab. The IRB is a committee monitoring, approving, and reviewing biomedical and behavioral research involving humans.

Before the onset of the experiment, investigators must obtain the informed and voluntary consent of the participants selected for the study. The American Psychological Association’s Ethical Principles of Psychologists and Code of Conduct\textsuperscript{6} specifies that participants have the right to informed consent—participants have the right to understand what will happen in the study (e.g., any known risks of harm, possible benefits, and other details of the experiment). Only after receiving such a briefing, can a participant agree to participate in the experiment. Thus, the details of the experiment should be written in clear, jargon-free language, and without reference to special technical terms. The participants must be able to easily understand the informed consent form. In addition, the form should enable prospective participants to determine for themselves whether they are willing to participate given his or her situation and personal tolerance for risk. We provide an example of an informed consent form in the Appendix.

IRB policies are subject to interpretation so when in doubt contact the IRB representative at your institution. They are more coaches then policemen.

In general IRB reviews fall under two categories, expedited or full review. Most behavioral science studies that do not involve the use of experimental drugs, radiation, or medical procedures can be considered for expedited review. Expedited review does not require full IRB approval, that is, the full IRB board to discuss your study, and can usually be accomplished within a few weeks (again this will vary by institution and other factors such as time of year). For all other cases, you will need to go through a full review—these are usually scheduled far in advance at specified dates, and this document does not attempt to cover such studies.

2.11 When don’t you need to go through an IRB

There are a few exceptions that are worth noting, where IRB approval is not required. If you are running yourself and only yourself, you do not need IRB approval. If you are running studies only for class work, or for programmatic improvement and not for publication, then IRB is not required. These exceptions are useful when you are piloting studies, or when you are teaching (or learning), or when you are developing software. Of course, you can in most cases still seek IRB approval advice or situations in these. The approval process offers you the opportunity for feedback on how to make your study more safe and efficient. Approval also allows later publication if the results are interesting.

\textsuperscript{5} This applies to research in the US. You should enquire locally because some countries do not see risk in routine cognitive experimental projects, or perform reviews in a more local or in a way adjusted more to the type of study.

\textsuperscript{6} \url{http://www.apa.org/ethics/code2002.html}
IRB approval is required before any aspect of the study is performed that will be published is conducted, including subject recruitment. Without exception, IRB approval cannot be granted once the study has been conducted. Consequently, you should seek IRB approval early in the process and keep your timeline and participant recruitment flexible. You do not need to seek new approval for enrolling fewer participants than requested or finishing early. You will, however, need to seek approval for finishing late or for enrolling a larger number of participants or substantially changing the study.

### 2.12 Further Readings

We list some reading materials that may help you plan and run experiments, and report results from the experiment.


  This book chapter provides you with background for using EEG and its processes, including physiological basis and frequency analysis of the EEG. In addition, Ray and Slobounov explain EEG research on motor processes in general and brain trauma.


  This is a specially edited article concerning usability laboratories. This special issue provides several representative usability laboratories—mostly computer, telecommunications, and consumer product companies (e.g., IBM, Symantec, SAP, Phillips, or Microsoft, etc.).


  This book provides comprehensive background in the area of human-computer interaction.


  *Psychophysiological Recording* is a very useful book for anyone who conducts experiments with human participants to measure their alternative hypothesis physiological responses. The book provides not only practical information regarding recording techniques but also the scientific contexts of the techniques.


  *Rival hypotheses* provides a set of one page mysteries about how data can be interpreted, and what are alternative hypotheses about how a study produced a result. Following the mystery is an explanation about what are other very plausible rival hypotheses that could lead to the result occurring. Reading it is engaging and teaches more critical thought about experimental results.
3 Potential Ethical Problems

There are several topics that you need to keep in mind when running subjects. 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.

3.1 Recruitment of a Broad Selection of Subjects

We would like to generalize the results that we find 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 outside your department, 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 the 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.

3.2 Coercion of Participants

Coercion is an ethical violation of the rights of human participants. It is necessary to avoid any procedures in a study that restrict participants’ freedom of consent regarding 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 classes where a lab component is integrated into the curriculum. In these cases, professors must not only offer an experiment relevant to the students’ coursework 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 example class with a lab component, 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.
3.3 Sensitive Data

When preparing to run a study, you should prepare how to deal with sensitive data. There are at least two issues here—data that you anticipate is sensitive and unexpected 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.

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 subject’s heart rate or blood pressure measurements, you may discover symptoms of underlying disease.

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 institution’s 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.

3.4 Plagiarism

Plagiarism refers to taking other’s work or ideas and using them as one’s own, that is, without attribution. Particularly in academia, this problem is taken seriously.

An individual might be tempted to steal others’ ideas, research methods, or results from unpublished or published works. Nowadays, manuscripts that are about to be submitted or already submitted for review, can be available online.

Why people are tempted to plagiarize others’ work? Generally, pressure to meet or surpass institutional standards causes people to plagiarize. To pass a programming class, students might copy another student’s code. A faculty member, facing review for tenure and stressed by the number of his or her refereed publications, or an RA trying to fill in a methods section all might be tempted to steal the work of others. Sometimes, the pressure to publish, is enough to tempt an academic to plagiarize other’s ideas and fabricate their data.

The integrity and development of scientific knowledge is rooted in the proper attribution of credit. In the APA’s publication manual (p. 349), you can find the APA’s guidelines for giving
credit. Direct quotes require quotation marks and citations while paraphrasing or in anyway borrowing from the work of others requires a citation. You may also need to acknowledge people who give you unpublished ideas for your research designs. In particular, you may have personal communications (e.g., email, messages from discussion groups on the net, letters, memos, etc.) that require acknowledgement. In this case, you will need to remember who gave you the idea (an email thanking them can be a good way to document this), and then cite them in the text with a date.

3.5 Fraud

We, sometimes, are shocked by news about research fraud. For example, if a researcher fabricates data and publishes a paper with the data, this is fraud. Other scientists trying to replicate the results are often the ones who find and reveal the initial findings to be fraudulent. While research fraud is unusual, we, nevertheless, must be aware that fraud can cause significant adverse effects not only for the perpetrator of the fraud but also often second or third parties such as his or her academic colleagues, institution, funding agency, or corresponding journal editor. Or, more distant people who base an educational system on a learning theory, or teaching strategies on incorrect data on memory.

If data is lost, it is lost, do not replace it. If you accidentally delete data, do not replace it. If you did not run a subject, do not run yourself. All of these practices undermine your study’s validity and are extremely egregious ethical violations. It is sad when you read in an article that “data from 3 subjects were lost”, but it is far better to write this phrase than to commit fraud.

3.6 Summary

This chapter notes a few of the most important ethical problems you might face. You may encounter others. If you have questions, you should contact the lead investigator or other senior personnel. In some cases, as in many ethical situations, there may not be a right answer, there may be several right answers, and often there are better answers and good, accepted practices.

3.7 Further Readings

Here is a list of further readings for you concerning this chapter.

- The APA’s webpage, *Ethical Principles of Psychologists and Code of Conduct*. This was published first in 1992, but has been superceded by a newer release issued in June 2003. Here is the link for the current code of conduct: http://www.apa.org/ethics/code2002.html


The APA publication manual provides useful guidance for reporting your experimental findings in a written paper.
4 Risks to Validity to Avoid While Running an Experiment

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 better 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 the same 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 to validity 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 (for example, the effect of historic world events on your study). 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 affect more blurred vision, irritated eyes, headache, and movement of torso and head (e.g., Rempel, Willms, Anshel, Jaschinski, & Sheedy, 2007). The factors of which can, thus, be risks to validity. 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. The experimental set up should not be sometimes one and sometimes the other.

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.

4.1 Validity Defined: Surface, Internal, and External

We refer to validity as the degree to which an experiment leads to an intended conclusion from the data. In general, two types of validity, internal validity and external validity, are of interest. Internal validity refers to how well experimental treatments explain the outcomes from the experiment. The experimental treatments indicate independent variables that you design. External validity, in contrast, refers to how well the outcomes from the experiment explain the phenomena outside the designed experiment. This is known as “generalizability”. 
Campbell and Stanley (1963) discusses 12 factors that endanger the internal and external validity of experiments. We need to consider how to reduce or eliminate the effects from these factors to guarantee valid results.

When you run studies you may notice factors that can influence the ability of the study results to be explicated (this is referred to as “internal validity”). Because you are running the subjects, you have a particular and in some ways not repeatable chance to see these factors. Good principle investigators will appreciate you bringing these problems to their attention. You should not panic—some of these are inevitable in some study formats, but if they are unanticipated or large then they may be interesting or the study may need to be modified to avoid them.

- History: Besides the experimental variable, a specific event could occur between the first and second measurements. Typically, this is some news item such as a space launch or a disaster that influences subjects in a global way leading to better or worse results than would occur at other times.
- Maturation: Participants can grow older, become more knowledgeable, or become more tired with the passage of the time. Thus, if you measure students at the beginning of the school year and then months later, they may get better scores based on having taken classes.
- Testing: The effects of taking a test on the scores of a second test. Thus, if you take an IQ test, the same test, a second time, you are likely to score better, particularly if you got feedback from the first taking.
- Instrumentation: It is required to calibrate a measuring instrument regularly. Some instruments need to be recalibrated with changes in humidity. Failure to recalibrate can affect an experiment’s results.
- Statistical regression: We need to avoid selecting groups on the basis of their extreme scores. If you select subjects based on a high score, some of those high scores will most likely not reflect the participant’s normal performance, but a relatively high score. On retests their performance will decrease not because of the manipulation but because the 2nd measure is less likely to be extreme again.
- Selection Biases: Differential selection of participants for the comparison groups should be avoided. Subjects that come early in the semester to get paid or credit are different from the subjects who put it off until the last week of the semester.
- Experimental mortality: There could be a differential loss of participants from the comparison groups. Some conditions could be harder on the subjects, and thus lead them to come back less.

As you run subjects you may also see factors that influence the ability to generalize the results of the study to other situations. The ability of results to generalize to other situations is referred to as external validity.

- The reactive or interaction effect of testing: A pretest could affect (increase or decrease) the participants’ sensitivity or responsiveness to the experimental variable. Some pre-tests disclose what the study is designed to study. If the pre-test asks about time spent studying math and playing math games, you can bet that mathematical reasoning is being studied in the experiment.
• The interaction effects of selection biases and the experimental variable: It is necessary to acknowledge that independent variables can interact with subjects that were selected from a population. In this case, the outcome or findings from the experiment may not be generalized to summarize a larger or different population.
• Reactive effects of experimental arrangements: An experimental situation itself can affect the outcome that cannot be generalized. That is, the outcome can be a reaction from the specific experimental situation.
• Multiple-treatment interference: If multiple-treatments should be applied to the same participant, the participant’s performance would then not be valid because of the accumulated effects from those multiple treatments. For example, if you have learned sample material one way, it is hard to tell if later learning is the result of the new learning method presented second, or the result of the first method, or the combination of the two.

Why mention these effects in a book on how to run subjects? Why not just let these be mentioned in experimental design? We mention them here because if you are new RA, you may not have had an experimental design class. And yet, many of these effects will mostly be visible to the person running the study. If there is an event in a country where you are running subjects like an election, and you will be comparing results to a different country where the PI is located, it is the RA that has the best chance of noticing that something unusual that is a threat to validity has happened in the study.

4.2 Risks to Validity

There are other issues that investigators need to consider, such as participants’ effects or experimenters’ effects. We will take these issues up in the following section.

4.2.1 Power: How many participants?

Performance is noisy. Differences that appear could be due to a theoretical manipulation, or it could be due to chance. We now discuss the power of a statistical test, and how a test’s power can influence its effectiveness. Calculating the test’s power can help maximize the benefits of an experiment by helping you decide how many subjects to run. For instance, while relatively rare, running too many subjects can be wasteful when the effect size is known to be large.

Testing a hypothesis produces two outcomes: (a) one outcome can be rejecting the null hypothesis ( ), while the other outcome (b) can be not rejecting the null hypothesis—that is accepting the alternative hypothesis ( ). When investigators decide to either accept or reject the alternative hypothesis, they can make two types of errors, known as Type I and Type II errors. Table 4.1 describes these errors.
Table 4.1. Type I and II error in testing the null ($H_0$) and experimental ($H_a$) hypotheses.

<table>
<thead>
<tr>
<th>Decision Made</th>
<th>True State</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$H_0$ is true</td>
</tr>
<tr>
<td>Reject $H_0$</td>
<td>Type I error (report a result, but no effect)</td>
</tr>
<tr>
<td>Fail to reject $H_0$</td>
<td>Correct decision</td>
</tr>
</tbody>
</table>

In fact, if the null hypothesis ($H_o$) is true, investigators should fail to reject the null hypothesis. When the null hypothesis is incorrectly rejected the null hypothesis, Type I errors occur. The probability of making a Type I error is denoted by alpha, written $\alpha$. On the other hand, if the alternative hypothesis ($H_a$) is true, in fact, investigators should accept the alternative hypothesis. When the alternative hypothesis is incorrectly rejected, Type II errors occur. The probability of making a Type II error is denoted by beta, written $\beta$. Experimenters will talk about Type I and Type II errors, so it's worth learning what they are.

The power of a test is defined as the probability of correctly rejecting the null hypothesis ($H_0$) when it is in fact true—this is denoted by $1-\beta$. In a practical sense, via the calculation of the power, investigators are able to make a statically supported argument that there is a significant difference when such a difference truly exists. Good sources of determining the size of a study a priori include: Cohen's books and explanations in stats books, and programs that can be found online for free, such as G*Power3.

4.2.2 Experimenter effects

When two or more experimenters are running the same experiment, effects or biases from experimenters can exist. One experimenter may unconsciously be more encouraging, or another be distracting in some way. Preventing possible experimenter effects is necessary for guaranteeing the validity of the experiment. Both the ability to repeat it and to generalize from it.

Mitchell and Jolley (2007) note some reasonable causes for error that investigators should avoid: (a) the loose-protocol effect, (b) the failure-to-follow-protocol effect, and (c) the researcher-expectancy effect.

First, to avoid the loose-protocol effect, when you run the experiment and particularly when a study is run by different experimenters, it is necessary to write down the procedures in detail. The protocol document should allow other experimenters to run the experiment in exactly the same way, providing a standardized way to run the trials. Once you finished a draft of the protocol document, you should test it with practice participants. An example is included as
Appendix <#>. Producing the final protocol document will require a few iterations of writing and then testing the protocols with practice participants.

The second cause of error, the failure-to-follow-protocol effect, results from an experimenter’s failure to follow the experiment’s protocols. There might be several reasons for not following the protocol—the reasons can include a lack of motivation to follow the protocol, or ignorance of the protocol, etc.

The third cause for error, the researcher-expectancy effect, arises from the influence of the experimenter’s expectations upon his or her interactions with the participants. For instance, I might be biased (consciously or unconsciously) in how I run the experiment if I know I am testing my hypothesis. After all, I have a personal incentive to reject the null hypothesis in this case. Therefore, it is preferable when possible that the experimenters interacting with the subjects be unaware of the hypothesis being tested. When this happens, it is called a double-blind study (American Psychological Association, 2001) where the experimenter and the subject both do not know what treatment the subject receives. An example would be when the RA does not know which amount of caffeine a subject received, or what condition the subject is in.

Following consistent written protocols in an unrushed manner is one way to avoid many of these errors. Please be patient and give the participants enough time to complete each procedure to the best of their ability.

4.2.3 Participant effects
Because personal characteristics and histories influence performance, it is important to try to methodically achieve a representative sample when selecting participants. Factors such as ethnicity, gender, age, experience, native language, or working memory capacity can all affect performance. Random assignment of subjects to conditions generally helps mitigate this effect. Random assignment, however, can go wrong (or be done incorrectly), or result in a suboptimal distribution. RAs often are the earliest, best, and often the only way to discover these problems.

4.2.4 Randomization
Randomization describes the process of randomly determining both the allocation of the experimental material and the order in which individual trials are to be performed (Montgomery, 2001). Random sampling, a related term, is a method for selecting the entire sample group. Ray (2003) states that one way to achieve external validity is to have the participants in the experiment constitute a representative sample of the entire population. In fact, it is very hard to accomplish true random sampling. After randomly selecting your participants from the population, you should randomly assign them to their experimental groups.

Statistical methods require that the observations should be independently distributed random variables. Proper randomization of the experiment makes the assumption that the independent distribution of observed data is valid and allows us to statistically analyze the behavioral data. Randomization also can be useful to alleviate bias of selecting participants.

In some situations, Montgomery (2001) states that it is difficult to randomize the experiment because of a hard-to-change variable (e.g., temperature in a chemical process, subject’s gender).
4.2.5 Not doing the experiment

During the experiment, subjects may wander away from the task. Few will stop using a computer-based task if you are in the room (so, if this problem occurs, stay in the room!). Some will pay less attention, and way to avoid this is to have shorter experiments, and, run a crisp experiment where you help them focus and indicate through your expectations that they will try.

In experiments using verbal protocols, the subjects may stop talking or talk about other topics. You should not let them sit without talking, and not let them talk about non-task related things. In the first case, you need to ask them to “keep talking” (Ericsson & Simon, 1993a, Appendix). In the second case, if they wander off from reporting their working memory onto other topics, you may have to ask them to focus on the task. Asking them to do these things is highly appropriate, and if you do not you will hurt the experiment. You might be more comfortable if you practice this with both a helpful (compliant) and unhelpful (uncompliant) friend as pilot subjects.

4.3 Example Problems

We can note a few problems that arose in previous studies with lessons they provide.

One study found that the subjects behaved unexpectedly, in that they had less problems than was expected. Upon further investigation, it turned out that the student research assistants were breaking up the lessons into subparts to facilitate learning. This is one effect that should be avoided to increase validity of research studies by eliminating lesson effects through instruction (e.g., VanLehn, 2007).

4.4 Further Readings

Here is a list of further reading materials concerning this chapter.


Cohen originated the statistical measure of power.


Howell’s book provides a useful summary of how to apply power written for those learning statistics. Other introductory statistics books will have similar treatments. They are useful introductions to this process.

How to run experiments: A practical guide

Ericsson and Simon in this book explain the theory of how verbal protocols can be used, and in what ways they are valid, and when they are invalid.
5 Running a Research Study

This chapter provides practical information on what to do when you run your experiments. We assume that you have developed your initial experimental design and are now ready to run a pilot study.

5.1 Script

Your research study will likely have a script of how to run the session. If it does not, it should, and it will help you run each subject in a confident and consistent manner. The script will often start with how to setup the apparatus. Before the subject’s arrival, the experimenter needs to setup the apparatus and should be ready to welcome the subject. Incorrect or inconsistently applied procedures of the apparatus setup can sometimes cause inconsistency of the study-running processes (e.g., omission of a step). Consequently, the script that appropriately represents required procedures plays an important role in conducting a successful experimental study. Appendix D provides an example script for a study.

The setup should include making sure that all materials that are used are available (e.g., forms, at least one back up copy), and that the apparatus is working. If batteries are used in any of the apparatus (e.g., a laser pointer, a VCR remote), spare batteries should be to hand.

5.2 Piloting

As mentioned earlier, 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 scientifically plausible 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 the opportunity to test the written protocols with practice participants, and are 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; please do 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. These suggestions can save you a lot of time, in that these reviews essentially constitute another way of piloting the study.
5.3 **Dress Code for Experimenters**

You should consider the impression you wish to make and will make when running your experiment. This consideration should include how your position, the type of experiment, and the type of participants in the experiment.

In most cases, we recommend wearing a semi-professional outfit, such as: a dress shirt with dress slacks, when running experiments. This helps you look professional and prepared but not intimidating. Semi-professional dress helps convey the experiment’s importance while not overwhelming the participant. Encouraging your subjects to take the experiment seriously should lead to more distinctive but still generalizable effects.

5.4 **Welcome**

As the experimenter you are taking on a role similar to that of a host, thus, it is appropriate to welcome participants to the study. Where it is appropriate, you might provide them materials to read if they have to wait, and to answer questions they have before the study begins. It is also very appropriate to confirm their names (for class credit), and to confirm for them that they are in the right place and at the right time. If the experimental protocol permits it, you might also indicate how long the study will take. This helps set the stage for the study itself.

5.5 **Missing Subjects**

In every study, there are two key parties—the experimenter and the subject or subjects. Inevitably, you will encounter a situation where a participant does not show up despite having an appointment. While participants should notify you in advance if they are going to be absent, keep in mind that missed appointments do happen, and plan around this eventuality. Participants are volunteers (even when you consider compensation). Therefore, it is appropriate to be gracious about their absence. Where possible, we recommend offering to reschedule once. However, when there are repeated absences, it is often not worth rescheduling.

In some cases, you as an experimenter may need to cancel an experiment. As an experimenter, it is not acceptable to simply not show up for an experiment. When you really have to cancel the experiment for any reason, you should do it in advance. Furthermore as the experimenter, you have the responsibility to cancel the experiment by directly contacting the participants.

5.6 **Decorum**

Be culturally sensitive and respectful to the participants. Consult with the lead investigator if you have general questions concerning lab etiquette, or specific questions related to the study.

5.7 **Talking with Subjects**

When you first welcome the subjects to your study and the study area, you might feel uncomfortable. After you have run a few sessions, this discomfort will go away. In a simple study, you can be quite natural, as there is nothing to ‘give-away’. In more complex studies, you will be busy setting up the apparatus, and this tends to make things easier.
In nearly all cases, abstaining from extraneous comment on the study is an important and useful practice that makes all parties concerned more comfortable. Many experimental protocols require not giving the subject feedback during the study. In these cases, your notes will probably indicate that you tell the participants at the beginning of the session that you are not allowed to provide them feedback on their performance. Generally, the debriefing can handle most questions, but if you are not sure how to answer a question, either find and ask the investigator, or, take contact details from the subject and tell them you will get them an answer. And then, do it! This also means that when you are running subjects for the first couple of times that someone who can answer your questions should be available.

In social psychology studies or where deception is involved, you will be briefed by the investigator and will practice beforehand. In this area, practice and taking advice from the lead researcher is particularly important.

### 5.8 Debriefing

The APA’s ethical principles offer a general outline of debriefing procedures. For many experiments, the lead researcher may provide additional guidance. Investigators should ensure that participants acquire appropriate information about the experiment - such as the nature, results, and conclusions of the research. If participants are misinformed on any of these points, investigators must take time to correct these misunderstandings. Also, if any procedures are found to harm a participant, the research team must take reasonable steps to report and to alleviate that harm.

The experiment’s procedures may cause participants to feel uncomfortable or be alarmed. After the experiment is finished, investigators or experimenters should listen to the participants’ concerns and try to address these problems. Mitchell and Jolley (2007) provide reasonable steps to follow when you need to debrief:

- Correct any misconceptions that participants may have.
- Give a summary of the study without using technical terms and jargon.
- Provide participants an opportunity to ask any questions that they might have.
- Express thankfulness to the participant.

When you have a study that can be perceived as being deceptive or when the study is a double-blind study, you should seek advice about how to debrief the participants. If deception is a procedural component, you will most likely have to explain this to the subjects, and ask that they not discuss the study until the study’s completion date. Requesting the participants to refrain from discussing the study will help keep potential subjects from being biased.

To review an example, double-blind studies prescribe that neither the subject nor the experimenter knows which treatment the subject has received. For example, the amount of caffeine any single participant has ingested in a caffeine study with multiple possible doses. In these cases, you will have to explain the procedures of the study, as well as provide a general rational for double-blind trials. Otherwise, participants may balk at being given a treatment in a sealed envelope, or by a person who is not the experimenter. Furthermore, events such as the
Tuskegee and Holmes Prison experiments underscore why procedural transparency is so essential.

5.9 Payments and Wrap-up

At the end of the session, you should be sure to compensate the subject as specified. Compensation can include monetary payment, credit towards a class, or nothing. If you are paying them monetarily, check with your supervisor, as there are nearly always detailed instructions for how to process such payments. In any case, you should make sure that they receive their compensation; you receive any required documentation such as receipts; and that you thank each participant for their assistance. Without them, after all, you cannot run the study.

At the end of the wrap-up, you should set the table for the next subject. Make sure that copies of forms are to hand, and that if you have used the spare batteries you get some fresh batteries.

5.10 Simulator Studies

You may find yourself running simulated subjects. User models and simulations are increasingly used, both as standalone objects, but sometimes as part of a study to provide a social context. For example, to model a social situation you might have two intelligent agents act as confederates in a resource allocation game (Nerb, Spada, & Ernst, 1997). These agents provide a known social context in that their behavior is known and can be repeated, either exactly or according to a proscribed set of knowledge.

When you run simulations as subjects, you should keep good notes. There are often differences between the various versions of any simulation, and this should be noted. Simulations will also produce logs, and these logs should be stored as securely and as accurately as subject logs. There may be more of them, so annotating them is very prudent.

If you create simulations, you should keep a copy of the simulation with the logs as a repeatable record of the results. You should keep enough runs that your predictions are stable (Ritter, Schoelles, Quigley, & Klein, in press), and then not modify those files of model and runs but only modify copies of them.

5.11 Problems and How to Deal with Them

For cognitive psychology and HCI studies, nearly all studies run smoothly and subjects are not exposed to any more risks than those that occur in everyday life. However, when you run an experiment, you can encounter unexpected situations in which a participant is exposed to some risk of harm. Investigators must be committed to resolving these problems ethically; recognizing that the well-being of the participants supercedes the value of the study. We recommend consulting your host organization (i.e., Office for Research Protection) in the event that you encounter problems that hinder conducting experiments by affecting either the experimenters or participants. Where these events are adverse enough, you are required to report these events to the IRB board.
5.12 Example Problems

We can note a few problems that we have encountered and some lessons learned.

In one study, we could not run a few subjects because they could not find the room in which we were conducting the experiments. A locked hallway entry door and no escort prevented the participants from finding the room. Ostensibly, the pilot study would have identified this problem; however, only intra-departmental personnel participated in the pilot study. Consequently, the need for an escort was not identified until the first experimental runs with paid subjects. This example highlights the importance of knowing your participant and their needs, as well as the limitations of internal pilot studies.

In another study, a colleague lost valuable data because of a hard drive failure. Spending years gathering data on children, he had not backed up his data. When his hard drive crashed, he was given the choice to spend an extra year rerunning subjects (if support was available), or go to industry. He went to industry where sadly he was very happy!
6 Concluding a Research Session and a Study

6.1 Data Care, Security, and Privacy

This section explains practical information about what you should do when you get done with your experiment.

6.1 Debriefing [Rich]

6.2 Data Care, Security, and Privacy

All information and data gathered from an experiment should be considered confidential. If others who are not associated with the experiment have access to either data or personal information, the participants' privacy could be violated. Thus, it is the responsibility of lead researchers and experimenters to ensure that all security assurance procedures are promulgated and enforced.

Researchers must safeguard against the inappropriate sharing of sensitive information. Personal information about the participants must not be shared with people not associated with the study. Thus, the data should not be left untended. In most studies, experimental data are kept in locked files or on secure computers. The level of security may vary with the type of data. Anonymous reaction time data, where the only identifying information is a subject ID, is low or no risk. Personal health records where the subjects might be identified are much more sensitive, and would require more cautious storage, perhaps being used only on a removable disk.

6.3 Data Backup

To protect against data loss, back up all of your data routinely (after running a subject, and every 10 days at minimum when you are doing analyses of the data). If your data is stored in electronic files, store them in a secure hard drive or burn them onto a CD. If you are using paper documents, they can be scanned and stored on a computer file as back up. We suggest that you back up your data after each subject rather than weekly while conducting a study.

6.4 Chance for Insights

Gathering data directly can be tedious, but it can also be very useful. Gathering data gives you a chance to obtain insights about aspects of behavior that are not usually recorded, such as the user’s affect, their posture, 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 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 mortgages are very common, whereas these words were less prevalent three or four years ago.
7 Example Research Studies [these are to get separated into breakout boxes]

We present here two example studies. In these examples, we show how to plan and prepare to run experiments with human participants. This section can help you obtain practical information for your own study.

7.1 Skill Retention Study

We draw our example from a study investigating the learning and forgetting performance of human participants across a set of procedural tasks. We present here specific procedures pertaining to the planning and running of the study.

Creation of Study Paradigm: We created a study paradigm to investigate how people forget what they have learned in a laboratory setting. For this study paradigm, we had to create a task that was novel enough to measure learning effects from participants (J. W. Kim, 2008). Thus, we chose a free spreadsheet called Dismal (Ritter & Wood, 2005) so that we could minimize any factors that would decrease the data’s validity. That is, the task of working with a spreadsheet was familiar to the learners, but the Dismal spreadsheet has not been used by many people. We also needed a tool to unobtrusively record the participant while he or she was performing the task; we chose RUI for this purpose. We then designed the experiment—the study uses a repeated measures design to allow multiple measures of learning and forgetting from each participant.

Here is a list of items that you can use to develop your own study.

- A task
- A task environment to perform the task in
- A tool to record behavior while performing that task

The First Pilot Test: We tested the study with a couple of pilot subjects. In the first pilot study, we had two male participants. We observed that participants had difficulty in learning the task.

Revising the Task Design: We decided to reduce the cognitive load of the target task so that we could measure learning effects in a restricted time. We reduced the task difficulty so that most of the learning was of the interface and not of the math in the task.

The Second Pilot Test: For the second pilot test, we had a female and a male participant who had no prior knowledge of the task. The task completion times for each were consistent with previous studies. In addition, the participants showed some forgetting over time, which was what we were studying. We concluded that the revised design was satisfactory. In some cases, it might be necessary to iteratively revise the study design further.

Getting IRB Approval: We prepared documents to receive IRB approval. To do this, we needed to provide detailed protocols specifying how we would run the experiments, as well as detailed methods for how we intended to recruit participants.

Start Running Experiments: After getting IRB approval, we started running the main experiment of approximately 30 subjects.
In addition to these sequential steps, we would like to share with you some practical information pertaining to the IRB process. When you prepare documents for the IRB, the forms will require you to address the following items. We also include the responses for the learning and retention study.

(a) The benefits of the study to participants:
From your participation, it is expected to obtain data representing how much knowledge and skills can be retained in the memory over time. This research can contribute to design a novel training program.

(b) Any known risks to the participant:
There is no risk to your physical or mental health. During the experiment, you can take a break at any time.

(c) How to achieve the participant’s privacy:
Your participation and data are entirely confidential. Personal identification numbers (e.g., PSU ID) will be destroyed after gathering and sorting the experimental data. Without personal identification, the gathered data will be analyzed and used for dissertation and journal publications. The following may review and copy records related to this research: The Office of Human Research Protections in the U.S. Department of Health and Human Services, the Social Science Institutional Review Board at PSU, and the PSU Office for Research Protections.

(d) Voluntary participation:
The participation of this study is purely based on volunteerism. You can refuse to answer any questions. At any time, you can stop and decline to continue the experiment. There is no penalty or loss of benefits if you refuse to participate or stop at any time.

(e) Compensation from the experiment:
Participants will receive monetary compensation of $25, $30, or $35 based on your total number of sessions, or extra credit (students registered in IST 331). The experiment consists of 5 to 7 trials ($5 per trial). The compensation will be given as one lump sum after all trials. For the amount of $30 and $35, participants will receive a check issued by Penn State. Others will receive $25 cash. Total research payments within one calendar year that exceed $600 will require the University to annually report these payments to the IRS. This may require you to claim the compensation that you receive for participation in this study as taxable income. For students in IST 331, you will receive 3% added to the total grade in the course. If you do not wish to take part in the research, you may earn the extra credit by completing the following:

- Choose a task to measure your learning and forgetting performance
- Gather your learning and forgetting data for 4 hours of study and test and 1 retention
- Analyze the data to show the total task time per study session and test session

One thing that we want to note here is that investigators should neither implicitly or explicitly force participants to participate in the experiment. As noted above, researchers recruiting students
who are enrolled in their classes must be particularly mindful of how they frame student participation in a study. When investigator use academic credits (extra or otherwise) for compensation, a balanced alternative must be made available.

7.2 Another example [to be caffeinezone, a smart phone app]
8 Afterword

There are many books available about research methods and related statistical analyses. We, however, realized that students usually do not have a chance to learn how to run their own experiments, and that there are no books that teach students practical knowledge about running experiments with human participants.

Students charged with running experiments frequently lack specific domain knowledge in this area. Consequently, young researchers chronically make preventable mistakes. With this book, we hope to assist students as they begin to obtain hands-on knowledge about running experiments. The topics and guidance contained in this book arise from the authors’ collective experience in both running experiments and mentoring students.

Further methods of gathering data are being developed. Though these changes will impact the development of future experimental procedures, the gross structures of a study and the aspects we have discussed here are not likely to change.

As you venture into research, you will find new topics that will interest you. In this text we are not able to examine all populations or touch upon measurements and tools that require additional training. Consequently, we are not able to cover in detail the collection of biological specimens, eye-tracking, or fMRI; however with further reading and consultation with colleagues, you will be able to master these skills.

Running studies is often exciting work, and it helps us understand how people think and behave. It offers a chance to improve our understanding in this area. We wish you good luck, bonne chance, in finding new scientific results.
Appendix A: A Checklist for Setting-up Experiments [Ritter would argue for keeping this as a checklist perhaps multipage, Illini Union used such things for their events…]

As an experimenter or a principal investigator for your project, you need to complete the items below to set up experiments to run.

- Prepare for the IRB form and submit it to office of research protection
- Run pilot tests to check your experimental design and apparatus
- Advertise your experiment to recruit participants (e.g., flyer, a student newspaper)
- Schedule participants
- Make sure a lab for the experiment is available for when you need to run
- Prepare how to debrief
Appendix A: Example Script to Run an Experiment

This is one page example script. While experiments will differ, this script includes many common elements.

Experimenter’s Guide

This is an example script for an experiment. Every experimenter should follow the procedures to run a user study about skill retention.

1. Check your dress code
2. Before your participants are coming in, you need to set up a set of the experiment apparatus.
   a) Start RUI in the Terminal Window. (see details ..)
   b) Start the Emacs text editor.
   c) Prepare disposable materials, handouts, such as informed consent form
3. Welcome your participants
4. Put a sign on the door indicating that you are running subjects when the experiment starts
5. Give the IRB approved consent form to the participant and have them read it
6. If they consent, start the experiment
7. Briefly explain what they are going to do
8. Give them the study booklet.
   a) Participants can use 30 min. maximum to study the booklet.
9. While participants are reading the booklet, you can answer their questions about the task.
10. Turn on the monitor that is located in the experimental room, so that you can monitor the participant outside the room.
11. When the experiment is finished, give an explanation about the payments or extra credit. Thank them; give them a debriefing form. Also, if there are any additional schedules for later measures, remind them.
12. Take down the sign on the door when the experiment is done
13. Copy the data to the external hard drive
14. Shut down apparatus
15. Make supplies for the next subject

Using RUI

RUI (Recording User Input) will be used to log keystrokes and mouse actions of the participant. RUI requires Mac OS X 10.3 (Panther) or later versions. It has been tested up to Mac OS X 10.4.3. (Tiger). In order for RUI to record user inputs, “Enable access for assistive devices” must be enabled in the Universal Access preference pane.

1. Launch Terminal
2. In Terminal, type the below information:
   ./rui –s “Subject Name” –r ~/Desktop/ruioutput.txt
3. You will get this message:
   rui: standing by – press ctrl+r to start recording…
4. Press “CTRL+r”
5. To stop recording, press “CTRL+s”

Note:
If you see the message of “bash: ./rui: Permission denied” in the Terminal window, you need to type “chmod a+x rui” while you are in the RUI directory.

**Measuring Learning and Forgetting**

Emacs is started by the experimenter for every session. The participants will start and stop RUI to record their performance. The experimenter needs to ensure that the participants cannot do mental rehearsal during the retention period.
Appendix D: Safety of Experiments

Some common safety concerns:
for cognitive psychology: there are none

for interesting things, see your IRB
for stress, see your IRB
for taking samples from humans, see your IRB
Appendix E: Example Consent Form

Here is an example of an informed consent form that you can refer to when you need to generate one for your experiment. This is taken from Kim’s thesis study (J. Kim, 2008).

Informed Consent Form for Biomedical Research
The Pennsylvania State University

Title: Investigating a Forgetting Phenomenon of Knowledge and Skills

Principal Investigator: Dr. Frank E. Ritter
316G IST Bldg, University Park, PA 16802
(814) 865-4453 frank.ritter@psu.edu

Other Investigators:

Dr. Jong Wook Kim
316E IST Building
University Park, PA 16802
(814) 865-6166; jongkim@psu.edu

Dr. Richard J. Koubek
310 Leonhard Building
University Park, PA 16802
(814) 865-7601 rkoubek@psu.edu

1. Purpose & Description: The purpose of the study is to investigate how much knowledge and skills are forgotten and retained in human memory after a series of learning sessions. Human performance caused by forgetting will be quantitatively measured. If you decide to take part in this experiment, please follow the experimenter’s instruction.

The experiment is held at 319 (Applied Cognitive Science Lab.) or 205 (a computer lab) IST building. During the experiment, the timing of keystrokes and mouse movements will be recorded.

A group of participants (80 participants) selected by chance will wear an eye-tracker to measure eye movements during the task, if you consent to wear the device. You can always refuse to use it. The eye-tracker is a device to measure eye positions and eye movements. The eye-tracker is attached to a hat, so you just can wear the hat for the experiment. The device is examined for its safety. You may be asked to talk aloud while doing the task.

2. Procedures to be followed:
   a. You will be asked to study an instruction booklet to learn a spreadsheet task (e.g., data normalization). Each study session will be 30 minutes maximum. For four days in a row, you will learn how to do the spreadsheet task.
   b. Then, you will be asked to perform the given spreadsheet tasks on a computer (duration: approximately 15 minutes).
   c. With a retention interval of 6-, 9-, 12-, 18-, 30-, or 60-day, after completing the second step, you will be asked to return to do the same spreadsheet task (duration: approximately 15 min/trial)
3. **Voluntary Participation:** The participation of this study is purely based on volunteerism. You can refuse to answer any questions. At any time, you can stop and decline the experiment. There is no penalty or loss of benefits if you refuse to participate or stop at any time.

4. **Right to Ask Questions:** You can ask questions about this research. Please contact Jong Kim at jongkim@psu.edu or 814-865-6166 with questions, complaints, concerns, or if you feel you have been harmed by this research. In addition, if you have questions about your rights as a research participant, contact the Pennsylvania State University’s Office for Research Protections at (814) 865-1775.

5. **Discomforts & Risks:** There is no risk to your physical or mental health. You may experience eye fatigue because you are interacting with a computer monitor. During the experiment, you can take a break at any time.

6. **Benefits:** From your participation, it is expected to obtain data representing how much knowledge and skills can be retained in the memory over time. This research can make a contribution to design a novel training program.

7. **Compensation:** Participants will receive monetary compensation of $25, $30, or $35 in terms of your total trials, or extra credits (students registered to IST 331). The experiment consists of 5 to 7 trials ($5 per trial). The compensation will be given as one lump sum after all trials. For the amount of $30 and $35, participants will receive a check issued by Penn State. Others will receive a cash of $25. Total research payments within one calendar year that exceed $600 will require the University to annually report these payments to the IRS. This may require you to claim the compensation that you receive for participation in this study as taxable income.

8. **Confidentiality:** Your participation and data are entirely confidential. Personal identification numbers (e.g., PSU ID) will be destroyed after gathering and sorting the experimental data. Without personal identification, the gathered data will be analyzed and used for dissertation and journal publications. The following may review and copy records related to this research: The Office of Human Research Protections in the U.S. Department of Health and Human Services, the Social Science Institutional Review Board and the PSU Office for Research Protections.

You must be 18 years of age or older to take part in this research study. If you agree to take part in this research study and the information outlined above, please sign your name and indicate the date below.

You will be given a copy of this signed and dated consent for your records.

_________________________________________            _____________________________
Participant Signature                    Date

_________________________________________            _____________________________
Person Obtaining Consent (Principal Investigator)                Date
Appendix F: Example Debriefing Form
[This is the debriefing form used in the study reported in Ritter, Kukreja, and St. Amant (2007).]

Human-Robot Interaction Study Debriefing Form

Thank you for participating in our human-robot interface testing study.

From your participation we will learn how people use interfaces in general and Human-Robot interfaces in particular. These interfaces are similar to those used to interfaces used to work in hazardous areas including those used in rescue work at the World Trade Center. By participating, you have been able to see and use a new technology. The results can lead to improved interfaces for robots that replace humans in hazardous conditions.

You may also find the Robot project overview page useful and interesting.

If you have any questions, please feel free to ask the experimenter. You can also direct questions to Dr. Frank Ritter, (frank.ritter@psu.edu, 865-4453).
Appendix G: Example IRB Application

Your Internal Review Board will have its own review forms. These forms are based on each IRB’s institutional history, and the types of studies and typical problems (and atypical problems) that they have had to consider over time. Thus, the form we include here can only be seen as an example form. We include it to provide you with an example of the types of questions and more importantly the types of answers characteristic of the IRB process. You are responsible for the answers, but it may be useful to see examples to see how long they are, and how detailed they need to be.

Following is a form used in one of our recent studies in the lab (Paik, 2011).

Submitted by: Jaehyon Paik  
Date Submitted: April 09, 2010 10:41:33 AM  
IRB#: 33343  
PI: Frank E Ritter

Study Title

1>Study Title  A New Training Paradigms For Knowledge and Skills Acquisition  
2>Type of eSubmission  New

Home Department for Study

3>Department where research is being conducted or if a student study, the department overseeing this research study.  Industrial and Manufacturing Engineering

Review Level

4>What level of review do you expect this research to need? NOTE: The final determination of the review level will be determined by the IRB Administrative Office.Choose from one of the following: Expedited

5>Expedited Research Categories: Choose one or more of the following categories that apply to your research. You may choose more than one category but your research must meet one of the following categories to be considered for expedited review.  
[X] Category 7 – Research on individual or group characteristics or behavior (including, but not limited to, research on perception, cognition, motivation, identity, language, communication, cultural beliefs or practices, and social behavior) or research employing survey, interview, oral history, focus group, program evaluation, human factors evaluation, or quality assurance methodologies.
Basic Information: Association with Other Studies

6> Is this research study associated with other IRB-approved studies, e.g., this study is an extension study of an ongoing study or this study will use data or tissue from another ongoing study?  No

7> Where will this research study take place? Choose all that apply.
   [X] University Park

8> Specify the building, and room at University Park where this research study will take place.
   If not yet known, indicate as such.  The research will be held in 319 Information and Science Technology Building.

9> Does this research study involve any of the following centers?
   [X] None of these centers are involved in this study

10> Describe the facilities available to conduct the research for the duration of the study.  We will mainly use a computer, keyboard, mouse, and joystick to test this study. Through the computer, participants can access the specific website that are developed by us.

11> Is this study being conducted as part of a class requirement? For additional information regarding the difference between a research study and a class requirement, see IRB Guideline IV, “Distinguishing Class-Related Pedagogical (Instructional) Assignments/Projects and Research Projects” located at http://www.research.psu.edu/orp/areas/humans/policies/guide4.asp.  No

Personnel

12> Personnel List

<table>
<thead>
<tr>
<th>PSU User ID</th>
<th>Name</th>
<th>Department Affiliation</th>
<th>Role in this study</th>
</tr>
</thead>
<tbody>
<tr>
<td>jzp137</td>
<td>Paik, Jaehyon</td>
<td>Industrial and Manufacturing Engineering</td>
<td>Co-Investigator</td>
</tr>
<tr>
<td>fer2</td>
<td>Ritter, Frank E</td>
<td>Information Sciences and Technology</td>
<td>Principal Investigator</td>
</tr>
</tbody>
</table>

- Role in this study  Principal Investigator

First Name  Frank  Middle Name  E  Last Name  Ritter  Credentials  PhD
PSU User ID  fer2  Email Address  frank-ritter@psu.edu  PSU Employment Status  Employed

[ ] Person should receive emails about this application

Mailing Address  316G IST Building
Address (Line 2)  863 3528  Fax number  Pager Number  Alternate Telephone

Department Affiliation  Information Sciences and Technology

Identify the procedures/techniques this person will perform (i.e. recruit participants, consent participants, administer the study):  This person will administer the whole process of experiments and he will help to recruit participants in his class.

Describe the person's level of experience in performing the procedures/techniques described above:  He has lots of experience doing this kind of experiment. Most of his students who already had a Ph.D degree did similar experiment from writing an IRB application to doing an experiment.

- Role in this study  Co-Investigator

First Name  Jaehyon  Middle Name  Last Name  Paik  Credentials
PSU User ID  jzp137  Email Address  jzp137@psu.edu  PSU Employment Status  Not
How to run experiments: A practical guide

Employed or Student
[ ] Person should receive emails about this application

Mailing Address 125 Washington Place

Mail Code City State College State Pennsylvania ZIP Code 16801

Phone Number 814 876 0984 Fax number Pager Number Alternate Telephone

Department Affiliation Industrial and Manufacturing Engineering

Identify the procedures/techniques this person will perform (i.e. recruit participants, consent participants, administer the study): This person designed the entire experiments and will perform recruiting participants, receiving consent form from participants, controlling the whole process of experiments, and gathering and analyzing data from participants.

Describe the person's level of experience in performing the procedures/techniques described above: This person is a ph.d student in IE department, and he has experience of experiments with human participants in his class. He conducted a similar experiments during his Master student. He also has 5 years in industry, so he has no problem to design and develop the environment.

Funding Source

13> Is this research study funded? Funding could include the sponsor providing drugs or devices for the study. No

NOTE: If the study is funded or funding is pending, submit a copy of the grant proposal or statement of work for review.

14> Does this research study involve prospectively providing treatment or therapy to participants? No

Conflict of Interest

15> Do any of the investigator(s), key personnel, and/or their spouses or dependent children have a financial or business interest(s) as defined by PSU Policy RA20, “Individual Conflict of Interest,” associated with this research? NOTE: There is no de minimus in human participant research studies (i.e., all amount must be reported). No

Purpose

16> Provide a description of the research that includes (1) the background, (2) purpose, and (3) a description of how the research will be conducted [methodology: step-by-step process of what participants will be asked to do]. DO NOT COPY AND PASTE THE METHODOLOGY SECTION FROM THE GRANT.

• Background/Rationale: Briefly provide the background information and rationale for performing the research study. Most research projects for exploring the effects on learning and retention by varying the training schedule have focused on two type of practice, distributed and massed. The results indicate consistently that the distributed practice has better performance on knowledge and skills acquisition than massed practice. However, a more efficient way might exist, and I assume that a more efficient way is the hybrid practice that uses the distributed practice and massed practice together. Through this study, I will explore more efficient practice strategy.

• Purpose: Summarize the study’s research question(s), aims or objectives [hypothesis]. This study has two objectives, in practical and theoretical way. The first objective is to explore the new paradigm of training strategy for tasks with declarative memory, procedural memory, and, perceptual-motor skill acquisition with different training schedules, such as distributed, hybrid 1 (massed placed in the middle of a regimen), and hybrid 2 (massed placed in the top of a regimen). And the results of each experiment are compared to verify which one is more efficient according to the task type. The second objective is to verify the results of three types of tasks with the learning and decay theories of the ACT-R cognitive architecture. The ACT-R cognitive architecture

57
provides learning and decay theories to predict human behavior in the ACT-R model. Using these theories, I will explore to verify and summarize the results of the tasks.

- **Research Procedures involving Participants:** Summarize the study’s procedures by providing a description of how the research will be conducted [i.e., methodology - a step-by-step process of what participants will be asked to do]. Numbering each step is highly recommended. **DO NOT COPY & PASTE GRANT APPLICATION IN THIS RESPONSE.**

  This research follows the order like below: 1. Participants have overall explanation of this research (the objective of the study, which data will be gathered, and so on). 2. After explanation, participants sign a consent form. 3. Participants will have a vocabulary word test for declarative memory, tower of hanoi game for procedural knowledge, and simple avoiding obstacle game for perceptual motor task game, each game takes no longer 5 minutes. 4. During the task, nothing will be asked to participants. 5. After experiments participants will be asked for not practicing the experiment until their second test.

17> **How long will participants be involved in this research study?** Include the number of sessions and the duration of each session - consider the total number of minutes, hours, days, months, years, etc. This experiment consists of 8 learning sessions and 1 testing session, and each session takes no longer than 20 minutes. The number of experiment days for participants varies according to the schedule type. Group 1 has 2 days, Group 2 has 8 days, and Group 3 has 4 days for the experiment.

18> **Briefly explain how you will have sufficient time to conduct and complete the research within the research period.** In the experiment day, Jaehyon will come to the office 1 hour early before the experiment to prepare the experiment, such as turn on the computer, launch the program, and launch a data correction program.

19> **List criteria for inclusion of participants:** 1. Participants should be older than 18 years. 2. Participants should have experience using a computer, keyboard, and mouse.

20> **List criteria for exclusion of participants:** 1. Participants should not have knowledge of Japanese vocabulary. 2. Participants should not have any experience of Tower of Hanoi game.

**Multi-Center Study**

21> **Is this a multi-center study (i.e., study will be conducted at other institutions each with its own principal investigator)?** No

**Participant Numbers**

22> **Maximum number of participants/samples/records to be enrolled by PSU investigators.**

   NOTE: Enter one number – not a range. This number should include the estimated number that will give consent but not qualify after screening or who will otherwise withdraw and not qualify for inclusion in the final data analysis. This number should be based on a statistical analysis, unless this is a pilot study, and must match the number of participants listed in the consent form. 30

23> **Was a statistical/power analysis conducted to determine the adequate sample size?** Yes

**Age Range of Participants**

24> **Age range (check all that apply):**

   [X] 18 - 25 years [X] 26 - 40 years

**Participant Information: Participant Categories**
How to run experiments: A practical guide

25> Choose all categories of participants who will be involved in this research study.
   [X] Healthy volunteers

26> Will Penn State students be used as study participants in this research study? Yes

27> Will students be recruited from a Subject Pool? No

28> Will participants be currently enrolled in a course/class of any personnel listed on this application? Yes

29> Describe the steps taken to avoid coercion and undue influence. We will not record any information of participants, so participants could decide to participate without any coercion.

30> Will participants be employees of any personnel listed on this application? No

31> Does this research exclude any particular gender, ethnic or racial group, and/or a person based on sexual identity? No

32> Could some or all participants be vulnerable to coercion or undue influence due to special circumstances (do not include children, decisionally impaired, and prisoners in your answer)? No

Recruitment

33> Describe the specific steps to be used to identify and/or contact prospective participants, records and/or tissue. If applicable, also describe how you have access to lists or records of potential participants. We will recruit participants with two ways. The first way is that participants will be recruited from class (IST 331). We will distribute experiment flyer for participating. The second way is that participants will be recruited by posting and emailing lists in department or college. We will also distribute experiment flyer to the department staffs, and we will ask them to distribute to students. In the experiment flyer, we describe that participants who have knowledge of Japanese vocabulary cannot participate this experiment for screening.

34> Will recruitment materials be used to identify potential participants? Yes

35> Choose the types of recruitment materials that will be used.
   [X] Letters/Emails to potential participants [X] Script - Verbal (i.e., telephone, face-to-face, classroom)

36> Describe how potential participants’ contact information (i.e., name & address) was obtained. We will ask department staff to broadcast our experiment.

37> Who will approach and/or respond to potential participants during recruitment? Jaehyon Paik

38> Explain how your recruitment methods and intended population will allow you access to the required number of participants needed for this study within the proposed recruitment period. This experiment is not a complex task. It takes no longer than 5 minutes each task, and it also has a simple game that can be attractive to the participants.

39> Before potential participants sign a consent document, are there any screening/eligibility questions that you need to directly ask the individual to determine whether he/she qualifies for enrollment in the study? [X] Yes
40> During screening/eligibility questions, will identifiable information about these individuals be recorded? No

41> Will investigators access medical charts and/or hospital/clinic databases for recruitment purposes? No

42> Will physicians/clinicians provide identifiable, patient information (e.g., name, telephone number, address) to investigators for recruitment purposes? No

43> Will researchers who are not involved in the care of potential participants review and/or use protected health information before a consent/authorization form is signed in the course of screening/recruiting for this research study (e.g., reviewing medical records in order to determine eligibility)? No

**Participant Consent/Assent**

44> When and where will participants be approached to obtain informed consent/assent [include the timing of obtaining consent in the response]? If participants could be non-English speaking, illiterate, or have other special circumstances, describe the steps taken to minimize the possibility of coercion and undue influence. The consent form will be given to participants at the first day in the experiment location. Participants should speak and hear English.

45> Who will be responsible for obtaining informed consent/assent from participants? Jaehyon Paik

46> Do the people responsible for obtaining consent/assent speak the same language as the participants? Yes

47> What type of consent/assent will be obtained? Choose all that apply.

[X] Implied consent – participants will not sign consent form (e.g., mail survey, email, on-line survey)

48> One of the following two conditions must be met to allow for a process other than signed informed consent to be utilized. Choose which condition is applicable. Choose only one.

[X] The research presents no more than minimal risk of harm to participants & involves no procedures for which signed consent is normally required outside of the research context.

49> Explain how your study fits into this condition. The experiment that we will have has not any harm for the participants. We just use a computer, mouse, and keyboard, that is, this experiment may part of our life.

50> If multiple groups of participants are being utilized (i.e., teachers, parents, children, people over the age of 18, others), who will and will not sign the consent/assent form? Specify for each group of participants. Participants should read the consent form, and do not need to sign, because we provide implied informed consent form.

51> Participants are to receive a copy of the informed consent form with the IRB approval stamp/statement on it. Describe how participants will receive a copy of the informed consent form to keep for their records. If this is not possible, explain why not. The implied informed form includes contents that "your participation in this research is confidential", and the form will be given to the participants before the experiment.
Cost to Participants: Compensation

52> Will the participant bear any costs which are not part of standard of care?  No
53> Will individuals be offered compensation for their participation?  No

Data Collection Measures/Instruments

54> Choose any of the following data collection measures/instruments that will be used in this study. Submit all instruments, measures, interview questions, and/or focus group topics/questions for review.
   [X] Knowledge/Cognitive Tests

55> Will participants be assigned to groups?  Yes
56> Will a control group(s) be used?  Yes
57> Choose one of the following:
   [X] Other control method

58> Describe the ‘other’ control method.  The difference variable is training schedule in this study.

Drugs/Medical Devices/Other Substances

59> Does this research study involve the use of any of the following? Choose all that apply.
   [X] None of the above will be used in this research study

Biological Specimens

60> Will biological specimens (including blood, urine and other human-derived samples) be used in this study?  No

Recordings - Audio, Video, Digital, Photographs

61> Will any type of recordings (audio, video or digital) or photographs be made during this study?  No

Computer/Internet

62> Will any data collection for this study be conducted on the Internet or via email (e.g. on-line surveys, observations of chat rooms or blogs, on-line interviews surveys via email)?  Yes
63> Is there a method in place to authenticate the identity of the participants?  No
64> Explain why an authentication method is not in place to identify respondents.  We do not collect information of participants.
65> Will data be sent in an encrypted format?  No
66> Explain why the data will not be sent in an encrypted format.  We do not record information of participants.
67> Will a commercial service provider (i.e., SurveyMonkey, Psych Data, Zoomerang) be used to collect data or for data storage?  No

**Risks: Potential for and Seriousness of**

68> List the potential discomforts and risks (physical, psychological, legal, social, or financial) AND describe the likelihood or seriousness of the discomforts/risks. For studies presenting no more than minimal risk, loss of confidentiality may be the main risk associated with the research. Memorize the Japanese vocabulary may discomfort participants.

69> Describe how the discomforts and risks will be minimized and/or how participants will be protected against potential discomforts/risks throughout the study (e.g., label research data/specimens with code numbers, screening to assure appropriate selection of participants, identify standard of care procedures, sound research design, safety monitoring and reporting). We assume that there is no risk in this experiment. However, if participants feel discomfort in experiment, they can quit the experiment immediately, and they can make a reschedule or they can give up the experiment.

70> Does this research involve greater than minimal risk to the participants?  No

**Benefits to Participants**

71> What are the potential benefits to the individual participants of the proposed research study?  (If none, state “None.”) NOTE: Compensation cannot be considered a benefit. none

72> What are the potential benefits to others from the proposed research study? The result may show the needs of new training paradigm.

**Deception**

73> Does this study involve giving false or misleading information to participants or withholding information from them such that their “informed” consent is in question?  No

**Confidentiality**

74> Describe the provisions made to maintain confidentiality of the data, including medical records and specimens. Choose all that apply. [X] Locked offices

75> Describe the provisions made to protect the privacy interests of the participants and minimize intrusion. First of all, we do not store any privacy information of participants, and the collected data will be stored in locked office. Only experimenter, Jaehyon Paik, can access the data.

76> Will the study data and/or specimens contain identifiable information?  No

77> Who will have access to the study data and/or specimens? Jaehyon Paik (only)

78> Will identifiers be disclosed to a sponsor or collaborators at another institution?  No

79> Will a record or list containing a code (i.e., code number, pseudonym) and participants identity be used in this study?  No

80> What will happen to the data when the research has been completed? Choose one. [X] Stored for length of time required by federal regulations/funding source and then destroyed
How to run experiments: A practical guide

[minimum of 3 years]

81>Is information being collected for this research that could have adverse consequences for participants or damage their financial standing, employability, insurability or reputation? No

82>Will a “Certificate of Confidentiality” be obtained from the federal government? No

HIPAA (Health Insurance Portability and Accountability Act)

83>Will participant’s protected health information (PHI) be obtained for this study? No

Radiation

84>Will any participants be asked to undergo a diagnostic radiation procedure while enrolled in this study? No

Physical Activity

85>Will participants be required to engage in or perform any form of physical activity? No

86>Will any type of electrical equipment other than audio headphones be attached to the participants (e.g., EMG, EKG, special glasses)? Submit a letter regarding the most recent safety check of the x-ray equipment being used with the supporting documents for this application. No

Document Upload

ICFS Document 1001 Received 03/22/2010 11:19:22 - Adult Form Revised version of consent form

INSTRUMENTS Document 1001 Received 03/22/2010 11:47:14 - For data collection - All data are recorded in webpage Document 1002 Received 04/09/2010 10:37:36 - The screenshots for the tasks. Document 1003 Received 04/09/2010 10:38:13 - Task2 Document 1004 Received 04/09/2010 10:38:51 - task3

RECRUITMENT Document 1001 Received 03/22/2010 11:20:24 - Recruitment Material Revised version of recruitment mat Document 1002 Received 04/09/2010 10:16:47 - Eligibility Screening This document for Eligibility Scr

SUBMISSION FORMS Document 1001 Received 03/22/2010 09:04:42 AM - Application Auto-generated by eSubmission Approval

• Click ADD to upload a new document for review
• Click REPLACE to upload a revised version of a previously submitted document (the radio button next to the document to be revised must be selected before clicking replace)
• Click REMOVE to delete a document. NOTE: Documents can be deleted at any time prior to submission. If an eSubmission is returned for additional information, only new uploaded documents can be deleted.
• To view a document just click on the document name. The following file types can be uploaded: .doc, .docx, .xls, .xlsx, .ppt, .pptx, .pub, .tif, .tiff, .txt, .pdf, .rtf, .jpg, .gif
References [ritter maintains with endnote, but some are from Kim?]


**Index**

How do I know my study measures what I want it to?

See:

How do I start to plan a study?

See:

Do I need to get IRB for my work?

See:

What should I do if I don’t need to get IRB?

See:

All authors of papers

Index terms: VPA, IV, DV, condition

Frank sez: I don’t know what to do with this, Rich or Jon, please merge into into or delete to leave here. This was the glossary, and we are (re)moving it?

<table>
<thead>
<tr>
<th>Independent variable</th>
<th>A variable that is manipulated in the study, either by assignment of materials or assignment of subjects.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dependent variable</td>
<td>A measurement that is taken during the study, such as reaction time, or percent correct. It depends on other things.</td>
</tr>
<tr>
<td>Pilot study</td>
<td>An abbreviated version of the study done to test the procedure and prepare for a larger study.</td>
</tr>
<tr>
<td>Power</td>
<td>The power in an experimental study indicates the probability that the test (or experiment) will reject a false null hypothesis. Failure to reject the null hypothesis when the alternative hypothesis is true is referred to as a Type II error. Thus, as the power of a study increases, the chances of a Type II error decrease.</td>
</tr>
<tr>
<td>IRB</td>
<td>Internal Review Board. They review study proposals to ensure safety and compliance with US federal regulations.</td>
</tr>
<tr>
<td>Informed consent form</td>
<td></td>
</tr>
<tr>
<td>Null hypothesis</td>
<td>The hypothesis that the treatment DOES NOT lead to differences. For example, the null hypothesis might be that two interfaces are equally easy to use.</td>
</tr>
</tbody>
</table>
How to run experiments: A practical guide