RESEARCH METHODS AND STATISTICS

A Unit Lesson Plan for High School Psychology Teachers

AMERICAN PSYCHOLOGICAL ASSOCIATION
TOPSS
TEACHERS OF PSYCHOLOGY IN SECONDARY SCHOOLS

Don Kober, MA; Scott Reed, MEd; Hilary Rosenthal, MEd; and Allyson J. Weseley, EdD
Revision Team

Stephen L. Chew, PhD, Samford University
Faculty Consultant

Developed and produced by the Teachers of Psychology in Secondary Schools (TOPSS) of the American Psychological Association, January 2017.

RESEARCH METHODS AND STATISTICS

A Five-Unit Lesson Plan for High School Psychology Teachers

This unit is aligned to the following content and performance standards of the *National Standards for High School Psychology Curricula* (APA, 2011):

CONTENT AND PERFORMANCE STANDARDS	SUPPORTING DOCUMENTS AND ACTIVITIES
CONTENT STANDARD 1: Research methods and measurements used to study behavior and mental processes Students are able to (performance standards): 1.1 Describe the scientific method and its role in psychology. 1.2 Describe and compare a variety of quantitative (e.g., surveys, correlations, experiments) and qualitative (e.g., interviews, narratives, focus groups) research methods. 1.3 Define systematic procedures used to improve the validity of research findings, such as external validity. 1.4 Discuss how and why psychologists use non-human animals in research.	Lesson 1 Content Outline Activity1.1: Sampling or Assignment? Activity 1.2 : A Tasty Sample(r): Teaching About Sampling Using M&M's Activity 1.3: Do Cookies/Donuts Improve Memory? Errors in Methodology Lesson 2 Content Outline Activity 2.1: Counting Fidgets: Teaching the Complexity of Naturalistic Observation Lesson 3 Content Outline Activity 3.1: Pattern Recognition Critical thinking exercises A, B, C, D, E
CONTENT STANDARD 2: Ethical issues in research with human and non-human animals Students are able to (performance standards): 2.1 Identify ethical standards psychologists must address regarding research with human participants. 2.2 Identify ethical guidelines psychologists must address regarding research with non-human animals.	Lesson 4 Content Outline Activity 4.1: Research Ethics Critical thinking exercises F and G
CONTENT STANDARD 3: Basic concepts of data analysis Students are able to (performance standards): 3.1 Define descriptive statistics and explain how they are used by psychological scientists. 3.2 Define forms of qualitative data and explain how they are used by psychological scientists. 3.3 Define correlation coefficients and explain their appropriate interpretation. 3.4 Interpret graphical representations of data as used in both quantitative and qualitative methods. 3.5 Explain other statistical concepts, such as statistical significance and effect size. 3.6 Explain how validity and reliability of observations and measurements relate to data analysis.	Lesson 5 Content Outline Activity 5.1: Statistical Significance Critical thinking exercise H
PROPOSED NUMBER OF DAYS/HOURS FOR LESSON: Recommended number of teaching hours: **5*** **6** days in 50-minute classes = **5** hours **3** days in 90-minute classes = **4.5** hours *See Introduction.	

CONTENTS

This project was supported by a grant from the American Psychological Foundation.

PROCEDURAL TIMELINE

LESSON 1: THE SCIENTIFIC METHOD AND PSYCHOLOGICAL SCIENCE

Activity 1.1: Sampling or Assignment?

Activity 1.2: A Tasty Sample(r): Teaching About Sampling Using M&M's

Activity 1.3: Do Cookies/Donuts Improve Memory? Errors in Methodology

LESSON 2: RESEARCH METHODS

Activity 2.1: Counting Fidgets: Teaching the Complexity of Naturalistic Observation

LESSON 3: RESEARCH METHODS, CONTINUED

Activity 3.1 Pattern Recognition

LESSON 4: ETHICAL ISSUES IN RESEARCH

Activity 4.1: Research Ethics

LESSON 5: STATISTICS

Activity 5.1: Statistical Significance

INTRODUCTION

This lesson plan provides content outlines, activities, critical thinking exercises, and resources for teachers to use to teach research methods and statistics, the foundation of any psychology course. Suggested activities are referenced within the lesson content outlines where appropriate, and all activities may be found together in a separate section after the content outlines. Additionally, a list of references and resources are found at the end of the lesson plan.

Some of the content provided in this lesson plan is too advanced for an introductory or on-level psychology class. We have highlighted this advanced content accordingly for introductory or on-level teachers to see.

This project was supported by a grant from the American Psychological Foundation. The authors and the APA TOPSS Committee thank Barney Beins, PhD, of Ithaca College, Julie Penley, PhD, of El Paso Community College, El Paso, TX, and Bonnie Green, PhD, of East Stroudsburg University, for their reviews of this lesson plan.

CONTENT OUTLINE

LESSON 1

The Scientific Method and Psychological Science

Overview: Psychology is a science because researchers establish knowledge and test hypotheses using scientific research methods. Understanding these research methods is essential to being able to think critically about psychology.

I. Psychology and the Scientific Method

A. The **scientific method** is the systematic, empirically based investigation of phenomena through objective observations and measurements and the formulation of testable and falsifiable explanations.

B. The scientific method is actually a set of multiple methods.

C. Based on the particular research question, a researcher will identify what type of method to use.

D. Psychologists, like all scientists, conduct research to describe, measure, predict, and explain the phenomena in which they are interested.

II. Basic Concepts of the Scientific Method

Using the scientific method, psychologists make systematic, precise observations to generate ideas about behavior and to test theories and hypotheses.

A. Theories and hypotheses

1. **Theories** are coherent sets of concepts that explain a phenomenon or set of phenomena.

2. **Hypotheses** are testable statements about the relationship between two variables. Hypotheses are based on observation, research, and theory and are not random guesses.

B. Variables: A **variable** is any factor that can take on different values. It is the opposite of a constant.

1. In research, variables can be manipulated, controlled, or measured.

2. There are two basic kinds of variables in research:

   a. **Independent variables** (hypothesized causes) are variables that the researcher manipulates or selects to test the hypothesis that the variable leads to a change.

   b. **Dependent variables** (hypothesized effects) are outcome or performance measures used to determine the effect of the independent variable.

ADVANCED

3. **There are two types of independent variables:**

   a. A *true independent variable* is one that the researcher manipulates. The researcher assigns participants to a condition.

   b. A *quasi-independent variable* is one in which the researcher selects people for having a certain trait or property (i.e., participants are chosen based on their gender, or because of a particular condition, such as brain trauma).*

C. **Sampling** is the way a researcher selects participants from a population.

1. Representative, unbiased sampling is critical for internally and externally valid results.

2. A biased sample undermines the validity of the results and limits how well the results might generalize to the intended population.

---

* Throughout this lesson plan, items boxed in gray indicate content for an advanced psychology course.

BACK TO CONTENTS

For example, say a researcher wants to determine college students' political beliefs. If the researcher only surveys psychology majors, the results might not reflect the views of the general college population.

3. Other important terms include:

   a. **Population**—the entire group of people one is interested in studying.

   b. **Sample**—the subset of participants selected from the population.

   c. **Representative sample**—this kind of sample reflects the characteristics of the population.

   d. **Random sampling**—in random sampling, every individual in the population has an equal chance of being selected for the sample, which helps researchers select a representative sample.

   e. **Convenience sampling**—in convenience sampling, the researcher selects participants who are available, such as members of an introductory psychology class.

   Convenience sampling does not result in representative samples, and generalizing from a convenience sample to the population may not be possible.

GO TO ACTIVITY1.1

Sampling or Assignment?

GO TO ACTIVITY1.2

A Tasty Sample(r): Teaching About Sampling Using M&M's

III. The Research Process

A. A researcher develops a hypothesis and designates the independent and dependent variables, which must be operationally defined.

An **operational definition** is a definition of a variable or condition in terms of the specific operation, procedures, or observable behaviors. The researcher must specify exactly how variables will be manipulated, controlled, or measured.

B. Before conducting research that involves human participants, the investigator must submit the detailed plans of the project to an **institutional review board (IRB)** to ensure that the method and procedures follow the ethical guidelines for conducting research.

C. After obtaining IRB approval, the researcher will then conduct the research, collect and analyze the data, and report the findings.

D. Once the data are collected, the researcher uses statistical techniques to analyze them.

1. **Descriptive statistics** (e.g., mean, median, and mode, the standard deviation, range) are used to characterize and summarize major trends in the data.

2. **Inferential statistics** (e.g., t-tests, ANOVA) are used to draw conclusions about the data and make generalizations from the results to the larger population.

E. The final step in any scientific process is making findings public through publication and/or presentation and open to scrutiny and **replication** by other scientists.

GO TO ACTIVITY 1.3

Do Cookies/Donuts Improve Memory? Errors in Methodology

IV. Use of Nonhuman Animals in Research

A. Psychologists conduct research with nonhuman animals to study a wide range of normal and abnormal behaviors, and the biological mechanisms underlying these behaviors.

1. About 7–8% of psychological research involves the use of nonhuman animals (APA, n.d.).

2. Basic research with nonhuman animals has enabled scientists to learn more about sensory processes, motivational systems, learning, memory, cognition, evolution, and development, all of which have contributed to the health and welfare of both humans and other animals.

3. Scientists have learned much about the central nervous system from conducting research with nonhuman animals.

4. Psychological scientists use a variety of animal models to study different behaviors and disorders. Rodents and birds, primarily mice, rats, and pigeons, make up approximately 90% of animals in psychological research. Nonhuman primates, such as monkeys and apes, are involved in research to a much lesser extent (APA, n.d.).

B. Reasons for conducting nonhuman animal research

1. Ethical considerations

2. Understanding behavior

3. Evolutionary change and influence

4. Increased control

5. Such research benefits nonhuman animals as well

Critical Thinking Exercise

A. You conduct an experiment where you take a sample of high school students and randomly divide them into two groups. Both groups view the same videotaped lecture. For the experimental group, a cell phone goes off at random intervals during the lecture. For the control group, there are no such distractions. After the video, both groups are given the same test of the material.

1. Name the independent and dependent variables.

2. Name two factors that were controlled in this study.

3. Identify two operational definitions of variables.

4. What kinds of descriptive statistics would you compute once you have the results?

5. Name your sample and your population in this study.

LESSON 2

Research Methods

Overview: In this lesson, research methods are introduced. This lesson covers descriptive research methods such as case studies, naturalistic observation, and surveys or questionnaires and also covers qualitative research methods such as interviews, narratives, and focus groups.

I. Research Methods Can Be Qualitative or Quantitative

ADVANCED

A. **Qualitative** research methods are those that collect and analyze non-numeric data. Researchers collect such data through interviews, focus groups, and narratives.

B. **Quantitative** research methods are those that collect numerical data to investigate phenomena of interest. Researchers gather such data through surveys, correlations, and experiments.

II. Descriptive Research Methods

Descriptive research methods are used to describe phenomena and can be either qualitative or quantitative. Descriptive methods **cannot** be used to establish cause-and-effect relationships.

A. A **case study** is an in-depth investigation of an individual or small group who may have a highly unusual trait.

1. For example, in *The Man Who Mistook His Wife for a Hat* (1985), Oliver Sacks presents case studies of some of his patients. An in-depth investigation of Phineas Gage would also be a case study.

2. Strengths include that case studies are especially useful for documenting phenomena that are rare or complex and that the research can lead to hypotheses and questions for further research.

3. Weaknesses include retrospective (hindsight) bias, where researchers inflate the importance of events that, in retrospect, seem relevant and overlook events that do not seem relevant. Case studies also have limited generalizability due to small sample size.

B. **Naturalistic observation** occurs when researchers collect observations of natural, ongoing behavior.

1. For example, researchers may observe people in multiple cultures and estimate the number of inches between two people having conversations; researchers may observe a street intersection to see how many people stopped at stop signs.

2. Strengths include the fact that all factors that might influence the behavior are present (as opposed to controlled experiments). Thus, the results have applicability to “real-world” questions.

3. Weaknesses include:

   a. The potential for both researcher and participant bias (sometimes also called subject bias, demand characteristics, and expectancy effects).

   b. Possible privacy issues due to the inability to obtain informed consent.

   c. An inability to establish causality because of lack of control over variables.

   d. The time required to conduct naturalistic observation, especially for rare events—a common practical difficulty.

GO TO ACTIVITY 2.1

Counting Fidgets: Teaching the Complexity of Naturalistic Observation

C. **Surveys or questionnaires** provide self-report data about attitudes, behaviors, or characteristics.

1. Strengths of surveys include that they can be easily administered and are efficient for gathering large amounts of data.

2. Weaknesses include:

   a. Biased samples either through poor sampling, selection bias, or low return rate.

   b. Self-presentation and social desirability biases in responding to questions.

c. Ambiguously worded or leading questions ("framing" of the questions) that bias responses.

d. Memory distortions in self-report.

ADVANCED

III. Qualitative Research Methods

A. Qualitative Data

1. Many complex behaviors do not lend themselves to accurate numerical measurement, such as multidimensional variables (e.g., interpretation of poetry) and categorical variables (e.g., beginner–novice–expert at a skill). In such cases, researchers employ qualitative research methods.

2. Often researchers measure both quantitative and qualitative data.

3. Qualitative research can capture complex experiences that numerical analysis does not, but the data cannot be summarized easily.

4. Types of qualitative research methods:

    a. In **interviews**, individual participants provide a verbal description of an event or behavior. For example, a researcher may interview blind participants to see how they navigate as pedestrians.

    b. **Narratives** are stories and in the narrative research method, the researcher collects the personal stories of participants.

    c. A **focus group** involves an interview with a sample of people that is representative of an important section of the population.

ADVANCED

Critical Thinking Exercises

B. You want to study how blind pedestrians navigate through neighborhoods with only a cane. Describe how you might use descriptive research methods such as case study, survey, and naturalistic observation. What would be the pros and cons of each method?

C. You are interested in studying attitudes about a controversial topic, such as euthanasia or abortion. What kinds of qualitative research methods might you use to study these issues? What would be the strengths and weaknesses of such methods, especially compared with quantitative methods, such as surveys?

LESSON 3

Research Methods, Continued

Overview: In this lesson, predictive (correlational) and experimental research methods are described, along with issues of validity.

I. Correlational Research Methods

Correlational methods measure the relatedness of two variables and attempt to predict the value of one based on the other.

A. **Correlations** examine the relationship **between two variables** without manipulating either one.

1. Strengths of correlational research include that it provides an index of strength of the relationship between variables and can be used to predict future behavior. Correlational research can be done with existing data or with variables that cannot be manipulated or are unethical to manipulate.

2. A weakness is that because there is no manipulation of an independent variable, the researcher cannot establish a cause-and-effect relationship using a correlation.

3. In a correlation, one variable might cause the other, but correlational designs do not give us enough information to make that determination. Correlation does not equal causation.

ADVANCED

4. The directionality problem is a limitation in correlations because even if one variable causes the other, we can't determine the direction of causation.

   For example, if a study found that people who retire later in life are healthier, it may imply that working longer leads to better health. However, it is equally likely that less healthy people retire earlier because of health problems.

5. The third variable problem refers to the fact that two variables that are correlated may not be directly related to one another; rather, a third variable might be affecting them both.

For example, many people believe that sugar is related to hyperactivity in children; research has shown no such relation. Rather, the activities (e.g., birthday parties) surrounding sugar consumption can be the causal factor.

ADVANCED

6. **Illusory correlations** are when two data sets show a correlation even though there is no relationship at all.

   For example, violent video game producers have tried to tie the increase in game sales to the decrease in violent crimes as reported by the F.B.I. While this is a negative correlation, it is doubtful that the two sets of data are related. Other examples are the perceived positive correlation between ice cream sales and murders (both are correlated to hot summer months), and the tendency to perceive members of minority groups as being correlated to unique, distinctive events.

B. **Correlation coefficients** are statistical measures of the relationship between two variables. Correlation coefficients are expressed from -1.00 (a perfect negative correlation) to +1.00 (a perfect positive correlation).

1. A correlation is reported as a decimal value (e.g., .60, -.70, .90). To interpret a correlation, break it down into two parts, the sign (positive or negative) and the absolute value or magnitude (between 0.0 and 1.0).

2. The direction of relationship is either positive or negative:

   a. In a **positive** correlation, as one variable increases, so does the other (and vice versa). For example, as class attendance increases, so do grades.

   b. In a **negative** correlation, as one variable increases, the other decreases (and vice versa). For example, as class absences increase, grades decrease.

3. The closer the absolute value of the coefficient is to 1.0, the stronger the correlation between the variables. The closer to zero, the weaker the relationship between variables. Therefore, a correlation of -.75 is stronger than a correlation of +.46.

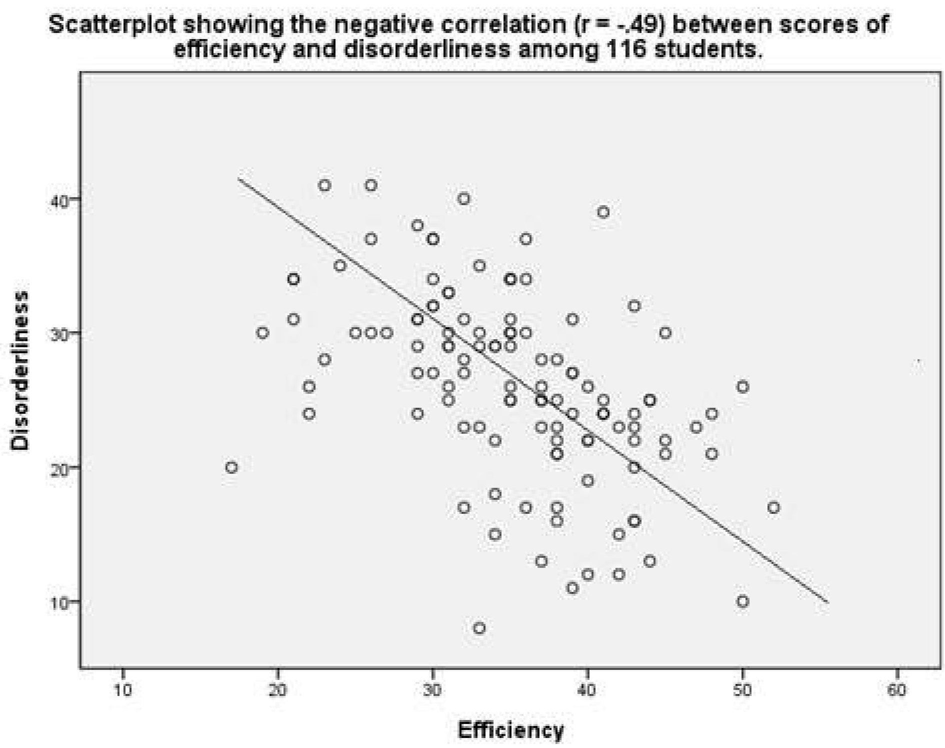
C. Scatterplots

1. **Scatterplots** are typically used to illustrate correlations. They show the relationship between two dependent variables.

2. The values of one variable are marked on the *x*-axis, and the values of the other variable are marked on the *y*-axis. Unlike other graphs, each individual data point is displayed as a point whose location is determined by its value on each variable. The closer the data points are to forming a line, the stronger the correlational relationship is.

3. Here is an example of a scatterplot showing a negative correlation. The figure below reflects a negative correlation between two traits, disorderliness and efficiency. In this example, as a person's disorderliness level increases, the efficiency score decreases:

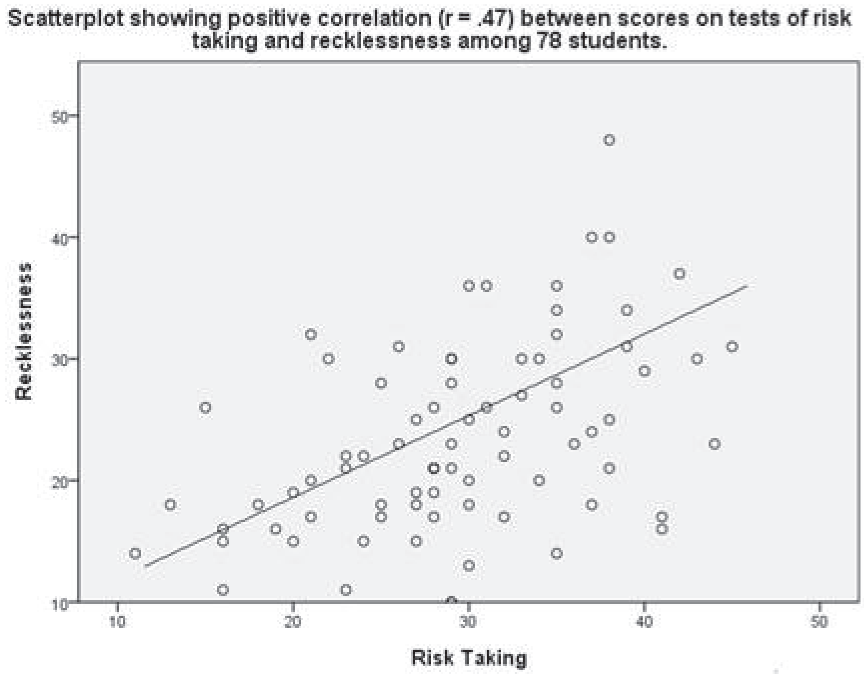
Figure 1



BACK TO CONTENTS

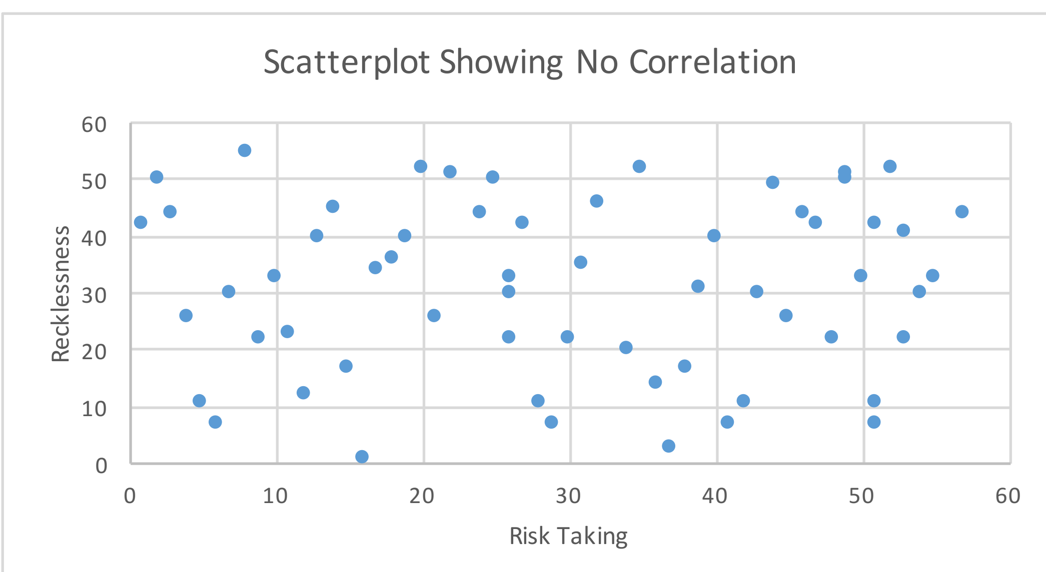
4. Here is an example of a scatterplot showing a positive correlation. The figure below reflects a positive correlation between scores on tests of recklessness and risk taking. In this example, as a person's recklessness increases, their risk-taking score increases:

Figure 2



Here is an example of a scatterplot showing no correlation:

Figure 3



II. Experimental Methods

Experimental methods involve both independent and dependent variables.

A. In **experiments**, manipulation or selection of one variable (the **independent variable**) takes place under controlled conditions to observe its effect on another variable (the **dependent variable**).

1. Experiments require multiple groups or conditions. At its most basic, an experimental group and control group is used:

    a. The **experimental group** is the group that receives the treatment.

    b. The **control group** is the group that does not receive treatment or receives a treatment presumed to be ineffective (e.g., a placebo). The control group serves as the basis for comparison of results from the experimental group.

        i. Proper control groups ensure that the impact of the independent variable can be assessed accurately. A control group should experience and do everything the experimental group does EXCEPT the independent variable.

        ii. Having a control group allows comparison of a treatment condition to a nontreatment condition to determine if the independent variable affected the dependent variable.

        iii. Administering a **placebo** (e.g., drug or behavioral treatment that the participant believes is a treatment but is actually not) is a common way to create a control group.

        iv. **Random assignment** is used to create comparable groups.

2. In experiments a distinction is made between true experiments and quasi-experiments.

    a. **True experiments** use true independent variables and are the only research method that allows a cause-and-effect relationship to be established.

BACK TO CONTENTS

b. **Quasi-experiments** treat subject variables (e.g., gender, race, age) like they are independent variables. They are like correlations in that no causality can be established.

ADVANCED

3. We can also distinguish between **lab experiments**, in which control is maximized but mundane realism (i.e., the real-world environment) is sacrificed, and **field experiments**, in which realism is maximized but control is sacrificed.

4. Strengths of experiments include:

   a. In experiments in general, the researcher has greater control over the whole situation than in any other research method.

   b. True experiments address causality.

5. Weaknesses of experiments include:

   a. **Generalizability**: Since experimental studies are often conducted in a controlled lab setting, the results may not reflect real-world events because some important factors may be missing from the experimental situation (compared with naturalistic observation).

   b. **Confounds or confounding variables**: These are uncontrolled variables that affect the outcome of the experiment; they are variables for which the researcher is not interested that covary with the independent variable and are almost always the result of a research design flaw. The researcher may believe the result is due to the independent variable, but it may really be due to an uncontrolled, confounding variable.

   c. **Ethical considerations**: All research has ethical dimensions, but this is especially true with experiments because the experimenter is manipulating the behavior of participants and causing them to do things they may not normally do.

   d. **Participant bias**: A participant knows he or she is in an experiment and may not act naturally.

   e. **Experimenter bias**: the experimenter may treat participants differently and influence their behavior according to the research hypothesis.

ADVANCED

f. **Carryover effect**: Sometimes what the participant does in one task may influence how they act in another task.

B. Longitudinal research and cross-sectional research methods are types of experimental methods typically used to study behavior over time.

1. In **longitudinal research**, the researcher studies the same group of people over a long period of time to see developmental changes.

2. In **cross-sectional research**, the research design involves the comparison of people of different age groups (cohorts) at the same time.

GO TO ACTIVITY 3.1

Pattern Recognition

ADVANCED

III. Validity in Research

Validity means that an operational definition is a true and accurate reflection of the phenomenon being studied.

A. Internal validity

1. **Internal validity** means that the results of a study reflect the effects of the variables being studied.

2. Here is how researchers try to control for threats to internal validity:

   a. In a **blind study**, participants are ignorant ("blind") to their group assignment—they are not told whether they are in the treatment or control condition, which controls for participant bias.

   b. In a **double blind study**, both participants and the experimenter are unaware of group assignment, which controls for both participant and experimenter bias. At least one person will know which condition each participant is in, which is critical for interpreting results.

ADVANCED

c. **Counterbalancing** is used to control for any order effects, where the impact of different levels of a variable depends on the order in which they occur. To counterbalance, the researcher would change the order of tests from one person to the next.

   For example, say a psychologist conducts an experiment on performance times (e.g., runners' sprint times). One trial may be done using the runner's typical preparation, with a second trial done using visualization (e.g., the runner is asked to visualize their peak performance). The researcher would want to counterbalance which preparation condition comes first. Half of the participants would do typical preparation first while the other half would do visualization first.

B. **External validity** is related to whether the research will generalize to other contexts outside the particular research setting. For example, if all participants in a given research study are White college age students from a private school, how can those results be generalized to a person who isn't college educated, or is 68 years old, or is African American or Latino?

Relevant questions include:

1. Is the sample in the research representative of the population to which the researcher wants to generalize the results?
2. Will the results of the research be the same if the psychologist repeats the study with different people or different nonhuman animals?
3. Will participants' behaviors recur if those behaviors are observed elsewhere—for example, outside of a laboratory rather than in a laboratory setting?

ADVANCED

Critical Thinking Exercises

D. Regarding the study of blind pedestrians (see Critical Thinking Exercise B), would it be better to use quantitative or qualitative methods? Describe both the qualitative and quantitative data you might measure.

E. You want to study the impact of video games on student learning. How might you study this with correlational methods and experimental methods? Design both studies and describe the strengths and weaknesses of each.

LESSON 4

Ethical Issues in Research

Overview: Research involves the systematic collection of data to further our knowledge about the world in which we live. Psychological research involves studying not only people but also other animals. In research, we observe the behavior of participants, and typically we control or manipulate their behavior. Such actions entail a responsibility on the part of psychologists to conduct research in an ethical manner. Conducting research is a privilege and investigators should be knowledgeable about effectively dealing with ethical issues that can arise throughout the research process.

I. Ethical Framework for Research With Human Participants

A. *Ethical Principles of Psychologists and Code of Conduct* (APA, 2010)

1. The APA Ethics Code governs the ethical conduct of psychologists both in practice and in research.

2. The code covers a wide range of ethical issues, from professional conduct of therapists to plagiarism in publications. Highlights of the code relating specifically to research include:

   a. informed consent

   b. freedom to withdraw

   c. protection from harm and discomfort

   d. confidentiality

   e. debriefing

3. APA also publishes ethical guidelines for a variety of specific situations and populations. There is a separate set of guidelines for high school students conducting research. See the Research Ethics section at **http://www.apa.org/science/about/publications/index.aspx** for more details.

B. Historical context: *Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research*

The current APA Ethics Code is strongly influenced by the Belmont Report, which was published in 1979 after public attention was drawn to abuses of human subjects, particularly in studies conducted by the Nazis around the time of World War II. The report provides an analytical framework to guide the resolution of ethical issues that arise during the conduct of research with human participants.

C. Oversight of research with human participants

ADVANCED

1. Federal regulations for the protection of human research participants were issued in 1981. These regulations specify in detail the basic requirements for ensuring that the rights and welfare of participants are protected, including:

   a. criteria for obtaining informed consent

   b. criteria for respecting the privacy of individuals

   c. the confidentiality of the data they provide

   Furthermore, these regulations prescribe how research with human participants should be monitored through the establishment of an institutional review board (IRB), which is charged with protecting the rights and well-being of research participants.

2. The IRB reviews research proposals, prior to the start of the study, to ensure that the study complies with all applicable federal regulations and institutional policies. The IRB has the authority to approve, require modifications to, or disapprove proposed research involving human participants.

3. IRBs must have at least five members with varying backgrounds, in both scientific and nonscientific areas, as well as one member who is not affiliated with institution.

II. Examples of Research With Ethical Concerns

A. The Tuskegee (Public Health Service) Syphilis Study (1932–1972)

B. Research on obedience to authority by Milgram

C. The Stanford prison experiment by Zimbardo

D. Although these classic examples might illustrate unethical research practices, ethical dilemmas in research typically may be much more nuanced. For example, in studying the creation of

BACK TO CONTENTS

false memories, is it ethical to plant a false memory of a negative experience in a child? In studying hand-washing behavior, is it an invasion of a person's privacy to observe and record his or her behavior in a public bathroom?

ADVANCED

III. Standards for the Humane Treatment and Care of Nonhuman Animals in Research

A. There are numerous safeguards, at various levels within the scientific community, to ensure the humane care and treatment of laboratory animals.

1. In 1966, Congress passed the Animal Welfare Act (AWA), which governs the care and use of many warm-blooded animals in research. Other federal policies include the *Policy on Humane Care and Use of Laboratory Animals* and the *Guide for the Care and Use of Laboratory Animals.*

2. An **Institutional Animal Care and Use Committee (IACUC)** is charged with reviewing all proposed research projects involving the use of nonhuman animals.

B. APA has also established guidelines for the care and use of nonhuman animals in research.

1. The *Guidelines for Ethical Conduct in the Care and Use of Laboratory Animals in Research* discusses all aspects of nonhuman animal use, including justification of the research, personnel, acquisition, care, and housing of laboratory animals, experimental procedures, field research, and the use of nonhuman animals in education. See **http://www.apa.org/science/leadership/care/guidelines.aspx**.

2. The *Guidelines for the Use of Nonhuman Animals in Behavioral Projects in Schools (K-12)* discusses the use of nonhuman animals in teaching demonstration and research projects at the elementary through high-school levels. See **http://www.apa.org/science/leadership/care/animal-guide.aspx**.

GO TO ACTIVITY4.1

Research Ethics

Critical Thinking Exercises

F. Assume that you are volunteering to participate in an experiment in the psychology department at a university. According the APA Ethical Guidelines, what must the researcher tell you before starting the experiment?

G. You want to test the idea that people are more likely to help others of their own race. You want to conduct an experiment where you will stage a kidnapping in front of a busy store to see who will seek help for the "victim." You will manipulate whether the person posing as the victim is of European or African descent. What are the ethical issues involved? If you were a member of the IRB, what questions would you ask of the researchers? Do you think such an experiment would be ethical? Can you design an experiment that tests the same idea but raises fewer ethical issues?

LESSON 5

Statistics

Overview: Psychologists use statistics to analyze research data. There are two basic kinds of statistics, descriptive and inferential. Descriptive statistics are used to summarize the major characteristics of a data set. Inferential statistics are used to draw conclusions about data (e.g., is the difference between means statistically significant?) and make generalizations from the sample to the population (e.g., if a researcher were to repeat this study with different participants, would he or she get the same results?). It is a critical component in psychological research, but it is a much more complex topic than descriptive statistics.

I. Descriptive Statistics

Descriptive statistics refer to a set of tools that permit the summary of the major characteristics of a large amount of data.

II. Frequency Distributions

ADVANCED

A. Frequency distributions are a simple yet effective way of summarizing a large data set and revealing the important characteristics of the data.

1. Instead of listing every data point, a frequency distribution shows the frequency of occurrence of any possible score or class of scores. Thus, each score is only shown once, followed by its frequency of occurrence.
2. For example, a data set of 1, 1, 2, 2, 2, 4, 4, 4, 4, 4, 5, 6, 6 is shown in the following frequency distribution. Note how each score is only listed once, and the frequency of occurrence makes it easy to count total frequency for each score:

Score	Freq.
1	2
2	3
3	0
4	5
5	1
6	2

ADVANCED

3. For data sets with a large range of scores from high to low, a researcher can create a grouped frequency distribution, which requires classes of scores and the number of cases in each class. To be effective, the classes must be mutually exclusive, or nonoverlapping, and of uniform width.

4. For example, for a set of data that range from 10 to 44, the grouped frequency distribution might look like the following:

Class	Freq.
10–14	1
15–19	3
20–24	6
25–29	4
30–34	4
35–39	1
40–44	2

5. Columns listing cumulative frequency, percent, and cumulative percent can be added to either grouped or ungrouped frequency distributions.

III. Quantitative Descriptive Statistics

When dealing with quantitative data, researchers often compute descriptive statistics that can summarize important characteristics of the data numerically. There are two basic classes of quantitative statistics: measures of central tendency and measures of variability or dispersion.

Note: Instruction on how to calculate statistics are not included here. If you wish to teach students to compute statistics, it would be best to consult a book on behavioral statistics for thorough instructions. Some resources are given at the end of this lesson plan.

A. Measures of central tendency

1. **Measures of central tendency** tell us how most people generally scored.

2. The **mean**, or arithmetic average, is the most commonly used measure of central tendency for quantitative data because it is the most sensitive measure.

   a. The mean is calculated by adding up all the scores and dividing this total by the number of scores.

b. The sensitivity of the mean can cause a problem if there are one or two extreme outlying scores compared to the rest of the sample (see skewed distributions below). When this is the case, the mean becomes misleading, and the median is preferred. This is why income is always reported in terms of the median.

3. The **median** is the score that divides a distribution in half. In other words, it is the middle score in an ordered set of scores. If the dataset includes an even number of scores, the median is determined by taking the average of the two middle scores. For example, in the dataset 1, 1, 2, 3, 4, 4, the median would be (2+3)/2 = 2.5. The median is also the 50th percentile.

4. The **mode** is the score that occurs most frequently in the distribution. Although there can only be one mean and one median in a dataset, it is possible to have multiple modes. There may also be no mode present.

B. Measures of variability or dispersion

Researchers also need measures of the spread or dispersion of scores. These indicate how different or varied scores are within the group. Here are the two most commonly used measures of variability:

1. The **range** is the distance between the highest and lowest score.

2. The **standard deviation** is an index of variability that reflects how widely distributed or clustered around the mean the scores tend to be. The standard deviation is found by calculating how far each score in the data set is from the mean.

IV. Describing Frequency Distributions

A. **Frequency distributions** and their resulting graphs often take on characteristic shapes.

1. A frequency histogram shows possible scores on the x-axis and the frequency of occurrence on the y-axis. If it is symmetrical, then the distribution on the low side mirrors the distribution on the high side. For symmetrical distributions, the mean equals the median.

ADVANCED

2. If a distribution is not symmetrical, it is said to be skewed. Data cannot be skewed; only distributions are skewed. Skewed means pulled; the distribution has been pulled—either positive or negative—from extreme observations.

    a. A **positive skew** is caused by one or a small group of unusually high value(s), where most of the data cluster near the lower scores in the distribution.

        i. The skew is named "positive" for the tail of the distribution, which is a high value relative to most other scores in the distribution. The mean is pulled in the positive direction due to the extreme high scores. The median is not affected as much and the mode is not affected at all.

        ii. An example of a positive skew is an extremely hard exam where most people do poorly but one or two students do very well.

    b. A **negative skew** is caused by one or a small group of unusually low value(s), where most of the data cluster near the higher scores in the distribution.

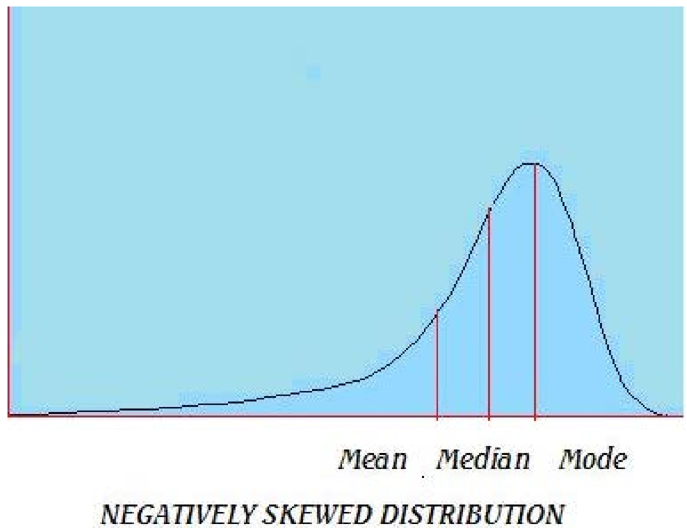
        i. The skew is named "negative" for the tail of the distribution, which is a low value relative to most other scores in the distribution. The mean in a negatively skewed distribution is pulled in the negative direction due to the extreme low scores. The median is less affected, and the mode is not affected at all.

        ii. An example of negative skew is an extremely easy test where most students do very well, but one or two still do poorly.

3. In a skewed distribution, outliers distort the mean, making the median the best measure of what constitutes a typical score in a highly skewed distribution.

    Here is an example of a negatively skewed distribution. The *x*-axis indicates scores and the *y*-axis indicates frequency of occurrence of the score:

Figure 4



B. The **normal distribution** is a specific form of a bell-shaped, symmetrical frequency distribution with mathematical properties that make it especially useful for statistics and measurement. It is sometimes called the “bell curve,” but that is misleading. While all normal curves are bell shaped, not all bell-shaped curves are normal.

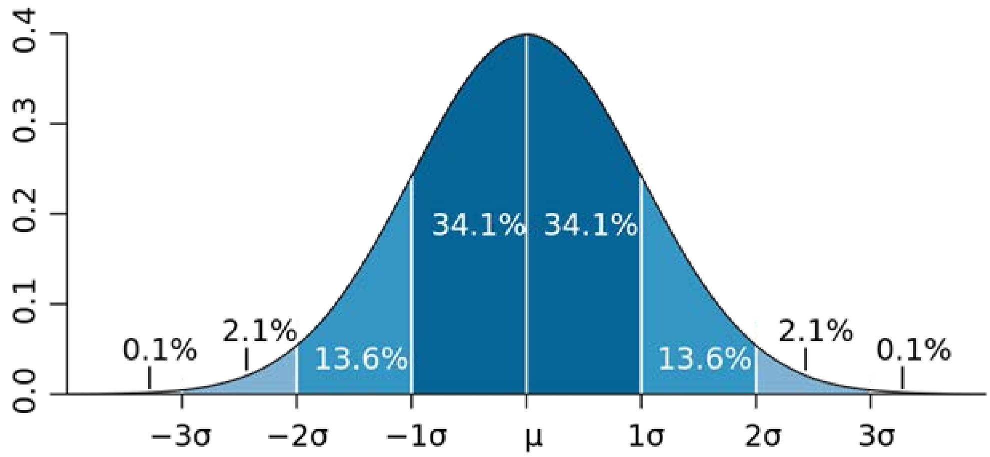
1. Like all bell-shaped and symmetrical curves, the mean, median, and mode in the normal distribution are all equal and at the highest point of frequency distribution.

2. Many physical and psychological traits approach being distributed in a normal curve. Psychological tests such as the IQ test are standardized so that the scores form a normal curve.

3. The distribution of cases between the mean and various standard deviations are the same for any normal curve. For example, 34% of data fall between the mean and the first standard deviation (either positive or negative), 14% fall between the first and second standard deviation, and 2% fall between the second and third standard deviation.

4. Here is a diagram of a normal curve, showing the distribution of cases as a function of the mean and standard deviations. Note that σ is the symbol for standard deviation. (Source: Wikimedia Commons)

Figure 5



V. Inferential Statistics

Inferential statistics allow the researcher to determine if results from the study are statistically significant and potentially generalizable beyond the research setting.

A. If a researcher concludes that the difference between groups is statistically significant, he or she is concluding that the results are not due to chance variation and that if the researcher were to conduct the research again, he or she could expect to get comparable results.

GO TO ACTIVITY 5.1

Statistical Significance

ADVANCED

B. The basics of inferential statistics

1. Inferential statistics are based on **probability**. They do not tell the researcher if a hypothesis is absolutely true or false. Rather, they allow the researcher to make conclusions based on probabilities, and there is always a probability that the researcher has made the incorrect conclusion.

2. Inferential statistics do not test the experimental hypothesis directly. Instead, they test the **null hypothesis**, which generally states that the results are due to chance factors and that the independent variable had no effect on the dependent variable.

BACK TO CONTENTS

ADVANCED

3. In addition to the null hypothesis, there is also an **alternative hypothesis** (sometimes called a **research hypothesis** or an **experimental hypothesis**) that says that the results are not due to chance—that is, the results are hypothesized to be due to the independent variable.

4. Researchers hope to reject the null hypothesis and indirectly support the alternative hypothesis.

5. If the probability is very small that the obtained results were due to chance, then the researcher will reject the null hypothesis and say the results are not due to chance but must reflect systematic differences, presumably from the independent variable.

    a. In other words, a test of significance indicates the probability of the null hypothesis being true.

    b. This is known as finding **statistically significant results**. It means the results are not due to chance and if the experiment were repeated, similar results would likely be obtained.

6. But how unlikely must it be that the null hypothesis is actually true in the population (the results are due to chance) before the researcher is willing to reject the null hypothesis? Before beginning the study, the researcher selects a level of significance that she or he will use to determine the cut-off for "unlikely." If the probability that the null hypothesis is true is less than or equal to the preselected level of significance, the null hypothesis is rejected. Otherwise, the researcher fails to reject the null hypothesis. The probability used to determine the decision regarding the null hypothesis is usually either $p = .05$ or $p = .01$. It may help students to conceptualize p as the percentage of the likelihood of the results being due to chance.

7. Note, however, that there is still a small probability that the results are due to chance and the researcher has made an error.

GO TO ACTIVITY 5.1

Statistical Significance

ADVANCED

C. There are many tests of significance that are used for different situations.

1. The most common are *t* tests and analysis of variance (ANOVA), but there are multiple kinds of each of these tests, and there are many other kinds of tests. The most appropriate test is determined by the type of research design and the number and type of variables measured.

2. No matter what kind of test is used, the test usually tells the researcher the probability that the results are due to chance, or the probability that the null hypothesis is true given the obtained results.

Critical Thinking Exercise

H. Collect some data from your class, such as favorite restaurant, number of shoes owned, amount of time spent on computer, favorite movie, highest bowling score, etc., and use the data to practice using descriptive statistics.

ACTIVITIES

ACTIVITY 1.1

Sampling or Assignment?

Allyson J. Weseley, EdD
Roslyn High School, Roslyn Heights, NY

BACK TO CONTENT OUTLINE

Sampling is when researchers select a group to study. While the word can apply to anything a researcher is studying (e.g., plants, animals), for the purpose of this exercise, we will assume we are talking about psychologists who are studying people. The word *sample* refers only to those people the researcher is studying.

Assignment, on the other hand, is the process by which researchers conducting an experiment decide which of their participants will be in each of the various treatment conditions. After picking their sample, experimenters must then assign the participants to conditions (e.g., experimental and control).

Students often confuse the terms “sampling” and “assignment.” For each of the scenarios described below, indicate which process is being discussed and how you can tell.

1. Keith exposes half of his participants to an episode of a sitcom and half to an episode of a violent television show and then observes them for signs of aggressive behavior.

2. Laurie picks 100 people to be in her study on the effects of listening to music while studying.

3. Danny picks 100 students to try a new AP Psychology text and compares them with 100 other students who are using the old text.

4. Chris puts 20 children in a drumming class and contrasts their drumming abilities with 20 children who have not had any drum instruction.

5. Tracey chooses 1,000 people to be in her study about the personalities of youngest children.

Answer Key

1. Assignment. Keith is dividing or assigning participants to groups for his study.
2. Sampling. Laurie is choosing 100 people—the sample—to be in her study.
3. Assignment. Danny has divided (or assigned) his sample of 200 students into two groups of 100 each and plans to compare them.
4. Assignment. Chris has split 40 students into two groups of 20 in order to look at the effects of drum instruction; he has assigned students to groups.
5. Sampling. Tracey has chosen (selected) 1,000 students to be her sample.

ACTIVITY 1.2

A Tasty Sample(r): Teaching About Sampling Using M&M's

Randolph A. Smith, PhD
Moravian College

This tasty demonstration exposes students to the concept of sampling and gives them a real-life sampling problem. Each student receives a small package of plain M&M's and quantifies the sample by color. Students use these data to hypothesize the population's color distribution. By pooling samples, students achieve closer approximations of the population distribution. This in-class activity is appropriate for classes in introductory psychology, statistics, and research methods. It requires about 15–30 minutes to complete, depending on discussion. Students (and faculty) find this demonstration compelling.

Concept

One concept that causes students some difficulty is sampling. Students do not always understand the need for sampling or the relation between a sample and its associated population. This knowledge is vital to understanding the research and inference process psychologists use.

BACK TO CONTENT OUTLINE

Materials Needed

Teachers will need a small package of plain M&M's for each student. You should also bring a napkin for each student. If students have calculators, the activity will be easier for them. Teachers can design a data sheet if they desire. (Note that students who are on special diets or who have food allergies may want to abstain from this activity.)

Instructions

This M&M sampling demonstration enlivens the presentation of sampling and makes it more relevant to students. Buy large sacks of fun-size packs of plain M&M's and allow each student to choose an "intact random sample" (one pack) from the population of samples. Students should sort their subjects by color, placing them on the napkin provided (much more sterile than the desk surface). Note that it is important to use plain M&M's. The peanut variety raises the potential problem of peanut allergies, but worse, they have a habit of trying to escape by rolling off the desk.

Students should make a simple frequency distribution of the six M&M colors (blue, brown, green, orange, red, and yellow) on a data sheet you will provide (see Activity 1.2 Appendix); scratch paper will also suffice. Note that it is possible that some M&M packages will not contain all six colors. You should caution your students that they are to complete their frequency distribution before any premature subject mortality occurs!

Because sample sizes typically vary somewhat (you can raise quality control as another interesting concept *and* practical application) and because you will want the students to make some inferences about the population on the basis of their sample, have them convert their raw data into percentages.

Ask each student to generate a hypothesis about the distribution of M&M colors in the population on the basis of the student's sample. These estimates generally vary considerably. Students then form pairs to pool their data (not literally, of course) and generate a joint hypothesis. Finally, we pool the data for the entire class to generate an overall hypothesis.

Discussion

Students learn some valuable lessons about sampling from this exercise. You can increase the sample size of M&M's (e.g., by using larger individual packages or 1-, 2-, or 3-lb bags) and demonstrate how larger samples typically yield better estimates of the population. Students gain an appreciation of statistics applied to real-life situations.

Because students individually generate hypotheses from small samples (usually about 24 M&M's in a fun-size pack), the hypothesized population parameters are usually low in accuracy. For example, it is not uncommon for one student to have eight of one color, say orange, whereas another student has only one orange M&M. Indeed, you will find that the bags, because the sample is so small, show considerable variability. However, as the students pair and combine their M&M's into larger samples, their estimates of the population proportions decrease in variability and more accurately approximate the population figures. When we combine the data for the entire class, variability decreases markedly, the samples become even better estimates of the population, and the hypotheses generally become more accurate.

Mars, Inc., is quite precise about the percentages of colors for the different M&M products, and the percentages are different for the various products. For plain M&M's, the current percentages are as follows: blue 24%, orange 20%, green 16%, yellow 14%, brown 13%, and red 13%. You can see those figures and the percentages for the other M&M products on the official M&M's website (http://www.m-ms.com). You should check percentages at the website before using the demonstration. When this activity was published less than 10 years ago, the percentage of brown M&M's was 30%, blue was 10%, and the other colors varied accordingly.

If you wish, you can compare the fit of your sample data to the population parameters using the chi-square statistic. I have collected large samples of data (more than 1,000 M&M's in each sample) on three different occasions. Interestingly enough, two of the three samples showed significant departure from the expected data p < .00l in each case).

Students react quite favorably to this technique, especially in light of the fact that I teach statistics immediately before lunch. I can also report that this class session is probably the liveliest of the semester. Of course, at the end of the activity you should tell your students that it is okay at that point to consume their subjects if they wish.

If you want to use a writing assignment with this activity, I suggest having the students write a letter to Mars, Inc., that describes the outcome of the class's findings. It is always challenging for students to attempt to communicate statistical findings in plain and easy-to-understand language. Such an assignment will help both the teacher and students discover whether they truly understand the concepts of sampling and of drawing inferences from samples. Students might also enjoy learning something about the history of M&M's from the company's website. It is likely that students and instructors will not know that the candies originated in 1941 and were sold to the military as a snack for American soldiers in World War II.

References and Suggested Readiing

M&M/Mars. (1993). *A little illustrated encyclopedia of M&M/Mars.* Hackettstown, NJ: Author.

Proctor, R., & Capaldi, E. J, (2006). *Why science matters: Understanding the methods of psychological research.* Malden, MA: Blackwell Publishing.

Pryczak, F. (2006). Making sense of statistics: A conceptual overview (4th ed.). Los Angeles, CA: Pryczak.

Smith, R. A., & Davis, S. F. (2007). *The psychologist as detective: An introduction to conducting research in psychology* (4th ed.). Upper Saddle River, NJ: Prentice Hall.

Smith, R. A. (2008). A tasty sample(r): Teaching about sampling using M&M's. In L. T. Benjamin Jr. (Ed.), *Favorite activities for the teaching of psychology* (pp. 8–10). Washington, DC: American Psychological Association.

Activity 1.2 Appendix

Frequency Distribution Data Sheet

Record your sample data and make a prediction of what you think the population of M&M's looks like:

	Blue	**Brown**	**Green**	**Orange**	**Red**	**Yellow**
Observed *f*						
Predicted %						

ACTIVITY 1.3

Do Cookies/Donuts Improve Memory? Errors in Methodology

R. Scott Reed, MEd
Hamilton High School, Chandler, AZ

The objective of this activity is to get students to think critically about the methodology of an experiment. I start by allowing the first half of the students to come in during the passing period and then locking the door. The students who arrived early are told to grab a cookie (or donut) and start eating before class starts. (Note that students who are on special diets or who have food allergies may want to abstain from this activity.) Since I have a window in my door, the students often will taunt the other students who are locked out. Once the final bell rings. I open the door for the others, and usually will say something like, “I wondered where everyone was.” Throughout the experiment I often will call this group the “late” and the first group the “early” group. This frustrates them even more.

I start the class and tell everyone we are doing some psychological research and we need to have a “testing environment.” I tell the class I am going to prove that the consumption of cookies increases memory in students. The students with cookies are given the list of words that are in a sentence and also the answer sheet numbered to 16. I give them 2 minutes to memorize the words. As the time comes to an end I tell them, “When you’re ready, flip the paper over and write as many words as you can remember.” I give them about 2 minutes to write the words and then have them grade their own paper.

BACK TO CONTENT OUTLINE

I then start giving the other list to the other students (no cookies) face down. I intentionally do not give them the paper to write the words down yet. Right after they start memorizing the words, I call my school phone. I feign being upset and yell into the phone. I pretend it is a friend telling me about a song on the radio, and then crank the song while the students are memorizing. I then go to the board and ask for students in the first group to yell out how many they got right.

After 2 minutes of memorizing, I tell students to immediately turn their papers over. I have not given them the paper to write the words on, so I rush around the room and "toss" the papers at them rudely. I tell them to write the numbers 1–16 before they start writing the words. Once they are done, they are told to give their papers to someone in the first group to grade. The graders are told to make sure the words are in the same form as on the list. The graders write the number correct and bring it up to me to do the statistics. The scores are written under "cookies" (which has a smiley face next to it) or "NO cookies" (frown), which has been on the board since the students walked into class.

I now look at the averages and talk about how obvious it is that cookies improve memory. I theorize that it is because of the glucose, since the brain needs glucose to run. I tell them that I am going to encourage the school to get cookies to students every day. Since the average scores are usually 15 for cookies and about 9 for the control group, I usually state that there's no question the cookies improved students' memory. At this point a lot of students start to raise their hands. I get them into small groups and tell them to come up with 10 different things that may have caused the difference in the two groups' statistics. At this point I also bring out cookies I had saved for the second group

Reasons for the Differences Between Groups

1. The cookies group was told they were going to do better.
2. The no-cookies group may have been frustrated being locked out of the room.
3. The no-cookies group may have been intimidated by hearing the other groups' scores.
4. The no-cookies group was treated poorly—called the "late" group, papers tossed at them, frown next to their group name.
5. The order of the words on the paper.
6. The distractions—phone, music, yelling of the scores.
7. Groups should never be organized so that the first half of those who arrive are placed in one group.
8. Students were told they were studying the same words, yet the no-cookies group was hearing 16s when it looked like their list only had 15.
9. The no-cookies group had a delay and then had to write numbers on their sheet.
10. The cookies group was told to turn over their papers and write the words in a more casual manner.

11. The cookies group graded their own papers.

12. The no-cookies group had words marked wrong if they had minor mistakes.

13. Since the cookies group shouted out their answers, some of them may have been lying about how many they really got right due to conformity.

Once we are done looking at the problems in the research, I introduce some aspects of statistics and hypothesis testing. I ask the students to come up with a possible hypothesis (i.e., that cookies cause improvement in memory). I then explain that statistics are used to test a hypothesis. I ask if the statistics would be the exactly the same between two groups if everything was done exactly right (random groups, same environment, same directions). Could all of the best students be randomly put into one group? I start writing numbers for the two groups (Group A–7.5 Group B–7.9) and asking if this is likely to happen. I am trying to get them to arrive at an early understanding of the numbers being so different that it is very unlikely it happened by chance, so this would then support the hypothesis. While my students are just using intuition and experience to conclude it is very unlikely one group would get 7.5 and the other 9.2, I then give some information about using statistics to describe the data and using statistics to draw conclusions or test hypotheses.

This activity is something the students will often refer to as we go through the concepts in the unit, and even at the end of the year study sessions. Sometimes we look at AP essays from prior years that include finding flaws in research, and I hear students say things like, “you cannot put all the first people in a group like we did in the cookie experiment.”

Word list (cookie group)	Word list (no cookie group)
1. Students	1. Influenced
2. Recall	2. Of
3. Ability	3. Grouping
4. Is	4. Your
5. Influenced	5. Recall
6. By	6. Together
7. Grouping	7. Area
8. Familiar	8. Students
9. Phrases	9. Phrases
10. Together	10. Ability
11. Into	11. By
12. Memory	12. Memory
13. Area	13. Is
14. Of	14. Brain
15. Your	15. Familiar
16. Brain	16. Into

Answer Sheet

1.	9.
2.	10.
3.	11.
4.	12.
5.	13.
6.	14.
7.	15.
8.	16.

ACTIVITY 2.1

Counting Fidgets: Teaching the Complexity of Naturalistic Observation

Bernard C. Beins, PhD
Ithaca College

The research tradition in psychology typically involves controlled laboratory settings. Nonetheless, naturalistic observation can generate important information. Unfortunately, most research methods textbooks devote only a single chapter to all of the descriptive techniques. This 5-min activity uses classroom observers to record fidgeting behavior and outlines a simple classroom technique that successfully conveys to students some of the complexities of naturalistic observation. [Author note: This activity can also be used to illustrate importance of operational definitions.]

Concept

Naturalistic observation can play a significant role in the study of social behaviors. However, students may not appreciate the complexities of this approach. In the activity, students count the number of fidgets they observe in classmates and discuss reasons why different observers in the same situation report different numbers of fidgets. Students then decide how, as researchers, they would solve the problems they identify.

BACK TO CONTENT OUTLINE

Materials

The only implements required for this activity are an ordinary watch, a classroom clock with a second hand or other timing device, and a sheet for tallying the fidgets in each of five 1-minute segments.

Instructions

Preparation

Solicit two student volunteers to participate in an as-yet undefined task. They need to have either digital watches or watches with a second hand, or they can borrow these items for the demonstration. It is helpful to know the volunteers because you can then select people who are likely to respond quite differently in the task; such variation enhances the pedagogical effectiveness of the activity.

Selecting one student who is energetic and another student who is calmer often leads to very different reports from the observers, which is the point of the exercise.

After choosing the two students, take them into the hallway briefly so you can explain their roles as student observers (see Activity 2.1 Appendix). Even though the directions are simple, the observers may have questions. Try to avoid answering questions about definitions of fidgets because that is part of the later discussion. In fact, I have found it is best not to let the observers ask any questions at all. Now return to the classroom.

Table 1. *Student Activity and Purpose for the 1-Minute Demonstration Period*

Minute	**Activity During the 1-Minute Period**	**Purpose**
1	I talk about a topic unrelated to systematic observation.	This generates a baseline period for number of fidgets.
2	I tell students I want them to close their eyes and imagine that insects are crawling on their skin.	This prepares them for the period in which the number of fidgets is likely to increase.
3	Students close their eyes and imagine the insects are present.	Students can concentrate on the insects without distraction. It helps generate fidgets.
4	We begin a discussion in which students speculate on the reason for the activity and the role the student observers played.	This creases a cool-down minute in which fidgets begin to decrease in number.
5	The discussion continues.	This provides another post-insect baseline.

Demonstration

The students in the class still do not know what is going on. They follow the directions as indicated in Table 1. The observation period consists of five 1-minute segments. During the observation time, the two observers are really the source of data to be discussed later. In general, the two observers will record very different numbers of fidgets within each 1-minute period across the entire time span. I have never failed to achieve notable differences between observers' counts. Often one student will record two or three times as many fidgets as the other. The discrepancy between observers illustrates the difficulty associated with monitoring a behavior as simple as fidgets. Trying to document more complicated psychological phenomena is enormously more difficult.

Discussion

Students are often not aware of the difficulties associated with naturalistic or systematic observation. During the discussion following the demonstration, ask them what could be done to improve data collection in observation studies.

I typically identify the following problems with the present methodology specifically and with observation techniques broadly; if the students do not generate these possibilities, note them and ask students to solve the problems.

1. The concept of a *fidget*, although intuitively clear, does not have a clear operational definition. Observations would be more reliable with a set definition.
   - Even though an operational definition would help, such a construct leads to missing some fidgets, whereas some movements that intuitively do not seem like fidgets would be recorded simply because of the definition used.
   - Training people until they are consistent would raise the low interrater reliability.
   - Creating a video of the scene to be recorded would allow observers to discuss their criteria so that all observers are recorded in similar ways.
2. The method of recording data might differ across observers. For example, some students log a fidget with every occurrence, taking their eyes off the class, whereas other students tally the movements in their heads and only enter them onto the data sheet when the 1-minute segment ends. Students in the latter group are less likely to miss movements while recording data.
3. The student observers may be sitting on different sides of the room, so their vantage points are not the same. As a result, they may not really be recording the same scene because of the possibility of partially blocked viewing conditions or differing perspectives.
4. Students in the class know they are being observed, even if they do not know the purpose. As a result, they may try to figure out the purpose and change their behaviors either to be helpful or to resist intru-

sive observation of their behaviors. Depending on the student's conclusion, that individual's behavior may not resemble that of the person in the next seat.

This activity is well received by students and generates meaningful discussion. Afterward, they are better able to recognize the pitfalls that arise during even simple observational techniques and appreciate the difficulties inherent in this approach.

Writing Component

Prior to the discussion of problems associated with observational research, students can generate their own list of pitfalls and the means to solve them. As a rule, any single student can produce a few of the problems, but the class discussion extends the listing greatly. As a final writing exercise, students can try to identify some of the strengths of observational research compared to experimental research. To complete the picture, they can also identify some of the limitations of the controlled experimental approach.

Suggested Reading

Babbie, E. (1995). *The practice of social research* (7th ed.). Belmont, CA: Wadsworth.

Goodwin, C. J. (1995). *Research in psychology: Methods and design.* New York, NY: Wiley.

Judd, C. M., Smith, E. R., & Kidder, L. H. (1991). *Research methods in social relations* (6th ed.). Fort Worth, TX: Holt, Rinehart, & Winston.

Activity 2.1 Appendix

Directions to Student Volunteers

I would like you to record the number of fidgets that the students in the class emit for a 5-minute period. Break the 5-minute period into separate 1-minute segments and keep a count of the number of fidgets in each segment. Keep a written record of the number of fidgets in each segment. You will need to scan the entire class, so sit at the front, facing the class.

When we go back to the classroom, take your observation seats and when I say “begin,” start recording the number of fidgets. For the first minute, I will be talking, for the second minute, I will explain that I want the students to sit with their eyes closed and imagine that insects are crawling on their skins. During the third minute, they will actually sit there with their eyes closed, imagining insects. During the fourth minute, they will begin a discussion of what they think is going on. The discussion will continue into the fifth minute.

Make sure you keep track of the time as accurately as you can and record the fidgets separately for each minute.

ACTIVITY 3.1

Pattern Recognition

Jane Halonen, PhD, and Cynthia Gray, PhD

One of the first decisions that psychology researchers must make is *how* to study their target behavior. *Correlational methods* involve studying behaviors as they are. These approaches involve careful observation, measurement, and interpretation of behaviors to uncover relationships among the factors that influence behavior. Correlational methods can be employed in case studies, surveys, and field work.

In contrast, *experimental methods* involve procedures that manipulate the conditions surrounding the behavior being studied. In effect, the experimenter interferes with normal behavior to narrow down the causes of behavior. By establishing equivalent conditions and systematically varying a specific factor, the experimenter determines the impact of that factor. This is the chief advantage of experimentation; experimental methods allow us to determine cause–effect relationships in a way that correlational methods cannot. Experiments can be conducted using a single subject or many participants. The key to determining cause–effect conclusions is providing good controlled conditions and making accurate comparisons across conditions.

BACK TO CONTENT OUTLINE

Read the following examples and judge whether the procedure described would be considered a correlational (noninterfering) approach or an experimental (interfering) approach.

The Problem of Child Abuse

Social scientists study the backgrounds of children who have been assigned to foster care. They discover that the majority of children who receive foster care have experienced physical punishment methods in their prior homes that would be severe enough to qualify as abusive.

Would this conclusion be derived from a correlational study or an experiment? Why?

BACK TO CONTENTS

Charlie's Weight Loss

Charlie is trying hard to lose his "spare tire" before his wedding. He has 2 months to get ready for the event. He decides to go about his weight loss systematically. The first week, Charlie exercises vigorously. The second week, he gives up meat. The third week, he drinks large amounts of water. The fourth week, he eats just bananas. He continues to vary his approach each week. At the end of his 8-week experiment, he concludes that his best weight loss comes from exercise.

Would Charlie's conclusion be derived from a correlational study or an experiment? Why?

The Curious Teacher

Ms. Tucker decides that she wants to evaluate which of her teaching methods might make the biggest impact on her students. For the first half of the semester, she teaches using a lecture format. She evaluates what students have learned using a 50-point multiple choice test. For the second half of the semester, she teaches using demonstrations and active learning exercises. She evaluates the second half using a 50-point multiple choice test. She discovers that her students have better test scores when using active learning strategies.

Would Ms. Tucker's conclusion be derived from a correlational study or an experiment? Why?

The Lucky Pen

Peter noticed that every time he used his special pen with green ink, something wonderful would happen to him. First, he got an A on his history exam. When he loaned his pen to Amy, she later agreed to go out on a date with him. While he was carrying his pen, he bought a lottery ticket that later won $2,000. He decided that his pen truly had lucky powers and he put it away so he wouldn't
lose it.

Would this conclusion be derived from a correlational study or an experiment? Why?

The Best Neighborhood

The local newspaper publishes the results of a finding that produce a significant impact on the real estate market. Their researchers identify various neighborhoods in the city and compare the SAT scores of students who live in different regions of the city. Based on these comparisons, they decide that the suburb of Suffolk has the best education. As a consequence, the real estate market in Suffolk booms.

Would this conclusion be derived from a correlational study or an experiment? Why?

BACK TO CONTENTS

Copycat Crime

A new popular movie depicts a grisly scene of teen violence against senior citizens. Although the criminals in the movie are caught and punished, city officials are dismayed to discover that similar acts of violence increase across the city shortly after the movie opens. A reporter decides to compare crime rates before and after the movie to determine if the movie has stimulated teen violence.

Would the reporter's conclusion be derived from a correlational study or an experiment? Why?

Doctor's Choice

A physician is intrigued by a new medication that has been developed for sleep disturbance, but he is concerned about its expense relative to the medication he usually prescribes. He decides to evaluate the success of the medication. He determines that the next 20 patients who complain of sleep problems will help him establish the effectiveness of the medication. Once he identifies the complaint, the physician slips out of the room and tosses a coin. If it comes up "heads," he prescribes the new medication. If it comes up "tails," he prescribes the medication he previously prescribed. At the end of a 3-month period, he asks his patients to report the degree of improvement on a 10-point scale. He discovers that the new medication provides no real advantage over the older, less expensive medication.

Would the physician's conclusion be derived from a correlational study or an experiment? Why?

After completing this exercise, you should be able to:

- Describe the key features of correlational research
- Explain why correlational studies cannot confirm cause-effect conclusions
- Discuss the key characteristics of experimental research
- Distinguish between scientific and nonscientific explanations
- Identify how correlational studies differ from experimental studies

Answer Key

The Problem of Child Abuse: *Correlational*

When social scientists study existing records of foster care children, they are not interfering in the normal process. They are making careful observations of the reports that describe the children's lives. In effect, they are combining many individual case studies and trying to define their common characteristics. Therefore, this approach is correlational. Although it will not produce strong cause–effect conclusions, a correlational study may provide some direction about the most important factors in the development of foster care children. Also, child abuse is an area in which it is extremely difficult to do experimental research. Scientists would be ethically opposed to constructing controlled conditions in an experiment that might place any child at risk for physical abuse or merely not exercise sufficient protection of children.

Charlie's Weight Loss: *Experimental*

Charlie's ambitious weight loss plan represents a good experimental approach to understanding weight management if he carefully imposes the limitations he describes from the outset of his plan. In effect, his approach simulates something called a single-subject design. In this case, Charlie's eating behavior is "interfered with" in a systematic fashion. He carefully records his progress and then is able to make a reasonably sound conclusion about the relative effectiveness of one method over another. This advantage assumes that Charlie is successful in abiding by the conditions he has established.

The Curious Teacher: *Experimental*

Ms. Tucker's strategy for evaluating teaching methods represents a good classroom experiment. There are some interesting problems with her design that threaten the validity of her conclusion. For example, she may be an advocate of active learning and may inadvertently construct her final exam to be easier than the midterm. Experimenter effects of this kind threaten the validity of conclusions. However, she does propose a controlled comparison, which would qualify as an experimental approach.

The Lucky Pen: *Neither Experimental nor Correlational*

Peter's observation that his pen is lucky is a good example of superstitious behavior. Human beings are inclined to reach unfounded conclusions of this nature because our brains are geared to find patterns even where patterns do not exist. Although this description sounds like it might be correlational because it does not show interference by a scientist, it does not qualify as correlational because it does not take a systematic approach to the problem. The observations are haphazard, and there is no attempt to quantify the relationships.

The Best Neighborhood: *Correlational*

Because the researcher is working with existing records, a careful comparison of SAT performance with geographic location represents a good correlational approach. The researchers are not interfering with the process but merely reporting existing relationships. However, as often happens when nonscientists try to make sense of the data, the researchers go too far by pronouncing the superiority of one school district over another. There are simply too many variables (e.g., parental income, access to computers) that might explain the superiority of one sector over another. This is the risk with correlational approaches: We cannot be confident in the cause–effect relationships that they suggest.

Copycat Crime: *Correlational*

It may be tempting to conclude that watching a violent film can induce violent behavior, but we do not have sufficient controlled evidence in this case to make that pronouncement. The reporter is not manipulating conditions in this situation (no interference), so the interpretation would be based on correlational methods. The relationship established would be that teens who see the violent example in the film would be more likely to commit a similar violent act than teens who had not seen the film. Obviously, we would be loath to construct an experiment to narrow down the cause–effect relationship suggested by this finding.

Doctor's Choice: *Experimental*

This interesting physician appears to be embarking on an experimental approach to evaluate the effectiveness of a new expensive medication. He does many things right, including random assignment to conditions by using a coin toss. However, no matter how good his intentions are, the physician does violate ethical protocol for the protection of research participants. His patients have the right to know that they are being "experimented" upon and that there are other treatment options.

ACTIVITY 4.1

Research Ethics

Allyson J. Weseley, EdD
Roslyn High School, Roslyn Heights, NY

All research involves ethical considerations. Such concerns do not mean that the research is unethical but rather that the researcher must do whatever she or he can to minimize ethical risks. Institutional review boards (IRBs) look over research proposals to safeguard participants and researchers. There are few hard-and-fast rules about what is and is not acceptable. IRBs generally engage in a kind of cost–benefit analysis.

Common ethical concerns include:

1. Informed consent

- People should not be forced to be into research.
- People have the right to withdraw from the research at any time with no penalty.
- There should be informed consent. If deception is involved, there should be a debriefing.
- To consent, people must be told something about the purpose of the research.

2. Anonymity/confidentiality

- The source of the data should be anonymous or kept confidential to protect people's privacy.

3. Long-term harm

- While it is acceptable to cause people minor discomfort during the research, no lasting physical or psychological harm should result from their participation.

BACK TO CONTENT OUTLINE

Instructions

Imagine you have been assigned the task of sitting on an Institutional review board (IRB) and have been asked to consider the following research proposals. Each proposal involves ethical issues. Read each proposal and answer the questions below.

Proposal One

Tyrone wants to study the impact of watching sexually suggestive/explicit television on people's attitudes toward sex. He plans to test ninth graders because he believes they are still young enough to be highly impressionable. He will solicit volunteers to come after school. Half will be assigned to watch one hour of sexually explicit clips from a cable TV show while the other half will view an hour of clips from the same show that deal with nonsexual topics. After watching the TV shows, all participants will fill out a questionnaire about the attitudes toward sex.

Questions:

What additional information might you want to know about the study in order to decide whether or not it should be approved?

What are the benefits that might result from this research? What are the potential harms?

If you were on an IRB reviewing this proposal, what would your recommendation be?

Proposal Two

Priya is interested in whether listening to music while working out makes people exercise harder. She plans to ask college students to come to the gym and run on a treadmill for half an hour either while listening to music or in silence. The dependent measure will be the number of miles run in that time period.

Questions:
What additional information might you want to know about the study in order to decide whether or not it should be approved?

What are the benefits that might result from this research? What are the potential harms?

If you were on an IRB reviewing this proposal, what would your recommendation be?

Proposal Three
Charlotte wants to research the effect of labeling students (gifted vs. struggling) on their achievement in second grade. She proposes that students in an elementary school's second grade be divided into reading groups in which ability levels (as determined by previous test scores) are evenly mixed. One group will be told they are gifted readers, another group will be told that they are struggling readers, and a third group will be told nothing at all. Charlotte theorizes that by the end of the second-grade year, the students in the "gifted" level group will outperform those in the "struggling" group on the same reading test.

Questions:
What additional information might you want to know about the study in order to decide whether or not it should be approved?

What are the benefits that might result from this research? What are the potential harms?

If you were on an IRB reviewing this proposal, what would your recommendation be?

ADVANCED

ACTIVITY 5.1

Statistical Significance

Allyson J. Weseley, EdD
Roslyn High School, Roslyn Heights, NY

One of the hardest concepts for students to grasp is the meaning of a statistically significant difference. A statistically significant difference is unlikely to be due to chance. Three factors contribute to whether a difference between groups will be statistically significant:

- the size of the difference between the group means
- the size of the sample
- the variance within the groups

While all three of these factors are reasonably intuitive, the third is more complicated.

After introducing students to the idea that a statistically significant difference is one that is unlikely to be due to chance, try the following exercise. Divide students into groups to read the example below and formulate an answer. Then, discuss the answers as a full class and try to draw out the answers to the discussion questions below.

BACK TO CONTENT OUTLINE

Example

Melissa is running a study to see if girls and boys average different amounts of participation in classrooms. She hypothesizes that girls participate more than boys and plans to observe students in various classes and record how often they raise their hands to answer the teachers' questions.

1. For the first part of her data collection, Melissa selects two boys and two girls to study. She finds that the girls raise their hands an average of 4.7 times per week while the boys only raise their hands an average of 1.3 times in the same classes. Do you think this difference is likely to be statistically significant? Why or why not?

ADVANCED

2. Next, Melissa expands her study to observe 50 boys and 50 girls. She finds that girls raise their hands an average of 3.1 times per week and boys raise their hands an average of 3.0 times per week in their social studies classes. Do you think this difference is likely to be statistically significant? Why or why not?

3. In a third data collection, Melissa studies another 50 boys and another 50 girls. This time, she finds that girls raise their hands an average of 3.2 times per week and boys only raise their hands an average of 2.4 times in their math classes. However, a closer look at her data reveals that there is tremendous variability between the participation of the students. Some students—both boys and girls—never raise their hands while others—again both boys and girls—raise their hands more than 15 times per week. How do you think this last factor—the variability in responses—affects the likelihood of the difference between girls' and boys' participation to be statistically significant?

Discussion Questions

1. Why does the small sample size in the first example increase the likelihood that Melissa's results are due to chance?

2. Why does the small difference between boys and girls in the second example increase the likelihood that Melissa's results are due to chance?

3. Why does greater variability among students' responses increase the likelihood that Melissa's results are due to chance?

4. Based on the data described above, what should Melissa conclude about her experimental hypothesis?

5. As a researcher, which of the three factors that influence statistical significance can you most directly influence?

KEY POINTS TO REMEMBER

1. Empirical research using the scientific method is a key difference between pop psychology and scientific psychology.
2. APA Ethical Principles address the need for care and concern for the participants in psychological studies, both human and nonhuman.
3. There are many research tools available in the psychologist's toolbox. An essential step in good research design is selecting the correct tool for the task.
4. Descriptive research and correlational research allow psychologists to describe and predict behaviors and mental processes, but only experimental research allows for explanation and control.
5. Psychological research generates data, and statistical analysis allows us to make sense of the data.

REFERENCES AND RESOURCES

REFERENCES

American Psychological Association. (2010). *Ethical principles of psychologists and code of conduct.* Retrieved from **http://www.apa.org/ethics/code**

American Psychological Association. (2011). *National standards for high school psychology curricula.* Retrieved from **http://www.apa.org/education/k12/national-standards.aspx**

American Psychological Association. (n.d.) *Research with animals in psychology.* Retrieved from **https://www.apa.org/research/responsible/research-animals.pdf**

General audience books about statistics, research, and quantitative literacy for background information and examples

Best, J. (2008). *Stat-spotting: A field guide to identifying dubious data*. Berkeley, CA: University of California Press.

Best, J. (2012). *Damned lies and statistics: Untangling numbers from the media, politicians, and activists*. Berkeley, CA: University of California Press.

Holland, B. K. (2002). *What are the chances? Voodoo deaths, office gossip, and other adventures in probability*. Baltimore, MD: Johns Hopkins University Press.

Huff, D. (1954). *How to lie with statistics.* New York, NY: Norton.

Books about behavioral statistics and research methods for more in-depth discussions of statistical concepts and computation as well as research methods and design

There are many textbooks on behavioral statistics and research methods in psychology that can provide sufficient technical information and examples. Here are two:

Smith, R. A., & Davis, S. F. (2005). *Introduction to statistics and research methods: Becoming a psychological detective*. Cherry Hill, NJ: Pearson.

Smith, R. A., & Davis, S. F. (2012). *The psychologist as detective: An introduction to conducting research in psychology* (6th ed.). Cherry Hill, NJ: Pearson.

ONLINE RESOURCES

TOPSS publishes teacher resources for supervising students in research for science fairs. In particular, they publish a booklet titled *Conducting Psychological Research for Science Fairs: A Teacher's Guide and Resource Manual*. This booklet explains many concepts of research methods and provides many good research examples. It is available here: http://www.apa.org/education/k12/science-fair-manual.pdf

TOPSS also provides an online module for teaching research methods, measurement, and statistics. It can be found here: http://www.apa.org/ed/precollege/topss/webcasts-modules.aspx.

TOPSS
TEACHERS OF PSYCHOLOGY IN SECONDARY SCHOOLS

American Psychological Association