Evaluating Research Summaries

Keith Millis, Department of Psychology, Northern Illinois University
Diane Halpern, Department of Psychology, Claremont McKenna College
Katja Wiemer, Department of Psychology, Northern Illinois University
Patricia Wallace, Department of Psychology, Northern Illinois University

Contents

Evidence-Based Teaching Strategies in this Resource

This booklet contains materials for students in Psychology to practice evaluating summaries that they find on the Web, in blogs, and in news summaries. Not all research summaries contain flaws but sometimes they do. It is important for students in the social sciences (and consumers more generally) to be able to critically evaluate research that they find because sometimes the claims made by authors and researchers are not warranted.

For each research summary, there is a list of potential flaws that the student can read for and identify. The answers are in another booklet called "How to Evaluate Research Summaries: The Answers".

Evaluating research summaries in this booklet has been shown to increase evaluation skills (Millis, Forsyth, Wallace, Graesser, & Timmins, 2019; Millis, Forsyth, Wiemer, Wallace, & Steciuch, 2019). In Millis et al., (2016), participants were presented with research summaries (many of which are in this booklet) and were asked to identify flaws. In the experimental conditions, students in research methods courses read the flaws and identified them using pen and paper or in a context of a computerized game. In a control condition, students were not shown the summaries. Based on evaluating other summaries before and after the intervention, there was significantly more learning in the experimental conditions than in the control condition which did not show evidence of learning (Partial Eta Squared = .63), and the presence of game-like features had minimal impact.

References

Millis, K., Forsyth, C., Wallace, P., Graesser, A. C., & Timmins, G. (2016). The impact of game-like features on learning from an intelligent tutoring system. *Technology, Knowledge and Learning*. doi:10.1007/s10758-019-09421-w

Millis, K., Forsyth, C., Wiemer, K., Wallace, P., & Steciuch, C. (2019). Learning scientific inquiry from a serious game that uses autotutor. In K. Millis, D. Long, J. Magliano, & K. Wiemer (Eds.) *Deep comprehension: Multi-disciplinary approaches to understanding, enhancing and measuring comprehension* (pp. 180-193). Routledge.

To the Instructor and Student:

In this booklet, there are 16 summaries of research conducted in Psychology and Biology. The summaries are not based on actual research, however. They were written by Psychologists to let students practice evaluating research that they find on the Web. The summaries represent how research may be summarized in Web articles, blogs, posts, etc.

There are many factors related to evaluating research. Based on feedback from scientists in different fields, we have pruned the number of factors down to 12 "flaws". Each flaw represents a different way that the research might be limited or flawed. The flaws are:

- No control or comparison group
- No random assignment
- DV (dependent variable) is not reliable, precise or accurate
- DV is not objectively scored (or subjective scoring without more than one scorer)
- DV is not valid
- Problems with Participant bias
- Problems with Attrition
- Small sample size
- Poor sample selection
- Problems with Experimenter bias
- Premature generalization of results
- Confuse correlation with causality

We start this document with a summary of each of the flaws titled "Common Flaws in Research". We then include a practice summary followed by the answers so you can check if you are on the right track.

The most helpful advice that we have is: if the summary does not contain information regarding whether a particular flaw is present, then do not list that as a flaw. So, if an author does not mention sample size, then that would not be a flaw. Of course, this can get tricky at times. For example, if the research describes a correlational study, then the flaw "No control or comparison group" is not a flaw, because that flaw is reserved for when researchers make a causal claim that would normally require a control or comparison group. We acknowledge that there are a few "grey" areas. That is, one may disagree with a few of the answers; however, we think that discussions around the provided answers (in another document) are fruitful for learning.

We hope you have fun reading and evaluating the research summaries. It is important that we all learn to carefully evaluate information that we find on the Internet, so that we can combat the harmful nature of intentional or unintentional misinformation.

Notes: These materials are provided free of charge. Please use them with your students and colleagues and forward them to anyone who might find them useful. Always use the citation showing our names as authors and agree to never sell or make money in any other way with these materials. We hope you will find them useful and that your students will enjoy reading and learning.

Some of the summaries also appear in a web-based program "Operation Aries!" which can be assessed at: https://app.skoonline.org/aries/text/ListOfCases.html.

Note: The summaries and Operation Aries was a result of a grant (R305B070349) awarded to Northern Illinois University from the Institute of Education Sciences, U.S.

Common Flaws in Research

Flaw: No control or comparison group

A control group or comparison group is similar to the experimental group, but it does not receive the manipulation or treatment. A control group is compared to the experimental group(s) to determine if the manipulation (or treatment or intervention) had an effect on the experimental group. The key point to remember is that in a good experiment, the treatment is varied for the different groups, but everything else in the study is held constant. **If the two groups were treated similarly except for the manipulation, the probability that the manipulation caused the results is increased.** For example, if we wanted to test the effectiveness of Drug X on headaches, the treatment group would get Drug X and the control group would get a sugar pill (or placebo). A control group in this situation would be treated identical to the experimental group except that it would get a sugar pill to ensure that the effect of Drug X in the experimental group was due to the drug and not just because the participants took a pill and thought they were supposed to feel different. Results of the experimental group would be meaningless unless there is a control or comparison group. Furthermore, a control or comparison group is a necessary component for making a causal claim.

Flaw: No Random Assignment

Random assignment is a procedure for assigning participants to the experimental and control groups so that each participant has an equal chance of being in the different groups. It is an important way of ensuring that the variability of participants in all groups is similar at the beginning of the experiment. If that's the case, then any differences between the groups on the dependent variable would be due to the independent variable (e.g., treatment, manipulation) instead of pre-existing differences (sometimes referred to as inherent differences) between the groups. If the experimental and control groups do differ in a systematic way, results could be biased and therefore unusable. For example, if a researcher put all participants who show up first to an experiment in an experimental condition and participants who show up later in a control group, and if the results show a difference between the conditions, the difference might not be due to the manipulation in the experiment. Instead, the results may be due to the systematic way in which the groups may differ, in this example, it might be the participants who show up on time may be more conscientious than participants who show up later. The level of conscientious might affect the results. Random assignment is very important because it helps experimenters to make causal claims about how the independent variable (e.g., treatment, manipulation) affects the dependent variable. Without a control group and random assignment, a causal claim cannot be made. **Random assignment refers to a procedure in which participants have an equal chance of being in the control and experimental groups. By giving participants an equal chance of being in either group, the original characteristics of the participants are spread out equally across the conditions, creating similar groups.**

Flaw: Participant Bias

Participant bias occurs when the participant's actions or behavior influence the results of a study in a way that was not intended. One type of participant bias occurs when participants act differently because they know that they are being studied. For example, if you advertise in the school newspaper that the psychology department is conducting a study on aggression in college students, it is more likely that college students will come in prepared to be more or less aggressive because they know this is what you are looking for. Human participants often try to "figure out" what the experiment is about or the goal of the experimenter. By doing so, they will pay attention to cues in the environment, which is called demand characteristics (e.g., the behavior of the experimenter, the materials or stimuli). They may even try to "please" the experimenter by acting in a way that they think they should. When these situations occur, the outcome of a study might be affected. **Participant bias refers to any influence the participants have on the results of a study that is not due to the manipulation alone. Because demand characteristics are cues that bias a participant, it is sometimes hard to know if the results of a study are solely due to the manipulation.** One way to reduce the effects of participant bias in an experiment is to make it "blind," which means the participants do not know the true purpose of the study and which manipulation they are receiving.

Flaw: Experimenter Bias

Experimenter bias refers to when the researcher, knowingly or unknowingly, provides subtle cues that would tell the participant how to behave. For example, let's say a researcher wanted to study the effect of showing emotionally charged videos on participants' moods. Now let's say that while testing participants, an experimenter acted cheerful prior to the viewing of a video that depicted happy events but acted depressed prior to the viewing of a video that depicted sad events. The way in which the experimenter acted in this example could possibly cue the participant to which type of video they were about to watch. This would be one type of experimenter bias. One way to reduce the effects of experimenter bias in an experiment is to make it "blind," which means the experimenter does not know what manipulation the participants are receiving. The experiment could also be "double-blind," which means both the experimenter and the participants do not know what manipulation the participants are receiving. Double-blind experiments eliminate both experimenter and participant bias.

Flaw: Small sample size

The sample size refers to the number of participants in a study. The participants can be people, animals or any other object that is being measured. If the sample size is small, then the results may not generalize to the larger population of interest. In statistics, the bigger the sample, the closer the average of the sample will be to the population average. Consider an experiment where a researcher is testing whether a new drug reduces anxiety. Imagine that only two people were in the experimental group (i.e., takes the drug) and only two were in the control group (i.e., takes the placebo). If you found a reduction in anxiety, how sure would you be it was due to the drug? You probably would not be very sure. The reason is that people differ on a number of things, including anxiety (and potentially other inherent differences). When scores differ among participants, it is called "variability". That is, there is variability among people and you have to decide whether the variability you found was just normal variability or variability due to the drug (independent variable). With only two people being compared in each condition, it is hard to know if the variability in their scores would have occurred even without the drug manipulation. However, if you randomly assigned 50 people to each group, and you found a difference between the averages of the two groups, then you would be more certain that it was due to the drug. The reason why is that with larger sample sizes, we are more confident that the results will be closer to the "true" score, which allows us to generalize the findings to the population. **There is no magic number of what makes a "good sample size" but we know that when there is more variability in the dependent variable (the scores vary from participant to participant), one needs a larger sample size. Generally speaking, though, one should have at least 25-30 or so participants in each condition.**

Flaw: Poor Sample Selection

The group of participants in an experiment is called a sample. Ideally, the sample should be representative of the population from which it was drawn. By "representative" we mean the individuals in the sample should have similar characteristics of the individuals in the population. By "population" we mean all individuals that the experimenter would like the results of the study to apply. If the population is "all college students," then the sample should be representative of all college students (e.g., different ages, different majors, and other inherent differences). If the sample contains only college students from one major, then it would not be representative of all college students, and the results could not generalize to all college students. **Ideally, participants should be drawn randomly from the population so that a representative sample is created.** If this occurs, we called it random selection. Unfortunately, this is usually not possible (just think about getting a random sample from college students from across the world!), but experimenters try their best to get a representative sample. Poor sample selection occurs when there is some obvious way that the sample may be different from the population. One way this occurs is if the participants that participate in the study are somehow different from other participants from the population. For example, if a researcher puts an advertisement for an experiment on twitter, then only participants with twitter accounts could see the ad and respond. Perhaps people who have twitter accounts could differ from people without twitter accounts (e.g., they could be younger). In this way, participants unknowingly "selected themselves" to take part in the study because

they have twitter accounts. People with twitter accounts may not be representative of all people. **Any time the sample may not be reflective of the population of interest because of how they were selected, we say that it was due to poor sample selection. Poor sample selection is important because it limits the generalizability of the findings.**

Flaw: Attrition

For researchers, **attrition** means participants dropped out of the study for some reason. When a researcher loses participants, it means that they can no longer use their data because the participant did not complete the study. Attrition may occur if the experiment occurs over hours or days. For example, if a study has two sessions, one on day 1 and another on day 2, and if someone does not show up on day 2, then this would be attrition. Attrition can become a flaw in research when something about the experimental procedures causes participants to drop out for systematic reasons. For example, if people did not show up for day 2 because they thought the study was boring, then the data might be biased because it only included people who were motivated enough to show up for day 2. These participants may be different than those who dropped out. **In general, attrition is a problem if it occurs unequally across the groups and when participants with extreme scores drop out. When participants with extreme scores drop out, then the people left in are no longer representative of the larger population. Attrition can make the groups unequal and therefore affect, or bias, the results.**

Flaw: DV is not reliable, precise, or accurate

Reliability of a measure is the consistency with which it measures something. For example, if you wanted to know how much you weigh and you got on a scale that gave you a different number each time (even when you weigh yourself seconds apart and you are sure your weight really did not change), the scale would be considered unreliable. Reliability is important to providing a valid conclusion from the results of a study. If a measure is unreliable it is difficult to determine what impact, if any, the independent variable had on the results. Another key concept, accuracy, means that a measure is interpreted correctly. Accuracy determines reliability. Precision is also important when measuring for experimental purposes; one must make sure that the units are small enough for what you are studying. For example, you wouldn't measure how far a slug moves an hour in feet when inches would be a more appropriate and precise unit of measurement. **Having a precise dependent variable means that your measurement is broken into many small units. With smaller units you are abler to measure small changes, making it a "sensitive" measure. Without a precise measure, you may miss differences between conditions because the measure is not sensitive to changes in the DV. You cannot have reliability without accuracy. A reliable DV is one that is consistent. Reliable dependent measures help experimenters make the claim that their measures are valid, and their results are due to what is being manipulated.**

Flaw: DV is not valid

It is important that dependent variables really measure what we want them to measure. If they do, then we say that they are valid. If they do not, then we say that they are invalid. For example, if we want to measure friendliness, then we would need to make sure that we measured friendliness correctly—such as the number of friends one has. If instead we measured friendliness by asking people about their happiness, then that would be an invalid measure since happiness is not the same as friendliness. It might be true that people who are friendly are happier than unfriendly people, but it is still a different variable. As another example, at one point in time people thought that the bumps on their skulls indicate a person's personality and ability. This was called Phrenology. We know now that bumps on the skull have no correlation with personality and abilities. Therefore, measuring bumps on the skull is an invalid way of getting at personality and abilities. Having a valid dependent variable is important because if it is invalid, then it is not testing the experimenter's hypothesis. **A measure is said to be valid if it measures the variable that the experimenter intends to measure.**

Flaw: DV is not objectively scored (or subjective scoring without inter-rater reliability)

When possible, a researcher wants to make sure to score the dependent variable objectively. **Objective scoring refers to when an experimenter does not need to make a judgment of the quality of a variable when scoring data**. For example, if an experimenter had a multiple choice test that was graded with an answer key—it would not matter who graded the test as long as they are careful. This is true because for a multiple choice test there are specific answers that are required and the answers should not vary. On the other hand, subjective scoring occurs when the judgment of the scorer is required. For example, when grading essay tests, though there may be certain aspects the essays have in common, scoring is much more open to interpretation. So if one experimenter grades the same essay as another experimenter, even if they are both very careful, they could come up with different scores. With subjective scoring, there is always the possibility that the bias of the scorer (or rater) will affect the scores. However, it should be noted that subjective scoring is not bad when there are more than one rater and that the raters agree on their judgments. **Because subjective scoring requires a scorer to make a judgment about a participant's behavior or some outcome, a single scorer needs to be careful not to bias the judgment that is made. It is typically better that the dependent variable is objectively scored than subjectively scored if possible in the study design.**

Flaw: Premature Generalization of Results

In good research, there is usually some acknowledgement of whether more data or research is needed to make a strong claim about the findings. It is important that researchers replicate findings so that we are more confident that the results will generalize to people who were not involved in the study. So, often research summaries say something about how the results will or will not generalize to other participants, settings, and related measures. **If a researcher makes a generalized claim based on the results of only <u>one</u> study, then the researcher commits the flaw of premature generalization of results.** If a researcher notes that the research needs replication or that it is being replicated, then this would not be a flaw.

Flaw: Confuse Correlation and Causality

Experiments are typically done to see if one thing causes another (e.g., whether a pill reduces hunger; a new technology increases learning; changes in temperature cause changes in mood, etc.). For a researcher to claim that A causes B, three things must be established: 1) A and B must be correlated, 2) A occurs before B, and 3) all other possible causes for the correlation are ruled out. Often people (sometimes even scientists!) confuse correlation with causality because they don't take into consideration the other two requirements. For example, it has been shown that post-menopausal women who take hormone replacement therapy are healthier than those that do not take the therapy. This in itself shows that health and undergoing hormone replacement therapy are correlated. To show a causal relationship, however, researchers must rule out other explanations for the correlation (condition 3). After doing more research, scientists have found that women who are on hormone replacement therapy are "different" from the women who did not use the therapy. More specifically, women who use the therapy are more likely to visit the doctor (perhaps this only includes women who could afford to see a doctor) and follow their doctor's instructions more closely than the women who do not use the therapy. It turns out that these and other differences accounted for the correlation. Consequently, the hormone replacement therapy was not shown to cause an increase in health. This demonstrates how one variable being correlated with another is insufficient to show that one variable causes the other variable. The phrase that some use for this is "correlation is not causation." In order to make causal claims, one needs random assignment to condition and other controls (such as a double-blind design) to rule out other explanations for the correlation. **If someone makes a causal claim (A causes B) only on the basis of a correlation, then that person is confusing correlation and causality.** Only an experiment can determine causality.

Instructions:

In the following pages, there are several written summaries of made-up studies from various sources. The summaries contain one or more "flaws" that are described above. Your job is to carefully read each study summary and identify the flaws that are present.

This is not an easy task. You will likely need some practice. The first research summary in this document ("Textbooks in the Classroom") contains the correct answer immediately following the summary. The answers to the other study summaries can be found in the "Evaluating Research Summaries: The Answers" document. In the answers, we provide you with the flaws that we think are present, along with an explanation of why the flaw is present or not present. You will note that we inserted space to the right of the flaws for students to provide an (optional) explanation for their answer.

By the way, one of the summaries does not contain any flaws.

Good luck!!!

Textbooks in the Classroom

One thing that many students think about when they register for classes at a university is how many textbooks they are going to have to buy for the class and how much the books are going to cost. To add to this, a lot of the students wonder if they are even going to use the books that they are required to buy. In fact, some students don't buy books for their classes because they are convinced that they don't really need them to achieve an acceptable grade.

This is exactly the line of thinking that textbook writers are afraid of—they want students to have to use their books to get good grades in their classes, and they want professors to think that students need their books so that they require them as part of their classes.

Even though textbooks have a definite value—they are available to students who use them when their professors are not—there is some debate on whether they are really needed as part of university classes.

Recently, a researcher conducted an experiment to address this question. In the experiment, the researcher compared two sections of his introductory statistics course, a course required for all liberal arts and sciences students. Students who were enrolled in the fall semester of the course were told that buying the textbook was optional, whereas students enrolled in the spring semester were told that buying the textbook was required. All 380 of the students (190 in the fall and 190 in the spring) completed the course, and they all took the final exam, which consisted of some calculations and several conceptual essay questions.

When the professor finished scoring the essays, he compared the final exam grades of both sections of the class. He found just what he thought he would—there were no differences in the scores on the exams between the section that thought the textbook was optional and the section that thought the textbook was required. The average grade for the fall semester was 84.3%, and for the spring semester it was 85.2%.

Based on this study, the researcher concluded that textbooks were not necessary or helpful for learning, since there were no differences in scores between the two sections.

The flaws (You may use this list to guide your evaluation)

- No control or comparison group; No random assignment; DV (dependent variable) is not reliable, precise or accurate; DV is not objectively scored or subjective scoring without more than one scorer; DV is not valid; Problems with Participant bias; Problems with Attrition; Small sample size; Poor sample selection; Problems with Experimenter bias; Premature generalization of results; Confuse correlation with causality

Check any flaws that apply: **Textbooks in the Classroom**

____ No control or comparison group: _____

____ No random assignment: _____

____ DV (dependent variable) is not reliable, precise or accurate: _____

_____

____ DV is not objectively scored or subjective scoring without more than one scorer: _____

_____

____ DV is not valid: _____

____ Problems with Participant bias: _____

____ Problems with Attrition: _____

____ Small sample size: _____

____ Poor sample selection: _____

____ Problems with Experimenter bias: _____

____Premature generalization of results: _____

____ Confuse correlation with causality: _____

Answers on next page

Flaw	X = Flaw in study	Explanation
No control or comparison group		The comparison group used in this study was the class that was not required to use a textbook.
No random assignment	X	To make a conclusion about the effectiveness of textbooks, you must randomly assign participants to conditions. That did not happen here.
DV could be more reliable, accurate, or precise		Considering the performance of both sections, it seems that there may have been room for improvement, but this is not a flaw.
DV is not scored objectively	X	The dependent variable was an essay test which lends itself to subjective grading by the professor. Furthermore, the professor was not blind to condition. The professor should have had two independent graders score each participant's essay test.
DV is not valid		Statistics is not just computational; it also involves conceptual knowledge. Therefore, the essay test is a valid measure.
participant bias		The students had no way of knowing the intentions of the professor or the point of the study. It is unlikely that they could have influenced the results with participant bias.
mortality or attrition		There were no students that did not finish the course or that did not take the final exam. Everyone followed instructions and there were no instances of attrition or mortality.
small sample size		There seemed to be an adequate sample size here.
poor sample selection		This statistics course is required for all majors so the participants do represent a wide variety of majors. Therefore, the sample selection does not seem to be a large problem here, even though there was no random assignment and the possibility of experimenter bias.
experimenter bias	X	It is a possibility that the professor taught the two classes differently because he knew which condition each class was assigned.
premature generalization of results	X	The researcher makes a very strong conclusion about his results without replicating his finding with another study. Premature generalization of results is a flaw here.
confuse correlation with causation		Considering that the researcher manipulated whether or not there was a textbook in use between two conditions, this is not a correlational study.

Studying with Music

"As a professor, I am interested in how my students study in general, and how they study for upcoming exams. One of the things that they often tell me is that listening to music helps them understand study material better than they would if they had to study in silence," said Dr. Robinson, who teaches at a large university. He continued on to say that, "I want to find out if this is really true."

In order to test whether or not his students were correct, that they were better able to understand material while listening to music, Dr. Robinson divided his advanced music theory class of music majors into two groups of 35 students each. On the day of testing, he instructed the students who sat in the front of the classroom to read a text on ancient music (a topic unfamiliar to with the students) while music played in their headphones, and the students who sat in the back of the classroom were instructed to read the same text in silence.

After 15 minutes had lapsed, Dr. Robinson had the students write down what line they were on in the text as a way of measuring reading comprehension. Dr. Robinson argued that reading is slowed when material is hard to comprehend. Unfortunately, the professor found that 20% of the students in the music group failed to record their final line.

For the students who did follow directions, Dr. Robinson found that the students who listened to music read more lines than those who read in silence. Additionally, he found that the students who listened to music comprehended more of the passage than the students who did not listen to the music.

Because he was surprised at the results he found, Dr. Robinson ran the study again later in the term. He found the same results.

When he was interviewed about his findings, he said that, "To my surprise, my students were right—they did understand more when they listened to music than when they studied in silence. I have actually started to suggest to my students that based on my results, they may want to start listening to music when they prepare for exams."

The flaws (You may use this list to guide your evaluation)

- No control or comparison group; No random assignment; DV (dependent variable) is not reliable, precise or accurate; DV is not objectively scored or subjective scoring without more than one scorer; DV is not valid; Problems with Participant bias; Problems with Attrition; Small sample size; Poor sample selection; Problems with Experimenter bias; Premature generalization of results; Confuse correlation with causality

Check any flaws that apply: **Studying with Music**

____ No control or comparison group: _____

____ No random assignment: _____

____ DV (dependent variable) is not reliable, precise or accurate: _____

_____

____ DV is not objectively scored or subjective scoring without more than one scorer: _____

_____

____ DV is not valid: _____

____ Problems with Participant bias: _____

____ Problems with Attrition: _____

____ Small sample size: _____

____ Poor sample selection: _____

____ Problems with Experimenter bias: _____

____Premature generalization of results: _____

____ Confuse correlation with causality: _____

Spontaneous Generation Exists!

"Spontaneous generation" (SG) is the name for the idea that living things can spontaneously be born from nonliving substances in decomposition, for example, from rotting meat or fruit. Visualize an alien creature crawling out of a rotten apple, and you get the idea.

So far, SG has been treated mostly as an entertaining science fiction concept. However, that could change soon, as a recent experiment suggests that SG might actually occur for real.

Inspired by a serendipitous finding involving fruit fly traps, Gina Morris, a journalist, conducted the following experiment. She purchased twenty jars with screw-on tops. For the experimental group, she placed a piece of cooked apple and a tablespoon of vinegar in each of ten jars. For the control group, the remaining ten jars were left empty.

For a period of 24 hours, all twenty jars were left open so fruit flies could enter.

After the 24-hour period, Ms. Morris discovered that there were insects present in one of the jars, an experimental group jar with cooked apple and vinegar in it. She emptied the fruit flies out of this jar but made sure to keep the fruit in. Then, all jars were sealed with fine gauze and left sitting for three days.

When Ms. Morris examined the jars on the fourth day, she noticed right away that the jar in which she had found fruit flies four days earlier now contained new fruit flies.

From this extraordinary finding, Ms. Morris concluded that the concept of spontaneous generation definitely exists. So, in the special report that she published the following week, she concluded that living things are spontaneously generated from nonliving substances, such as rotting fruit.

The flaws (You may use this list to guide your evaluation)

- No control or comparison group; No random assignment; DV (dependent variable) is not reliable, precise or accurate; DV is not objectively scored or subjective scoring without more than one scorer; DV is not valid; Problems with Participant bias; Problems with Attrition; Small sample size; Poor sample selection; Problems with Experimenter bias; Premature generalization of results; Confuse correlation with causality

Check any flaws that apply: **Spontaneous Generation**

____ No control or comparison group: _____

____ No random assignment: _____

____ DV (dependent variable) is not reliable, precise or accurate: _____

_____

____ DV is not objectively scored or subjective scoring without more than one scorer: _____

_____

____ DV is not valid: _____

____ Problems with Participant bias: _____

____ Problems with Attrition: _____

____ Small sample size: _____

____ Poor sample selection: _____

____ Problems with Experimenter bias: _____

____Premature generalization of results: _____

____ Confuse correlation with causality: _____

Which Is Stronger—A Gorilla Or A Chimp?

Everybody loves going on field trips to the zoo. But field trips to the zoo provide more than just a day off from the classroom for kids. It provides an opportunity to learn about science!
Mrs. Young, a fifth grade teacher in our district, was especially excited to bring her class to the zoo this year because she knew that they were going to get to see a zoologist test the strength of the different primates. As Mrs. Young observed, "Not only is it interesting for the students to get to see all the different animals, but they are going to be learning about how strong the different animals are, as well."

The main research goal for the zoologist was to compare the strength of chimpanzees and gorillas with the expectation that there would be a difference. First, she trained 4 chimps and 4 gorillas to lift up different sized, light-weight objects by rewarding them with food. She then tested the animals' strength by giving them heavier objects that ranged from 25 to 100 pounds. During this testing, every time the primates lifted the weights for the food, the zoologist increased the weight.

"The children and I were surprised to learn that both the chimps and the gorillas could lift the 100-pound weight. We thought that one group would be stronger than the other, but that wasn't the case," Young said.

The zoologist conducted this experiment again, and found the same results. Therefore, she was able to make the conclusion that chimps and gorillas do not differ in strength.

The flaws (You may use this list to guide your evaluation)

- No control or comparison group; No random assignment; DV (dependent variable) is not reliable, precise or accurate; DV is not objectively scored or subjective scoring without more than one scorer; DV is not valid; Problems with Participant bias; Problems with Attrition; Small sample size; Poor sample selection; Problems with Experimenter bias; Premature generalization of results; Confuse correlation with causality

Check any flaws that apply: **Which is Stronger – A Gorilla or a Chimp?**

____ No control or comparison group: _____

____ No random assignment: _____

____ DV (dependent variable) is not reliable, precise or accurate: _____

_____

____ DV is not objectively scored or subjective scoring without more than one scorer: _____

_____

____ DV is not valid: _____

____ Problems with Participant bias: _____

____ Problems with Attrition: _____

____ Small sample size: _____

____ Poor sample selection: _____

____ Problems with Experimenter bias: _____

____Premature generalization of results: _____

____ Confuse correlation with causality: _____

Pesticides

Whether people have a green thumb or not, organic gardening has become such a hotly debated topic over the recent years and it is hard to escape the controversy over this topic.

If you have considered organic gardening and have compared organic and chemical gardening products in stores, you will have noticed that as a rule, organic products come at a higher price. So, the interesting question for consumers is what benefits do these products have for the environment as well as for us gardeners? In what ways, if at all, will the extra cost bring us a benefit?

You may be interested in a study that my neighbor recently conducted in his own garden. Now, my neighbor is both a professional biologist and an obsessive gardener – so he knows what he is doing.

He wanted to compare the effects of using organic pesticides and chemical pesticides on produce quality. I asked him his 'operational definition' of 'quality' and he said healthy coloring and nutrient value. In a controlled, greenhouse environment, he raised two beds of tomato plants for the study. Each bed contained fifteen tomato plants. On one bed, only organic pesticides were used. The other bed was treated chemically. Both beds were on identical water and light schedules.

When the thirty plants started producing tomatoes, my neighbor compiled two measures. First, he counted the number of tomatoes yielded by the fifteen plants in each bed. Second, after harvesting all ripe fruit, he weighed the tomatoes from the two beds of plants separately. Both measures were accumulated until the season came to a close.

He found that while the plants in both beds produced a similar number of tomatoes (organic: 87; chemically treated: 90), the average weight of the tomatoes was significantly higher in the chemically treated bed (124g) than in the organic culture (82g).

My neighbor concluded that the use of chemical pesticides does indeed increase the quality of produce, compared to organic gardening practices.

The flaws (You may use this list to guide your evaluation)

- No control or comparison group; No random assignment; DV (dependent variable) is not reliable, precise or accurate; DV is not objectively scored or subjective scoring without more than one scorer; DV is not valid; Problems with Participant bias; Problems with Attrition; Small sample size; Poor sample selection; Problems with Experimenter bias; Premature generalization of results; Confuse correlation with causality

Check any flaws that apply: **Pesticides**

____ No control or comparison group: _____

____ No random assignment: _____

____ DV (dependent variable) is not reliable, precise or accurate: _____

_____

____ DV is not objectively scored or subjective scoring without more than one scorer: _____

_____

____ DV is not valid: _____

____ Problems with Participant bias: _____

____ Problems with Attrition: _____

____ Small sample size: _____

____ Poor sample selection: _____

____ Problems with Experimenter bias: _____

____Premature generalization of results: _____

____ Confuse correlation with causality: _____

Can A Memory Pill Enhance Your Short-Term Memory??

No one can remember everything. Many people forget things everyday—where they put their keys, scheduling a doctor's appointment, calling a friend back. But maybe, just maybe, researchers are one step closer to helping people to solve their memory problems with a new memory pill.

Dr. Andrew Phillips, a neurologist from Bluegrass State University, thinks there is definitely a place in today's society for such a pill. He says that he believes a memory pill "could be very effective in helping people remember activities and plans they make on a daily basis." He goes on to say that if a memory pill is shown to be effective, "there would be widespread interest from the public and a huge market for the pill."

In one study, Phillips tested the effectiveness of one such memory pill. He had 49 adults complete an interview as a health check. During the interview, the participants were given 3 lists of 20 items to remember. They were immediately asked to recall the lists. All the words on the list were common, well-known objects. Their scores were recorded.

All participants were required to take the memory pill once a day for 3 weeks. To ensure that they really did take the pill each day, they were required to receive the pills directly from the researcher every day. The participants were led to believe they were taking a pill to lower their blood pressure.

Twenty-four of the participants were randomly assigned to take a sugar pill, while twenty-five took the memory pill. After 3 weeks, the participants all returned to the laboratory to complete another interview similar to the first one. Again, they were asked to recall 3 lists of 20 items each.

When the researchers calculated the scores on the recall tests for pre- and post-treatment, they found that the participants only recalled one more item in the second interview than they did in the first. Furthermore, the scores were the same for the placebo group as they were for the group that took the memory pill. These results were the same as those they got in a prior study.

Of the results, Dr. Phillips said the following: "Unfortunately, this study did not give us the results we were hoping for. It seems that the memory pill used in this study was not effective in improving participants' short-term memory capabilities." And so, the search continues for the magic pill that will improve our memories.

The flaws (You may use this list to guide your evaluation)

- No control or comparison group; No random assignment; DV (dependent variable) is not reliable, precise or accurate; DV is not objectively scored or subjective scoring without more than one scorer; DV is not valid; Problems with Participant bias; Problems with Attrition; Small sample size; Poor sample selection; Problems with Experimenter bias; Premature generalization of results; Confuse correlation with causality

Check any flaws that apply: **Can a Memory Pill Enhance Your Memory?**

____ No control or comparison group: _____

____ No random assignment: _____

____ DV (dependent variable) is not reliable, precise or accurate: _____

_____

____ DV is not objectively scored or subjective scoring without more than one scorer: _____

_____

____ DV is not valid: _____

____ Problems with Participant bias: _____

____ Problems with Attrition: _____

____ Small sample size: _____

____ Poor sample selection: _____

____ Problems with Experimenter bias: _____

____Premature generalization of results: _____

____ Confuse correlation with causality: _____

Heavy Metal Music: A Teenager's Perspective

Like a lot of other kids my age, I like to listen to heavy metal music. Like a lot of other parents, mine complain about the type of music I listen to…..all the time. They think that my listening to heavy metal music is going to make me sad and depressed. I always tell them they're wrong—my friends and I all listen to heavy metal music, and we are some of the happiest people I know!

I knew that the only way to get them to back off was to prove to them that they were wrong, so I did some research. I found the one and only article that seems to be out there about the effects of heavy metal music on mood.

In this study, 80 participants were randomly put into one of two groups. In one group, the participants were told that they could not listen to heavy metal music for one whole month. In the other group, the participants had to listen to heavy metal music for two hours every day for a month.

After the month was over, everyone had to fill out a survey with a lot of different rating questions. There were so many different questions that no one who participated could have guessed that depression was being studied.

The main question that the researchers were interested in was how the participants rated their anxiety/emotional level. So the key question was "What is your current emotional state," and the participants answered by circling either "anxious" or "not anxious" (these were the only options given). The researchers were unaware of which condition the participants were in when they did the final data analyses.

The researchers found that 60% of the participants who listened to heavy metal music reported feeling anxious, as did 58% of the participants who did not listen to the music at all. This was not a statistically significant difference.

So, what does all this mean for me? Well, it means that even though my parents admitted to me that I was right—I'm not going to become depressed from listening to heavy metal music—they still don't want me to listen to it. But, the take-home message for all the other teens and their parents out there is this: Listening to heavy metal music is not going to lead to depression, so listen to it as much as you like!

The flaws (You may use this list to guide your evaluation)

- No control or comparison group; No random assignment; DV (dependent variable) is not reliable, precise or accurate; DV is not objectively scored or subjective scoring without more than one scorer; DV is not valid; Problems with Participant bias; Problems with Attrition; Small sample size; Poor sample selection; Problems with Experimenter bias; Premature generalization of results; Confuse correlation with causality

Check any flaws that apply: **Heavy Metal Music: A Teenager's Perspective**

____ No control or comparison group: _____

____ No random assignment: _____

____ DV (dependent variable) is not reliable, precise or accurate: _____

_____

____ DV is not objectively scored or subjective scoring without more than one scorer: _____

_____

____ DV is not valid: _____

____ Problems with Participant bias: _____

____ Problems with Attrition: _____

____ Small sample size: _____

____ Poor sample selection: _____

____ Problems with Experimenter bias: _____

____Premature generalization of results: _____

____ Confuse correlation with causality: _____

Gender and Aggression

Aggression is a behavioral response that many people show on a regular basis. People yell, shout, push and fight. Often times, people perceive that men are more aggressive than women, perhaps because this is what is portrayed most often in the media.

Knowing that both men and women are capable of having aggressive tendencies, Dr. Maye, a member of a research institution which focuses on aggression, wanted to find out if one gender is really more aggressive than the other.

To conduct his experiment, he placed an ad in a newspaper and asked for volunteers to participate in a study that was going to explore gender differences in aggression. All interested people were asked to report to the institution conducting the research.

Thirty men and thirty women volunteered to participate in the study. When they arrived at the study, they were exposed to multiple situations that were supposed to elicit aggression (an accomplice posing as a participant was used to provoke the actual participants by gently pushing them away while saying something rude). After they were put in this situation, the participants were given the opportunity to write a message to the person who provoked them.

The messages that participants wrote were coded by two independent researchers who were not aware of the participant's gender. The messages were coded on a 7-point scale for the degree of verbal aggressiveness that was used. The results showed that women provided more aggressive messages than men.

In a follow-up study, the researchers found the same results—women were found to be more verbally aggressive than men. The researchers used these results to draw the conclusion that, contrary to popular belief, women are actually more aggressive than men.

The flaws (You may use this list to guide your evaluation)

- No control or comparison group; No random assignment; DV (dependent variable) is not reliable, precise or accurate; DV is not objectively scored or subjective scoring without more than one scorer; DV is not valid; Problems with Participant bias; Problems with Attrition; Small sample size; Poor sample selection; Problems with Experimenter bias; Premature generalization of results; Confuse correlation with causality

Check any flaws that apply: **Gender and Aggression**

____ No control or comparison group: _____

____ No random assignment: _____

____ DV (dependent variable) is not reliable, precise or accurate: _____

_____

____ DV is not objectively scored or subjective scoring without more than one scorer: _____

_____

____ DV is not valid: _____

____ Problems with Participant bias: _____

____ Problems with Attrition: _____

____ Small sample size: _____

____ Poor sample selection: _____

____ Problems with Experimenter bias: _____

____Premature generalization of results: _____

____ Confuse correlation with causality: _____

"Pounds Off" Helps People Lose Weight

Do you need to lose a few pounds? Are you struggling to lose weight quickly?

Well, if so, this may not be true for much longer. New research has shown that you may be just a few months away from having the trimmer figure that you've always wanted.

So, you may be wondering, what is this new pill? It's called "Pounds Off," and it was scientifically tested for its effectiveness by a group of independent researchers.

Thirty people were randomly selected to participate in a study about the side effects of a medication. They were all told that the study would last for 3 months and that they would have to have a full physical at the beginning and end of the 3 months. Their weight was recorded each time by a doctor who was unaware of the purpose of the study.

Participants had to take a pill every day during the 3-month period. Immediately after they took the pill, they had to call or email the researcher to let him know; however, 13 out of the original 30 participants failed to do this. Those participants were dropped from the study because the researchers claimed that the effectiveness of the pills could not be tested if they were only taken sometimes or not at all.

The researchers for "Pounds Off" found that, at the time of the second physical, the participants who completed the study had lost an average of 25 pounds. At the end of the study, one participant, Megan Smith, said, "I tried 'Pounds Off' and it really worked for me! It seems like I tried every diet pill on the market before this, and nothing worked, but this time I actually lost weight. 'Pounds Off' is definitely one diet pill that works!" It should be noted that the results of this study were successfully replicated with another group of participants.

So, if you are ready to shed some weight, "Pounds Off" may be the right pill for you because it has clearly been shown to help healthy adults lose a significant amount of weight. What have you got to lose by giving it a try (besides a few pounds, of course)?

The flaws (You may use this list to guide your evaluation)

- No control or comparison group; No random assignment; DV (dependent variable) is not reliable, precise or accurate; DV is not objectively scored or subjective scoring without more than one scorer; DV is not valid; Problems with Participant bias; Problems with Attrition; Small sample size; Poor sample selection; Problems with Experimenter bias; Premature generalization of results; Confuse correlation with causality

Check any flaws that apply: **Pounds Off Helps People Lose Weight**

____ No control or comparison group: _____

____ No random assignment: _____

____ DV (dependent variable) is not reliable, precise or accurate: _____

_____

____ DV is not objectively scored or subjective scoring without more than one scorer: _____

_____

____ DV is not valid: _____

____ Problems with Participant bias: _____

____ Problems with Attrition: _____

____ Small sample size: _____

____ Poor sample selection: _____

____ Problems with Experimenter bias: _____

____Premature generalization of results: _____

____ Confuse correlation with causality: _____

Let's Dance

Stephanie Webber, a dance instructor at a local college, is always looking for new ways to help her dance students improve their techniques. One way she thought to do this was to show her students a dancing video that she strongly believed would help them improve their dance ability. The video showed dance performances on the popular TV show "Dancing with the Stars."

During the first week of a new dance schedule, she asked students who were interested in helping her with a project to join her that weekend at the studio. Ten of her students were able and willing to join her that weekend.

Without telling them what she was doing, she measured each of the participant's dancing ability using a test in which she videotaped them dancing alone while holding a broom. While sounding a bit funny, the test has been validated in a number of studies using blind scoring. The scorer uses an objectively designed coding scheme to code the dancer's movements using slow motion and pausing of the videotapes.

Prior to this study, Webber was trained to use the coding scheme. She coded each dancer's videotapes as they performed the "broom test."

Ten students completed the broom test and then watched the "Dancing with the Stars" video. According to post-experimenter questionnaires, they were all unaware that the video was supposed to help them improve their dancing ability. After they watched the video, they took the broom test again while being videotaped. Webber collected the tapes and coded them herself.

The dance instructor found that her students' dancing ability was significantly better after they watched the video than it had been before. In order to be certain that the video was truly effective, she repeated the study with another ten dance students and found the same thing. As a result, she felt confident concluding that watching the dancing video will indeed improve the dancing skills of students in her classes.

The flaws (You may use this list to guide your evaluation)

- No control or comparison group; No random assignment; DV (dependent variable) is not reliable, precise or accurate; DV is not objectively scored or subjective scoring without more than one scorer; DV is not valid; Problems with Participant bias; Problems with Attrition; Small sample size; Poor sample selection; Problems with Experimenter bias; Premature generalization of results; Confuse correlation with causality

Check any flaws that apply: **Let's Dance**

____ No control or comparison group: _____

____ No random assignment: _____

____ DV (dependent variable) is not reliable, precise or accurate: _____

_____

____ DV is not objectively scored or subjective scoring without more than one scorer: _____

_____

____ DV is not valid: _____

____ Problems with Participant bias: _____

____ Problems with Attrition: _____

____ Small sample size: _____

____ Poor sample selection: _____

____ Problems with Experimenter bias: _____

____Premature generalization of results: _____

____ Confuse correlation with causality: _____

Video Helps Participants Improve Conversation Skills

Is initiating and maintaining a conversation difficult for you? Do people say that you talk over their heads? Under their heads? Stammer? If you said yes, then you might want to watch a video that could improve your conversation skills, at least according to recent research.

The researcher, Dr. Marie Vallus, gave 64 participants a standardized test of conversation ability that covered topics such as starting and maintaining conversations, being able to engage less responsive people, being sensitive to the knowledge of the other person, and changing topics. Based on the results, half were assigned to a low-ability group and the other half to a high-ability group. The participants had been selected via an advertisement asking for people who wanted to improve their conversation skills.

Each participant was then given an envelope which contained a word like "computers" or "love." The participants in each group paired up in dyads (couples). Each person was given 2 minutes to talk about his or her topic to the other person. The person that was listening rated the person talking on "speech fluency" using a 7-point scale. Speech fluency was defined by the number of pauses. One participant remarked "It was easy and fun. If I sensed the person paused a lot, then that would be a high-ish score. If the other person had few pauses then that would be a low-ish score." The participants were asked not to show their ratings to their partners so as not to influence their behavior.

After the participants spoke about their assigned topics and completed their ratings, all of them watched an hour video designed by an independent lab to help them in conversations. It covered topics like engaging people, being sensitive to others, changing topics, and reducing pauses.

Following the video, they were given additional envelopes with new topics and repeated the same procedure they did before the video. The same topics were used before and after the video, but distributed in such a way that no one saw the same word more than once. That way, the results could not be biased by the words themselves.

The researcher found that the ratings of pauses significantly decreased after watching the video. Watching the video seemed to help participants decrease their number of pauses, and this occurred in both low-and high-ability groups.

"These results suggest the video improves the many facets of conversation skills," said Vallus, "However, it probably needs further testing."

The flaws (You may use this list to guide your evaluation)

- No control or comparison group; No random assignment; DV (dependent variable) is not reliable, precise or accurate; DV is not objectively scored or subjective scoring without more than one scorer; DV is not valid; Problems with Participant bias; Problems with Attrition; Small sample size; Poor sample selection; Problems with Experimenter bias; Premature generalization of results; Confuse correlation with causality

Check any flaws that apply: **Video Helps Participants Improve Conversation Skills**

____ No control or comparison group: _____

____ No random assignment: _____

____ DV (dependent variable) is not reliable, precise or accurate: _____

_____

____ DV is not objectively scored or subjective scoring without more than one scorer: _____

_____

____ DV is not valid: _____

____ Problems with Participant bias: _____

____ Problems with Attrition: _____

____ Small sample size: _____

____ Poor sample selection: _____

____ Problems with Experimenter bias: _____

____Premature generalization of results: _____

____ Confuse correlation with causality: _____

Cell Phones and Driving

Talking on cell phones while driving would not have been an issue twenty years ago, but with developing technology, it is something that many people do every day without giving it a second thought.

Many people, on the other hand, believe that talking on cell phones while driving is detrimental to one's ability to drive. Indeed, drivers have less attention to devote to looking at the road and watching for traffic signs if they are talking on the phone while driving. A recent report has also indicated that 1.4 million car accidents are caused by drivers talking on their cell phones each year, but the question remains: Are these accidents really caused by talking on the phone, or could something else be playing a role?

Researchers from Highland University recently conducted an experiment about using cell phones while driving. Sixty people participated in the study, and all of them admitted to frequently talking on the phone while driving. They were all told before the onset of the study that the goal was to find out how cell phone use influences driving.

Each participant drove through a closed driving course that involved navigating between traffic cones. Participants were randomly assigned to either a cell phone or no cell phone group. Those in the cell phone group were required to drive through the course while talking on the cell phone. Those in the no cell phone group drove through the course without talking on the cell phone.

In order to determine participants' driving ability, there were electronic sensors in the cones which enabled the researchers to calculate the distance (in centimeters) between the car and the cones. When all drivers had completed the driving course and the distances between the car and cones were calculated and averaged, researchers found that the two groups (cell phones and no cell phones) were not different from each other. That is, they found that people using their cell phones while driving did not come any closer to and were not any farther from the cones when compared to people who did not use their cell phones while driving.

Recently, the researchers who conducted this study found similar results in another study, leading them to believe that this is evidence that talking on cell phones may not impede driving ability.

The flaws (You may use this list to guide your evaluation)

- No control or comparison group; No random assignment; DV (dependent variable) is not reliable, precise or accurate; DV is not objectively scored or subjective scoring without more than one scorer; DV is not valid; Problems with Participant bias; Problems with Attrition; Small sample size; Poor sample selection; Problems with Experimenter bias; Premature generalization of results; Confuse correlation with causality

Check any flaws that apply: **Cell Phones and Driving**

____ No control or comparison group: _____

____ No random assignment: _____

____ DV (dependent variable) is not reliable, precise or accurate: _____

_____

____ DV is not objectively scored or subjective scoring without more than one scorer: _____

_____

____ DV is not valid: _____

____ Problems with Participant bias: _____

____ Problems with Attrition: _____

____ Small sample size: _____

____ Poor sample selection: _____

____ Problems with Experimenter bias: _____

____Premature generalization of results: _____

____ Confuse correlation with causality: _____

Butterflies are Not Free

In the United States, butterflies represent big bucks. Wedding planners buy butterflies by the thousands so that newly married couples can witness thousands filling the sky as they say their vows. Wealthy home owners buy them to fill their gardens. Butterflies are becoming an industry to themselves.

But, there is a problem. That is, although butterflies grow on trees, many desirable species are in low supply. Here enters Jim Cowan, a high school biology teacher in Florida. Mr. Cowan used his biology class to test whether planting several sweet pepperbrush shrubs in a 20 mile area near his school would attract butterflies and keep them in the area.

"These shrubs were going to be planted by the park district anyway so I thought I would test whether they would increase the butterfly population in the area," explained Cowan. "If they did, then the kids would be happy knowing that they would have butterflies in their near future and even at their weddings."

At the beginning of the semester (1 month before the shrubs were planted) his class placed insect friendly traps throughout the area. Food and water were placed inside of the traps to keep the insects alive. Students working in teams recorded the number of trapped butterflies at weekly intervals. They were careful not to touch the butterflies or mark them in any way so they would not disturb them. Weekly recordings continued after a whopping 10,000 shrubs were planted in the park, and they ended at the completion of the semester, about 2 months later.

However, something occurred in the middle of the semester that did not ensure butterflies in the students' future other than in their stomachs.

"We experienced a devastating drought coupled with high temperatures. It only lasted a week, but it adversely affected the plants in the areas. Many of the pepperbrush shrubs died," Cowan lamented.

Fortunately, the study did not die out like the shrubs. "It appears that the butterflies in fact must have been attracted to the shrubs because when the shrubs died out, so did the butterfly population. There was a huge statistically significant drop in the number of butterflies before and after the shrubs died," revealed Cowan.

Cowan is so excited about his discovery that pepperbrush shrubs attract butterflies that he plans to raise butterflies at his uncle's farm and sell them to nearby wedding planners.

The flaws (You may use this list to guide your evaluation)

- No control or comparison group; No random assignment; DV (dependent variable) is not reliable, precise or accurate; DV is not objectively scored or subjective scoring without more than one scorer; DV is not valid; Problems with Participant bias; Problems with Attrition; Small sample size; Poor sample selection; Problems with Experimenter bias; Premature generalization of results; Confuse correlation with causality

Check any flaws that apply: **Butterflies are Not Free**

____ No control or comparison group: _____

____ No random assignment: _____

____ DV (dependent variable) is not reliable, precise or accurate: _____

_____

____ DV is not objectively scored or subjective scoring without more than one scorer: _____

_____

____ DV is not valid: _____

____ Problems with Participant bias: _____

____ Problems with Attrition: _____

____ Small sample size: _____

____ Poor sample selection: _____

____ Problems with Experimenter bias: _____

____Premature generalization of results: _____

____ Confuse correlation with causality: _____

Bulk-Up (ON-LINE AD)

I have been going to the gym for two months now in an effort to get into shape. However, progress has been disappointingly slow.

While I hang off of the weight machines, dripping sweat and hurting, this young guy with balloon arms comes in. You've all seen him. He seems to live in the gym. It's just demoralizing.

Well, if you're like me, you don't have time to hang out at the gym often enough to get that perfect shape. So, are we stuck with our drumstick muscles? Well, my friend, it appears that there is good news for the likes of us.

BulkUp is the answer! It contains a chemical that increases testosterone production. And testosterone can be used by our bodies to build muscle. That key chemical in BulkUp is naturally produced by humans as well as by some trees. So it's really just considered a dietary supplement.

Athletes who have tried BulkUp claim that it helps them to gain bulk. But, you don't have to take their word for it. In a recent study, a researcher tested BulkUp on a sample of thirty nationally rated body builders during their training for a national competition. He gave each athlete a different amount of BulkUp on a daily basis as part of their training regimen for a total of one month.

After that month, the researcher confirmed the effectiveness of BulkUp. When they stepped on the scale, the readouts revealed that the athletes on higher doses of BulkUp had gained more weight than athletes on lower doses.

So you see, muscle growth for you and me is just around the corner - it works, and you can afford it more easily than that expensive gym membership. And, BulkUp comes in three delicious flavors. Order your sample today - a one-month supply for a special rate of only 49.99!

[ORDER YOUR BOX HERE]

The flaws (You may use this list to guide your evaluation)

- No control or comparison group; No random assignment; DV (dependent variable) is not reliable, precise or accurate; DV is not objectively scored or subjective scoring without more than one scorer; DV is not valid; Problems with Participant bias; Problems with Attrition; Small sample size; Poor sample selection; Problems with Experimenter bias; Premature generalization of results; Confuse correlation with causality

Check any flaws that apply: **Bulk-Up**

____ No control or comparison group: _____

____ No random assignment: _____

____ DV (dependent variable) is not reliable, precise or accurate: _____

_____

____ DV is not objectively scored or subjective scoring without more than one scorer: _____

_____

____ DV is not valid: _____

____ Problems with Participant bias: _____

____ Problems with Attrition: _____

____ Small sample size: _____

____ Poor sample selection: _____

____ Problems with Experimenter bias: _____

____Premature generalization of results: _____

____ Confuse correlation with causality: _____

Facilitated Communication Helps Autistic Children

Autistic children face severe social, communicative and behavioral impairments. Not surprisingly, researchers of autism have begun to cast a wide net for finding ways to improve the communication skills of people with autism.

One promising procedure is known as *facilitated communication*, in which a person gently holds or touches an autistic child's hand while he/she is writing. It is believed that autistic children respond positively to physical contact resulting in improved communication.

The positive effects of facilitated communication were recently demonstrated in a study published in the journal *Autism*. Professor Art Jones, a leading autism researcher who ran the study, said that he was "interested in finding out if something as simple as physical contact with another human would help autistic children become better at something they have so much trouble with and something that is so essential to everyday life—communication with others."

The study employed parents of autistic children to serve as experimental assistants. These volunteers had been previously trained in facilitated communication and believed in its therapeutic value. "I needed help to run the children through the study, and the volunteers were happy to do so," said Jones.

Forty-four autistic children from Illinois participated in the study. These children were randomly assigned to either an experimental or a control group. In the experimental group, the volunteers gently held the children's hands as they wrote. In the control group, the same volunteers were present and sat next to the children as they wrote, but they did not have any physical contact with them.

Following the experiment, researchers who were unfamiliar with the hypothesis analyzed what the children wrote. Specifically, they looked at the complexity of the children's writing. Sentence complexity was defined on a number of variables, including the length of the sentences, and the number of words at or beyond the grade level of the child, both standard measures of writing complexity.

The results showed that the children in the experimental group—those exposed to facilitated communication—wrote significantly more complex sentences than those in the control group.

For Jones, the results of this study are very encouraging, but he is planning to replicate the study later this year before promoting the program as a therapy.

The flaws (You may use this list to guide your evaluation)

- No control or comparison group; No random assignment; DV (dependent variable) is not reliable, precise or accurate; DV is not objectively scored or subjective scoring without more than one scorer; DV is not valid; Problems with Participant bias; Problems with Attrition; Small sample size; Poor sample selection; Problems with Experimenter bias; Premature generalization of results; Confuse correlation with causality

Check any flaws that apply: **Facilitated Communication Helps Autistic Children**

____ No control or comparison group: _____

____ No random assignment: _____

____ DV (dependent variable) is not reliable, precise or accurate: _____

_____

____ DV is not objectively scored or subjective scoring without more than one scorer: _____

_____

____ DV is not valid: _____

____ Problems with Participant bias: _____

____ Problems with Attrition: _____

____ Small sample size: _____

____ Poor sample selection: _____

____ Problems with Experimenter bias: _____

____Premature generalization of results: _____

____ Confuse correlation with causality: _____

Are There Real Benefits to Using Antibacterial Soap?

The purpose of all soaps is to clean our skin. But clearly, not all soaps are created equal. Shelves are packed with soaps in all colors and shapes, some containing lotion, some containing exfoliating agents; some scented, some hypoallergenic.

If you have felt overwhelmed by the choices, you are not alone. Most people have expressed uncertainty when shopping for soap. One question that has plagued consumers for some time now is whether it makes a difference to buy antibacterial soaps.

Antibacterial soaps seem to be the default choice of public offices and institutions in order to curb the spread of harmful bacteria. But, is this kind of soap really superior to regular soap in effectively getting rid of harmful germs?

The answer appears to be no. Dr. Jelex and her team from a well-respected university recently obtained evidence refuting the higher effectiveness of antibacterial soap over plain soap.

The researchers set out to compare the number of colony forming units (CFUs; a standard bacteria count) in the palms of eight participants before and after they washed their hands with either antibacterial or plain soap, as determined by random assignment, for two minutes.

The soap was presented in no-name containers, thus, the participants were not aware of the condition to which they were assigned.

The number of CFUs was estimated based on samples taken by swab from a predetermined part of the palm for all participants. The swabs were labeled with numbers that allowed for blind scoring and later matching to the condition.

The CFU numbers of the four participants in the antibacterial soap group were reduced by 76; however, there was a reduction of 74 for the four participants in the plain soap group. The difference was not significant, according to Jelex.

On the basis of this finding, "we strongly recommend that consumers save their money and reach for the plain soap options," advises Dr. Jelex.

The flaws (You may use this list to guide your evaluation)

- No control or comparison group; No random assignment; DV (dependent variable) is not reliable, precise or accurate; DV is not objectively scored or subjective scoring without more than one scorer; DV is not valid; Problems with Participant bias; Problems with Attrition; Small sample size; Poor sample selection; Problems with Experimenter bias; Premature generalization of results; Confuse correlation with causality

Check any flaws that apply: **Are There Real Benefits to Using Antibacterial Soap?**

____ No control or comparison group: _____

____ No random assignment: _____

____ DV (dependent variable) is not reliable, precise or accurate: _____

_____

____ DV is not objectively scored or subjective scoring without more than one scorer: _____

_____

____ DV is not valid: _____

____ Problems with Participant bias: _____

____ Problems with Attrition: _____

____ Small sample size: _____

____ Poor sample selection: _____

____ Problems with Experimenter bias: _____

____Premature generalization of results: _____

____ Confuse correlation with causality: _____