



Evaluating Research Summaries: The Answers

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

Copyright 2021 by **Keith Millis, Diane Halpern, Katja Wiemer, Patricia Wallace**. All rights reserved. You may reproduce multiple copies of this material for your own personal use, including use in your classes and/or sharing with individual colleagues as long as the author's name and institution and the Society for the Teaching of Psychology (STP) heading or other identifying information appear on the copied document. No other permission is implied or granted to print, copy, reproduce, or distribute additional copies of this material. Anyone who wishes to produce copies for purposes other than those specified above must obtain the permission of the authors.

Contents

Evidence-Based Teaching Strategies in this Resource..... 3

To the Instructor.....4

Summary by Flaw Matrix.....5

Answers.....6

 Study with Music.....6

 Spontaneous Generation Exists!.....7

 Which is Stronger --- a Gorilla or a Chimp?.....8

 Pesticides.....9

 Can a Memory Pill Enhance your Short-term Memory?.....10

 Heavy Metal Music: A Teenager's Perspective.....11

 Gender and Aggression.....12

 "Pounds Off" Helps People Lose Weight.....13

 Let's Dance.....14

 Video Helps Participants Improve Conversational Skills.....15

 Cell Phones and Driving.....16

 Butterflies are Not Free.....17

 Bulk-Up (an online ad).....18

 Facilitation Communication Helps Autistic Children.....19

 Are There Real Benefits to Using Antibacterial Soap?.....20

Evidence-Based Teaching Strategies in this Resource

This booklet contains the answers to the booklet "How to Evaluate Research Summaries."

The activities included in this and the "Evaluate" 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 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, 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*.
- 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:

Here are the answers to the research summaries. We believe that the answers are accurate but in a few cases, one may disagree. There are many nuanced issues regarding any research. However, we think that discussions around disagreements may be very fruitful to understanding research methods.

An instructor might use the evaluation tasks in different ways. One is to assign some as homework or as in-class assignments. One might use the booklets electronically or as hard copies. Another option is to have students write out an explanation for each potential flaw (why it is or not a flaw) or only to those that they identify as flaws (why it is a flaw). Students can grade them by themselves or with a partner, using the answer key. Another option is to copy and paste the summaries (and answers) into a multiple answer questions within a teaching platform, such as Blackboard. Still another option is for students to write their own research summaries with flaws that others in the class can evaluate. While we think this is an engaging task for students (as demonstrated in an unpublished study), it will be necessary for them to learn how to evaluate summaries first.

On the next page, there is a matrix indicating which flaws correspond to which summary. This can be a quick guide for scoring or as a resource for teaching because one might want an example of a particular flaw.

Table 1: Summary by Flaw Matrix

Summary	Control Group	Random Assignment	DV not reliable, precise, accurate	DV not objectively scored	DV is not valid	Participant bias	Attrition	Sample size	Sample selection	Exp. Bias	Premature generalization	Confuse correlation & causation
Study music		x			x	x	x					
Spont. Gen.											x	
Gorilla/chimp			x					x				
Pesticides					x			x			x	
Memory Pill												
Heavy Metal			x		x						x	
Gender & Agg					x	x			x			
Pounds off	x						x					
Let's Dance	x							x	x	x		
Video & Conv.	x			x	x	x						
Phones & Driving						x						
Butterflies			x								x	x
Bulk Up					x				x		x	x
Fac. Comm.										x		
Antibacterial Soap								x			x	

Studying with Music

Flaw	Flaw in study	Explanation
No control or comparison group		The comparison group was the group of students that did not listen to music.
No random assignment	X	There was a problem with random assignment because the students were assigned to condition by where they sat in the classroom. There could be something different about how the students who sit in the front of the class study compared to those who sit in the back of the class.
DV could be more reliable, accurate, or precise		Marking a completed line of reading is a commonly accepted measure of reading rate. There was not a problem with sensitivity, precision, or accuracy in this study.
DV is not scored objectively		This is an objective measure because the students recorded their own objective data, which did not require any interpretation.
DV is not valid	X	How fast someone reads is not a valid measure of comprehension.
Participant bias	X	Because the students recorded their own data and they were aware of the hypothesis, their personal beliefs regarding this study may have influenced the results. So there is a problem with participant bias.
mortality or attrition	X	There is a problem with attrition here. When participants are dropped from a study it is possible that the differences found are due to the differences of the characteristics between those who were dropped and those who completed the study.
small sample size		Thirty-five students per condition seems like enough to find significant differences between groups.
poor sample selection		The researcher was interested in how listening to music affected music students' comprehension of material. The participants in this study were music students taking a music theory class, and therefore the sample tested is reflective of the population of interest.
experimenter bias		Experimenter Bias is not a flaw here because Dr. Robinson did not have the opportunity to influence the results in any way.
premature generalization of results		Premature generalization of results is not a problem here because it was reported that these results were reproduced.
confuse correlation with causation		This is not a correlational study. The manipulation of the independent variable and the conclusion indicates this.

Spontaneous Generation

Flaw	Flaw in study	Explanation
No control or comparison group		There are sufficient comparison groups in the study.
No random assignment		There is no indication that the assignment was not adequately random.
DV could be more reliable, accurate, or precise		There does not appear to be a problem with the accuracy or sensitivity of the count.
DV is not scored objectively		No subjective scoring is necessary in this study, so this flaw does not apply here.
DV is not valid		The measurement is pretty straightforward, there does not seem to be a problem with validity.
participant bias		The subjects in this case are fruit flies, which we can assume are not likely to be biased.
mortality or attrition		Of course in this study, mortality is tricky to track. However, given the study objective and the outcomes, if mortality occurred, it does not affect the conclusiveness of the study, so it is not a problem here.
small sample size		In this study, sample size is adequate, as we are dealing with numerous flies and ten jars per condition.
poor sample selection		There are no apparent problems with sample selection here, even though using only one species imposes some limitations.
experimenter bias		The description of the study does not give any indication that the experimenter was biased.
premature generalization of results	X	If you thought “the flies had laid eggs” in the jar, you would be correct. This is particularly important here because fruit flies are inclined to lay eggs in fruit; therefore, there is an obvious confound in this study. However, we do not have a category that fits this explanation. But what is certain is that the conclusion seems premature. Usually, a number of different studies are conducted before a strong conclusion is drawn.
confuse correlation with causation		The data are not correlational, thus this flaw does not apply here.

Which is Stronger: A Gorilla or a Chimp?

Flaw	Flaw in study	Explanation
No control or comparison group		The chimps and the gorillas are being compared so this is not a flaw.
No random assignment		This would be a flaw if the researcher was trying to make a causal statement. But since they are comparing two naturally occurring groups, they cannot randomize animals to a condition.
DV could be more reliable, accurate, or precise	x	It is definitely a problem to stop at only 100 pounds. Because there was no difference, this study needs a more sensitive measure.
DV is not scored objectively		Lifting the weight is an objective measure of strength. There is nothing subjective about this type of measurement.
DV is not valid		Lifting the weight is a valid measure of strength.
participant bias		The subjects were unaware of the experimenter's hypothesis.
mortality or attrition		All the participants were able to complete the study, so mortality is not a problem.
small sample size	x	Only 4 of each group is not enough, and this is especially a problem since they did not find a difference between groups.
poor sample selection		This is likely not a problem because they just used regular animals from their zoo. It would be a problem if zoo animals differ from wild animals on strength.
experimenter bias		There doesn't appear to be a way for the experimenter to influence the study, and she doesn't seem to have a vested interest in a particular result.
premature generalization of results		She replicated and found the same results, so this seems ok.
confuse correlation with causation		This is not a correlational study.

Pesticides

Flaw	Flaw in study	Explanation
No control or comparison group		The design is adequate for comparison.
No random assignment		The plants were randomly assigned to the test beds, so this flaw does not apply here.
DV could be more reliable, accurate, or precise		The measurements of numbers and weight seem adequately precise and sensitive.
DV is not scored objectively		There was no need for the biologist to make judgments for the scores, so there is no problem with objectivity.
DV is not valid	X	The biologist says he wants to test the effect of the pesticides on the quality of the produce, but he is using the quantity measures of count and weight. Quality would involve other aspects, such as the nutrient content.
participant bias		Tomatoes, like other nonhuman participants, are not affected by participant bias.
mortality or attrition		There is no indication that plants were lost during the study, so this flaw does not apply here.
small sample size	x	The sample size of fifteen for each condition is a bit small.
poor sample selection		There are no obvious flaws in how the plants were selected, so we will assume that the selection was appropriate.
experimenter bias		The biologist did his research on his own plants, so he obviously wanted a rich yield, but there is no indication that he favored one condition over the other.
premature generalization of results	x	The biologist concluded that pesticides increases the quality of produce based only on this one study.
confuse correlation with causation		The data used in this study are not correlational in nature, so this flaw is not a concern here.

Can a Memory Pill Enhance Your Memory? [NO FLAWS]

Flaw	Flaw in study	Explanation
No control or comparison group		In this study, half of the participants took a placebo pill and the other half took the memory pill. So, there was a good comparison group used here.
No random assignment		Participants were randomly assigned to one of two conditions. Either they took the memory pill, or they were given a placebo pill.
DV could be more reliable, accurate, or precise		Three lists of twenty items was an adequate amount of items to use for a measure of memory. So, there is not a problem with sensitivity, precision, or accuracy here.
DV is not scored objectively		The dependent variable was performance on a memory test, and this was an objective measure that needed no further interpretation.
DV is not valid		The experimenter was interested in how a pill would influence memory, so a recall task that tests short-term memory is a valid measure to use here.
participant bias		The participants were under the impression that they were being tested for their blood pressure and that the recall task was part of a health check. They did not know that the study was investigating memory. Participant bias was not an issue here.
mortality or attrition		All of the participants completed the study. Attrition or mortality was not an issue with this study.
small sample size		Forty-nine participants is a good amount for finding significant differences between groups, which did not happen here. Sample size is not an issue in this study.
poor sample selection		Sample selection is not an issue here. These participants should adequately reflect the general population with regards to memory.
experimenter bias		It was not possible for Dr. Phillips to influence the results of the study, so experimenter bias is not an issue here.
premature generalization of results		This study was a replication of previous findings regarding this memory pill. So, it is not premature to conclude that this pill is not effective.
confuse correlation with causation		This study did manipulate whether or not participants received the memory pill. So, the concluding statement refers to causality and not to correlation. Confusing correlation with causality is not an issue here.

Heavy Metal Music: A Teenager's Perspective

Flaw	Flaw in study	Explanation
No control or comparison group		There was a comparison group that was instructed not to listen to music for one month. This was a good comparison group to use for this study.
No random assignment		Participants were randomly assigned to either the control group or the heavy metal group.
DV could be more reliable, accurate, or precise	X	The anxiety rating scale that was used only allowed for two response options: not anxious or anxious. It would have been better to have a series of response options to tap into a variety of scores. This lack of precision creates a flaw related to the sensitivity of the dependent variable.
DV is not scored objectively		The measure was self-report so the participants made mood judgments, not the researchers. All the researchers had to do was count the number of anxious and non-anxious scores. This is objective.
DV is not valid	X	Although anxiety and depression tend to co-occur in individuals, they are different constructs. For this study, the researcher needed to measure depression in order to draw any conclusions about depression. This study is flawed because the failure to measure depression means the study lacks construct validity.
participant bias		The subjects did not know the purpose of the study and could not guess it from the survey. The survey included many distractor questions to disguise its purpose.
mortality or attrition		The researchers do not report a problem with participants dropping out of the experiment or failing to do what they were instructed to do.
small sample size		Forty participants per group should be more than enough to find a significant difference if there really were one between the heavy metal and no heavy metal conditions. It is also large enough to lead to a generalizable conclusion.
poor sample selection		Sample selection is not a flaw in this case because it is unclear how the sample of participants was obtained.
experimenter bias		The experimenters did not know which condition participants were in when they did the scoring. Such blind scoring is used to reduce the opportunity for experimenter bias.
premature generalization of results	X	Well, the writer made a hasty conclusion because there was only one study. Besides, we should always be careful making strong conclusions when you find no significant difference. The conclusion that it will not lead to depression is too strong. One can say that there is no evidence that it leads to depression but not that it doesn't lead to depression.
confuse correlation with causation		This is an experiment not a correlational study.

Gender and Aggression

Flaw	Flaw in study	Explanation
No control or comparison group		Males are being compared to females, so there is a comparison group.
No random assignment		This is not the type of variable that one can be randomly assigned to participants, and since a causal statement is not being made, this study does not have the flaw of no random assignment.
DV could be more reliable, accurate, or precise		The 7-point scale seems to be sensitive enough because they found a significant effect.
DV is not scored objectively		While the coding is subjective, it is coded consistently by two researchers reading the messages independently so this is not a flaw of the behavior not being scored objectively.
DV is not valid	X	Verbal aggression is only one form of aggression. This form alone is not a valid measure of the complete sense of aggression.
participant bias	X	This study has participant bias because the volunteers knew they were signing up for a study on gender differences in aggression.
mortality or attrition		All the participants completed the study so this is not a problem.
small sample size		25 males and 25 females is a good sample size for this study.
poor sample selection	X	This study has a problem with poor sample selection because they recruited people interested in participating in an aggression study.
experimenter bias		The scoring was done blind to condition so the experimenter could not bias the results.
premature generalization of results		The findings were replicated with the same results so this is not a problem.
confuse correlation with causation		This is not a correlational study, and in fact, there is no causal statement made.

Pounds-Off

Flaw	Flaw in study	Explanation
No control or comparison group	X	Notice that the hypothesis states that the pill will help a person lose weight. In order to accurately test this question, you would need a comparison group that does not receive the pill. The flaw is an absence of a comparison group.
No random assignment		Since there was only one group, there was no need for random assignment to condition. This is not a flaw in this study.
DV could be more reliable, accurate, or precise		Measuring the number of pounds lost during a 3-month period seems sensitive enough to find an effect. People can lose about 1-2 pounds per week, in general.
DV is not scored objectively		Measuring weight on a scale is an objective measure especially if the scale is digital.
DV is not valid		Weighing people on a scale is a valid measure of weight loss.
participant bias		The participants were unaware that the purpose of the study was to test a weight loss pill. It is a good thing they didn't just eat more when they noticed they were losing weight.
mortality or attrition	X	Whenever we lose participants from a study, there is a chance that any differences that are found may be due to characteristics of those who dropped out and those who remained. Several people were removed from the study because they did not take the pill every day; therefore, there is a flaw of attrition.
small sample size		There were 30 participants, and this seems like a large enough sample to attempt to find a significant difference in weight loss in human adults. It also seems like enough participants to be able to generalize from the sample to the population.
poor sample selection		The participants were randomly selected to participate in the study. This is an unbiased method of selecting participants.
experimenter bias		The researchers used measures to limit their influence on the participants by having an independent physician measure weight during the full physical.
premature generalization of results		The result was successfully replicated, and the conclusion that the use of this pill appears to aid adults in losing a significant amount of weight is presented tentatively and not as a proven fact.
confuse correlation with causation		This is not a correlational study.

Let's Dance

Flaw	Flaw in study	Explanation
No control or comparison group	X	To be able to make a claim about how effective the dance video is, there needs to be a group that did not watch the dance video. This did not happen here, and it is a problem with this study.
No random assignment		Because there was no control group, participants were not randomly assigned to condition. So, random assignment could not be a flaw here.
DV could be more reliable, accurate, or precise		There is no reason to believe that the coding technique was not sensitive, accurate or precise.
DV is not scored objectively		The measure used in this study was an objective score coded by the experimenter.
DV is not valid		The "broom test" used in this study is a valid measurement of dance skill.
participant bias		As the questionnaire indicated, participants did not know that the video was supposed to improve their dancing ability. So, participant bias was not a flaw here.
mortality or attrition		All of the participants completed the study. Attrition is not a flaw.
small sample size	X	This study only used ten participants, which is too small of a sample to make claims that could generalize to the population.
poor sample selection	X	It is important to ensure that the sample tested is reflective of the population you are making claims about. Considering the sample chosen for this study were self-selected dance students and not necessarily representative of dance students or the general population, it is an example of a poor sample selection.
experimenter bias	X	There was a problem with experimenter bias here. Ms. Webber obviously had strong feelings about the possibility of the effectiveness of this video and she did have an opportunity to influence the results of the study.
premature generalization of results		This study was replicated and the results were the same. So disregarding the other problems, premature generalization is not a flaw here.
confuse correlation with causation		Correlation is not confused with causation here because Ms Webber is not making a strong causal claim regarding the effectiveness of the dance video.

Video Helps Participants Improve Conversation Skills

Flaw	Flaw in study	Explanation
No control or comparison group	X	To evaluate Vallus's claim that the video improves conversation skills, the study needed a comparison group that did not watch the video or did some other treatment.
No random assignment		Technically, that's true, but it doesn't apply to this study because everyone was in the same treatment.
DV could be more reliable, accurate, or precise		Given the nature of the scale that was used, and the fact that they found differences between pre and post testing, this does not seem to be a problem.
DV is not scored objectively	X	The subjects scored their partner. They were not trained in this type of scoring and they were not blind to condition. The researchers should have had two independent raters score each participant's fluency.
DV is not valid	X	Speech fluency (how smoothly a person speaks) is only a small part of what is needed to be skilled at making conversation. Because of this, the construct of conversation skills is not being adequately measured. Conversation skills would need to be assessed in ways that cover all aspects of this construct.
participant bias	X	Because participants knew how their speech fluency was being evaluated, they might have been careful not to pause or say um during the second testing.
mortality or attrition		There is no problem with attrition or mortality since no one failed to complete the pretest, treatment, and posttest.
small sample size		64 participants seems like a large enough sample to attempt to find a significant difference and to be able to generalize from the sample to the population.
poor sample selection		It is uncertain whether this is a problem or not because they never mention how they recruited their sample.
experimenter bias		The researcher did not have a real opportunity to bias the outcome of the study.
premature generalization of results		The researcher's conclusion was not too strong, and she warned that more research should be done.
confuse correlation with causation		This study was not correlational. But the researcher should not make a causal statement for other reasons.

Cell Phones and Driving

Flaw	Flaw in study	Explanation
No control or comparison group		The group that drove without talking on the cell phone is the comparison group so this is not a flaw.
No random assignment		Participants were randomly assigned to groups so this is not a flaw.
DV could be more reliable, accurate, or precise		Measuring in centimeters appears to be sensitive, accurate, and precise enough.
DV is not scored objectively		Driving ability was determined by calculating distance measured by sensors which is an objective measure.
DV is not valid		Distance from an object is a valid measure of driving control.
participant bias	x	Because participants knew that they were being tested for the negative effects of driving while talking on the cell phone, this may have changed the way they drove normally.
mortality or attrition		All participants finished the course so attrition is not a problem with this study.
small sample size		Thirty participants per group is a good size for generalizing and detecting a difference between groups.
poor sample selection		We cannot tell if this is a flaw because the study doesn't say how participants were selected.
experimenter bias		The researcher did not have a chance to bias the outcome of the study.
premature generalization of results		The author's conclusion is reasonably tentative.
confuse correlation with causation		This study was not correlational.

Butterflies are Not Free

Flaw	Flaw in study	Explanation
No control or comparison group		There is no control group because this is a correlational study. Therefore, this is not a problem.
No random assignment		This is not a problem because there are no groups or conditions.
DV could be more reliable, accurate, or precise	x	I think accuracy is a flaw because there is no guarantee that each butterfly was caught just once. Therefore, it would be difficult to determine the exact number of butterflies.
DV is not scored objectively		Counting butterflies in teams seems objective to me.
DV is not valid		The number of butterflies is generally a valid measure of the butterfly population.
mortality or attrition		Mortality and attrition are problems only when the remaining subjects are different from the ones that dropped out. This results in a biased sample that affects the interpretation of the findings. However, in this case, the dependent variable was the number of butterflies. So, we would not be able to explain the decrease in butterflies due to the flaw of mortality or attrition.
small sample size		It didn't mention any sample size, so this would not be a flaw.
poor sample selection		This doesn't seem to be an issue because the sample was naturally occurring.
experimenter bias		There was no suggestion that the students who recorded the butterflies knew the hypothesis.
premature generalization of results	x	He is generalizing too soon because he is starting a business on the basis of a single study.
confuse correlation with causation	x	This is an important issue here. This was a correlational study but he is making a causal claim from it.

Bulk-up

Flaw	Flaw in study	Explanation
No control or comparison group		This study does not use experimental groups. Instead, it varies the bulk-up across all participants.
No random assignment		Random assignment does not apply here because everybody is in a different condition.
DV could be more reliable, accurate, or precise		The use of a weight scale seems appropriately accurate and sensitive for assessing weight.
DV is not scored objectively		Weight does not require a scoring judgment, so we do not need to worry about objective scoring here.
DV is not valid	X	We cannot judge to what extent the weight that was gained was actually due to muscle gains.
participant bias		There is no indication from the description that subjects may have been biased in this study.
mortality or attrition		It does not look like any athletes dropped out from the study.
small sample size		The sample size seems adequate for this study.
poor sample selection	X	To generalize to the overall population, we would want to see a study performed on participants who are not all nationally rated body builders but who are representative of the general population.
experimenter bias		We have no indication that the experimenter was biased.
premature generalization of results	X	This report endorses bulk-up as if its effectiveness was clearly supported. Given that every study has some limitations, we would want to see a replication study to be performed before endorsing any product.
confuse correlation with causation	X	Just because the athletes who took more bulk-up weighed more, we cannot conclude that the additional gain was due to the higher dosage. These athletes may have, for example, eaten more. It is not appropriate to make causal claims from correlational research.

Facilitated Communication Helps Autistic Children

Flaw	Flaw in study	Explanation
No control or comparison group		One group of autistic participants did not receive contact, so this is not a flaw.
No random assignment		Participants were randomly assigned to groups, so this is not a flaw.
DV could be more reliable, accurate, or precise		Because the results were significant, the measure was clearly sensitive enough.
DV is not scored objectively		The measures were counts of things which can be scored objectively. This is not a flaw.
DV is not valid		The measures used were standard measures of writing complexity. The measures included many ways of defining complexity, not just length.
participant bias		The autistic children did not know the purpose of the study and did not know they were in different conditions. Participant bias is not a flaw.
mortality or attrition		All the children completed the study, so there is no flaw here.
small sample size		22 children per condition is enough to generalize, and because there were significant results it is not a problem here.
poor sample selection		It is unclear how they selected the children, so we won't count that as a flaw in this case.
experimenter bias	x	The volunteers served as experimenters. They may have been biased in that they could have guided the children's hands, thus helping them write complex sentences. This is a likely possibility because they believed in the procedure and volunteered.
premature generalization of results		The researcher does not overstate the results and even noted that future testing should be done.
confuse correlation with causation		This is not a correlational study, so this is not a possible flaw with this study.

Antibacterial Soap

Flaw	Flaw in study	Explanation
No control or comparison group		The main question in the study is whether antibacterial soap is more effective than other soap. The experimental conditions allow for that comparison. But generally, it is desirable to include a no-treatment condition.
No random assignment		Participants were randomly assigned to the soap conditions, so this is not a concern.
DV could be more reliable, accurate, or precise		We can assume that the dependent variable was measured with adequate accuracy and sensitivity, because the researchers employed a measure that is accepted as a standard count.
DV is not scored objectively		There is always a chance that data are not scored entirely without error, but this study involves the objective scoring of bacteria counts. Objectivity is not a problem.
DV is not valid		The CFU count used in this study is a standard measure for bacteria, and represents the variable the researchers are interested in.
participant bias		Participants were not aware of the experimental condition they were assigned to, so there should be no participant bias affecting the data.
mortality or attrition		There are no drop-outs from this study that could have compromised the results.
small sample size	X	The sample size is a concern here. A total of eight participants, or four per group, is not a sufficient sample to produce trustworthy results.
poor sample selection		We have no information on how the sample was selected; we shall assume that the participants were representative of the larger population.
experimenter bias		There is no indication that the experimenters were biased in their analysis and interpretation. In fact, they were blind to the experimental condition while analyzing the bacteria counts.
premature generalization of results	X	The explicit recommendation of antibacterial soaps based on these findings is not warranted. Dr. Jelex is drawing overly strong conclusions from the study.
confuse correlation with causation		Since the researchers controlled for which kind of soap was used, and how long participants washed hands with it, we are dealing with a controlled experiment that generally permits conclusions about causal relations between the variables.