v 1.0

ACTIVITY 7

Improving Experimental Design

COMPUTER SIMULATION

ACTIVITY 7

Improving Experimental Design

ACTIVITY SUMMARY

Students use a computer simulation to see how adjustments to experimental design affect the reliability and interpretation of the results. First, they manipulate sample size to understand how results can vary widely even with the same experimental design. Then, they apply their understanding of sample size to explore different treatments and methods for assigning participants to experimental and control groups. As students progress, they explore how reliable results depend on large sample sizes and reducing confounds, which prepares them to design their own randomized controlled trial in Activity 8.

ACTIVITY TYPE
COMPUTER SIMULATION

NUMBER OF 40–50 MINUTE CLASS PERIODS
1-2

KEY CONCEPTS & PROCESS SKILLS

1 Incorrect conclusions about causation can happen when the result is due to an alternative explanation, such as chance or a confound. Careful study design and analysis can reduce the likelihood of the occurrence of alternative explanations.

2 Randomized controlled trials (RCTs) are the most reliable method for identifying cause-and-effect relationships because they reduce the likelihood that alternative factors are influencing the effect.

NEXT GENERATION SCIENCE STANDARDS (NGSS) CONNECTION:
Analyze data using tools, technologies, and/or models (e.g., computational, mathematical) in order to make valid and reliable scientific claims or determine an optimal design solution. (*Science and Engineering Practice: Analyzing and Interpreting Data*)

CONCEPTUAL TOOLS

CAUSAL REASONING

RANDOMIZED CONTROLLED TRIALS

VOCABULARY DEVELOPMENT

control group
the group in an experiment that does not receive the treatment and is compared to the experimental group

experimental group
the group in an experiment that receives the treatment and is compared to the control group

treatment
the procedure or situation that is changed only for the experimental group in a scientific study

TEACHER BACKGROUND INFORMATION

Random Sampling

Random sampling, also known as random selection, is a method in which a smaller group is chosen from a larger group in a way that gives every member an equal chance of being selected—like drawing names from a hat. This helps ensure that the sample represents the whole population fairly and without bias. However, if the population is very diverse and the sample size is small, the results may not be accurate. In a small sample, each individual has a greater impact on the overall average, which can lead to an unrepresentative outcome just by chance.

Random sampling of people is often more challenging than conducting controlled experiments in fields such as chemistry and physics where researchers directly manipulate variables instead of relying on participant selection. In human studies, selecting participants is not always truly random because it depends on who is willing to participate and how they are recruited. For example, national opinion polls try to gather a representative sample that mimics random selection, but this is becoming more difficult in recent years because fewer people have been agreeing to participate. Those who do participate tend to differ from the general population—they may be less busy, older, or more trusting of institutions such as polls—which makes the sample less representative of the general population.

MATERIALS & ADVANCE PREPARATION

FOR THE TEACHER

- VISUAL AID 7.1 "Random Sampling" (OPTIONAL)
- VISUAL AID 7.2 "Comparing Trials"
- VISUAL AID 7.3 "Example Results at Different Sample Sizes" (OPTIONAL)
- VISUAL AID 7.4 "Scoring Guide: Analyzing and Interpreting Data (AID)"
- ITEM-SPECIFIC SCORING GUIDE: Activity 7, Build Understanding item 3

FOR EACH PAIR OF STUDENTS

- COMPUTER WITH INTERNET ACCESS

FOR EACH STUDENT

- STUDENT SHEET 7.1 "Comparison of Study Designs"
- SCORING GUIDE: Analyzing and Interpreting Data (AID) (OPTIONAL)

Review the Dapple Experiments simulation before class so you are familiar with how it functions and are prepared to answer questions.

TEACHING NOTES

Suggestions for **discussion questions** are highlighted in gold.

Strategies for the **equitable inclusion of diverse students** are highlighted in lime.

GETTING STARTED (10 MIN)

1 **Prepare students to explore a computer simulation about experimental study design.**

- In Activity 6, students learned about key characteristics of observational studies. In this activity, students will be introduced to another type of study that can provide better evidence for a cause-and-effect relationship between variables. Students will explore the key elements of randomized controlled trials (RCTs) as they review how sample size and confounds can affect study results. At this point, do not define RCTs for students, as the goal is to allow them to discover this on their own as they work through the simulation. In Activity 8, students will receive a formal definition.
- Have students share examples of games they've played that they think improved their moods. Ask, Why do you think those games improved your mood? Ask students to describe any evidence they have—perhaps they felt more relaxed, had fun, or connected with friends while playing. Some students may also mention that games served as a distraction from other tasks they could have or should have been doing. Ask, How would you design a study to determine how playing a game affects mood? At this point, allow for all responses. Let students know that they will be using the Dapple Experiments simulation to investigate how playing a game affects well-being. Their goal is to design a study to answer the research question: *Does playing Dapple increase well-being for kids in Salas?*

PROCEDURE SUPPORT (40 MIN)

2 **Students begin to explore experimental study design in Part A.**

- The fictional scenario presented in Procedure Step 1 can be shared with the class in multiple ways. Read the scenario aloud to the class or have individual students read it aloud while others follow along with the text (either as a whole class or in small groups). Reading the scenario aloud can better support comprehension for many students, including neurodiverse students and emerging multilingual learners who often have more highly developed listening and oral skills than reading comprehension skills. Alternatively, students can read the scenario independently.

- Support students as they work through Procedure Steps 2–4. The simulation is designed to get students thinking about what information they need to evaluate a study with a focus on sample size and effect size.

- Support students, particularly emerging multilingual learners, in sensemaking and language acquisition as they read the text in the simulation. Circulate around the room and check in with students as they use the strategy to decode scientific ideas and construct meaning as they progress through the simulation screens.

- After Procedure Step 4, you may want to point out that each trial run in the simulation uses random sampling, meaning that even if the same sample size is used, different people are randomly selected to participate each time. You may wish to display optional Visual Aid 7.1, "Random Sampling," to explain this concept.

- When students have completed Part A, review their responses to Procedure Step 5. Ask, Why aren't the results of the eight trials in Part A the same, even though they had identical study designs? Students should recognize that each trial is an independent event, and therefore, the results will vary. Students may also notice that the variation in results is influenced by both the small sample sizes and random sampling, which means that different people are selected for each trial. If students don't identify this on their own, don't explain it yet as they will continue to explore how sample size impacts effect size in Part B of the simulation.

  Sample Student Response, Procedure Step 5

  The results are different because each trial is separate, and the people chosen are random. The numbers vary so much because each person is different and has their own individual well-being level. We only tested six individuals in each trial, so it is hard to know whether it worked or not.

  TEACHER'S NOTE: If students have completed Unit 3, Scientific Uncertainty and Probabilistic Reasoning, in the *Scientific Thinking for All: A Toolkit* curriculum, they may recognize that random errors are responsible for variability (noise) in the results in Parts A and B of the simulation. Random sampling combined with a low sample size increases the potential for random error due to greater variability between the sample and the population. Likewise in Unit 3, students explore how larger sample size and averaging reduce uncertainty from random errors. Connect students' inquiry in this activity to the concepts of signal and noise from the other unit.

3 In Part B, students use the Effect Sizes for Different Sample Sizes graph in the simulation to see how sample size affects the trial results.

- Before students start Part B, help them interpret the Effect Sizes for Different Sample Sizes graph shown in the lower section of the simulation by displaying Visual Aid 7.2, "Comparing Trials." Ask, Which sample size, A or B, shows more consistent results? Students should see that Sample Size B shows more consistent results because the arrows are closer to one another, are similar in length, and are going in the same direction.

- Review what the arrows represent in the simulation by using the following descriptions:
  - arrows: a single trial
  - arrow location on the x-axis: the sample size of that trial
  - tail of an arrow: the average well-being level of participants in the control group
  - tip of an arrow: the average well-being level of participants in the experimental group
  - length of an arrow: the effect size, which is the same as that shown in the corresponding bar graph in the upper right of the simulation
  - arrow color: blue shows that the Dapple players had higher well-being on average; red shows that the non-players in the control group had lower well-being on average

- Students can explore the results at each sample size in any order but will need to run trials at every sample size in the graph (from 6 to 120) to see how the reliability of results slowly improves with more samples. Encourage students to repeat trials multiple times at each sample size to find a pattern.

- After completing the graph by running trials with a sample size of up to 120, students will be able to select the Run 25 Trials At A Time button. This allows them to see 25 trials at every sample size, helping them to better estimate the smallest sample size where the results become consistent and reveal the true effect size of playing Dapple. The true effect size is the actual strength of a relationship or difference in the population, not just what a small number of trials finds. Share with students that the true effect size can be estimated from the average arrow length at the sample size where the results become consistent across trials..

- After Procedure Step 7, ask students to share their responses about the sample size and true effect size where the trial results become consistent. Students may have different ideas about what is consistent enough, depending on how precise an estimate of the effect size they want. Emphasize that there is not a correct answer. Answers will vary but generally, the more precise and certain students want to be about the results, the bigger the sample size needed. Students may suggest a sample size between 72 and 114 as the point where results become reliable and recognize that the true effect size is 1.5 points on the 10-point well-being scale.

  Student Sample Response, Procedure Step 7

  The smallest sample size that gives reliable results is 102, because that's when the blue lines are similar lengths, which is about 1.5, so this is probably the true effect size.

- Facilitate a class discussion about what the Results Across Dapple Trials graph suggests. Ask, Why do you think that smaller sample sizes result in more varied and less reliable results than a large sample size? Students should understand that this happens because some individuals are very different from the average in a population (i.e., they are statistical outliers). In a small sample, these individuals have a stronger influence on the results; in a larger sample, the effects of individual differences are balanced out.

4 **In Part C, students explore six possible study designs and choose the best one.**

- There are 12 possible configurations for the study designs (2 treatments x 2 controls x 3 methods of assignment). While students do not need to do all the combinations, they should explore at least 6 and understand that there are additional possibilities.

- In Procedure Step 11, hand out Student Sheet 7.1, "Comparison of Study Designs." Results for Procedure Step 12 will vary from group to group because the simulation does not yield identical results across trials. However, if students choose adequate sample sizes, the same study design should give similar results across trials. A sample student response to Student Sheet 7.1 is shown at the end of this activity. A few of the more notable results students should notice include the following:

  - The experimental group that plays the game once results in little or no effect size since the treatment may not be long enough to make a difference.

  - There are three confounding variables.

    1. Assigning groups by neighborhood places the control group in a southern area with many parks and the experimental group in a northern area without parks. This confound results in lower well-being in the experimental group because of less exposure to green spaces, not because of the game.
    2. Assigning groups by grade places middle school students in the control group and high school students in the experimental group. This confound is less obvious, making it a good example of a hidden confound. Students will only recognize it as a confound if they compare a study design with grade-based assignment to one with random assignment (the coin flip) and see that assignment by grade level inflates the difference in well-being (implying that middle schoolers in Salas have lower well-being than high schoolers).
    3. The control group with the condition "sits on the sidelines" creates a confound by lowering non-players' well-being simply because they don't get to play, which could distort the true effect size.

  - The study design with the biggest effect size has the following settings:
    - The experimental group plays the game once a day for a week.
    - The control group sits on the sidelines.
    - Group assignment is based on grade.

    However, as just described, both "sitting on the sidelines" and "assignment by grade" are confounds that can falsely increase the effect size.

  - Larger effect sizes or results opposite to the true effect may also arise from small sample sizes, which can be pulled to extremes by outliers.

- In Procedure Step 13, students will be asked to label their best study design with a star. The best study design has the following settings to avoid confounds and false results from low sample size:

```
- Sample size: 120
- Experimental group: Plays the game once a day for a week
- Control group: Gather in same size groups without playing Dapple
- Assignment to group by coin flip
```

5 **Optional: Support students' inquiry into the error bars displayed on the graphs.**

- The error bars in the simulation represent standard error, a statistic that estimates the variation across multiple samples of a population. Smaller standard errors occur from lower variation in a data set and larger sample sizes. While students don't need a formal understanding of standard error, the Error Bar toggle in the simulation allows them to explore how variation impacts the size of the error bars and confidence in the true effect size.

- In Part C, the error bars indicate whether differences between groups might be due to chance. If the bars overlap, the difference could be random. If they do not overlap, the difference is less likely to be due to chance. Use optional Visual Aid 7.3, "Example Results at Different Sample Sizes," to show that larger sample sizes have smaller error bars because extreme values tend to average out at large sample sizes.

SYNTHESIS OF IDEAS (10 MIN)

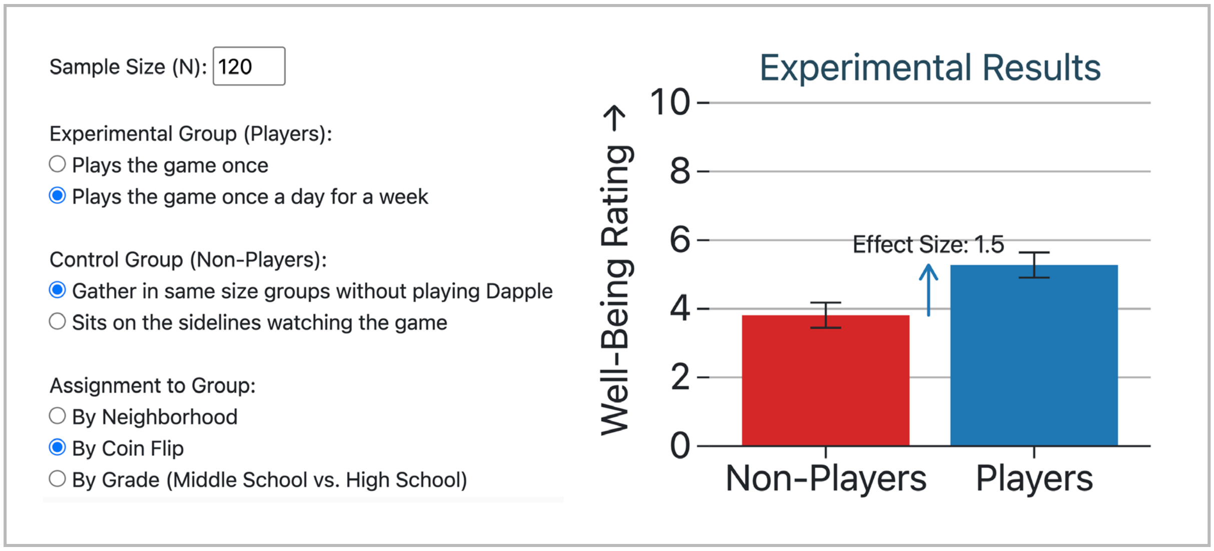
6 **Facilitate a discussion about the best study design for studying the effects that playing Dapple has on well-being.**

- Have a few students share with the class what they thought was their best study design and have them explain their reasoning.

- In Procedure Step 14, facilitate a discussion to review the strengths and weaknesses of the different study designs. Ask the following questions to guide the discussion and check students' understanding:
  - What is the best sample size for the investigation? Students should recall from Parts A and B that larger sample sizes produce more consistent results by minimizing the impact of individual outliers, leading to more reliable findings.
  - Which control condition was better—having kids gather in the same size groups somewhere else and not play or sitting on the sidelines watching? Students should recognize that having kids gathering somewhere else and not playing is the better control condition. Sitting on the sidelines could introduce a confound, as frustration and boredom from being unable to talk or do anything else might affect results.
  - What was the best way to assign participants to the experimental and control groups? Students should identify random assignment by coin flip as the best method. This approach prevents confounds related to factors such as neighborhood or grade level.
  - Was it better to have kids play the game once or once a day for a week? Students should find that playing daily for a week produced a stronger effect, while playing only once was too weak a treatment to show a noticeable impact in the experimental group.

- Why do scientists repeat experiments many times with different people and how does this help scientists be more confident in identifying a cause-and-effect relationship? Students should recognize that scientists repeat experiments with different people to ensure that the results are not just due to random chance. If the results are consistent across multiple trials, scientists can be more confident that the observed effect is real and not a fluke.

- Help students summarize what they have discovered about study design by reviewing the elements that contributed to the best design. The optimal study design settings and the true effect size are shown in the following screenshot from the simulation.



- Revisit the research question that began the activity, *Does playing Dapple increase well-being for kids in Salas?* If playing Dapple does increase well-being for kids, by how much? Students should respond that yes, the results show an increase in well-being from playing Dapple. Ask students to reflect on how the best study design impacted their confidence in this conclusion. Students should recognize that the best study design produced more reliable results by trying to avoid confounds and by using a large enough sample size to accurately measure the effect size.

7 **Review student responses to Build Understanding items 1 and 2.**

- Students can complete Build Understanding items 1 and 2 individually, in pairs, or in small groups. It is important to be sure that students' responses to item 1 indicate that they understand that larger sample sizes increase confidence by reducing random variation but do not eliminate confounds, so accuracy is not guaranteed. For item 2, responses should show that confounds create uncertainty about whether the treatment or another factor caused the effect. It may help to discuss that in Part C of the simulation, assigning groups by neighborhood may have made the control group seem to have higher well-being, not because they didn't play the game, but because they had more access to parks.

- Optional: If you discussed error bars with your students in Teaching Step 5, ask them how error bars and sample sizes are related. Show optional Visual Aid 7.3, "Example Results at Different Sample

Sizes," again. Students should explain that larger sample sizes result in smaller error bars because they produce more consistent and reliable results, which leads to a narrower range of possible true values for the effect and increases confidence in the findings.

8 Assess student growth, using the Analyzing and Interpreting Data (AID) Scoring Guide for Build Understanding item 3.

- Remind students of the Analyzing and Interpreting Data Scoring Guide. You may wish to project Visual Aid 7.4, "Scoring Guide: Analyzing and Interpreting Data (AID)," for your students to review each level and clarify your expectations.
- Do not share the item-specific version of the Scoring Guide (Item-Specific Scoring Guide: Activity 7, Build Understanding Item 3) with students as it provides specific information on how to respond to the item prompt.
- Remind students that you expect to see them demonstrate growth in their understanding and explanation of analyzing and interpreting data and that they may want to review their responses to the assessment in Activity 4 (Build Understanding item 2). You may also want to let students know that they will have one more opportunity in the unit to be assessed (Activity 9, Build Understanding item 2).
- Sample responses for Levels 1–4 are provided in the Build Understanding section that follows. Review these responses to get an idea of what is expected for each level alongside the Item-Specific Scoring Guide. See Appendix 2: Assessment Resource at the end of the Teacher's Guide for more guidance and information on using the Scoring Guides and assessment system with your students.
- Depending on your students, you may want to have them provide feedback on one another's work for revision prior to turning in their work to you for scoring. Alternatively, consider having students turn in a rough draft to you for feedback and revision.

9 Revisit the Guiding Question for the activity.

Finish the activity by revisiting the Guiding Question, *How do changes in study design affect the results?* Use responses to this question to formatively assess students' understanding of the key concepts and process skills related to experimental design such as sample size, effect size, and confounds.

SAMPLE STUDENT RESPONSES

BUILD UNDERSTANDING

1. **Think of a study in which you would want a really large sample size, such as studying a new medicine for asthma.**

   a How would a bigger sample size affect your confidence in the results? Explain your answer.

   A bigger sample size would increase my confidence in the results because it helps reduce the random differences that can happen when only a few people are involved.

   b Does having a really big sample size guarantee that your results will be accurate? Why or why not?

   A really big sample size doesn't guarantee accurate results because it does not get rid of confounds, other factors that could still affect the results.

2. **Why do confounding variables make it harder to interpret the results of a study? Use the example of green spaces from the simulation to illustrate your answer.**

   `HINT: Consider the results in Part C of the simulation when assigning participants by neighborhood.`

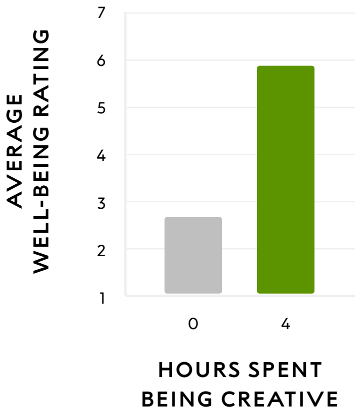
   Confounding variables make it harder to interpret results because you can't tell if the difference between experimental and control groups is because of the treatment or because of the confounding variable. For example, when we studied the game in Part C of the simulation, the parks were a confound if you chose the groups by neighborhood. The control group could have had higher well-being than the experimental group, not because of the game but because they lived near the park and had more time to spend in green spaces.

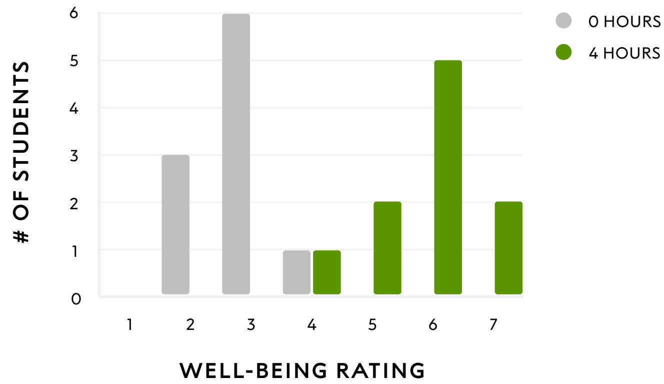
3. **AID Assessment**

   Ms. Lee's class decided to conduct an experiment to test the effect of time spent being creative on a person's well-being. They had 20 students from the class choose to do something creative for 4 hours that week or do nothing creative that week, even if they normally would. Then they asked the participants to rate their well-being. Their results are shown in Figure 7.2.

FIGURE 7.3

Ms. Lee's Class Results Testing Creativity and Well-Being





a **Does the data support the class's idea that creativity improves well-being? Why or why not?**

b **What elements of the study design could be improved? Explain how changing these elements would improve your confidence in the study results.**

c **If you were going to run this experiment in your class, would you change the study design? If yes, how would you change it? If no, why not?**

Level 4 response

a *The data does support their idea because almost all the students who spent time being creative rated their well-being much higher than those who didn't spend time being creative.*

b *There are several elements that could be improved to increase my confidence in the study results. The sample size is small, and a bigger sample size would improve the study because if the results are similar with a bigger sample size, it's less likely that the results are from chance. Also, they didn't assign people randomly to groups, which means that students might have chosen something they prefer to do, and that could have affected their well-being. Third, they didn't get the participants' well-being ratings before the experiment, so they don't have any way to know if there is really an effect. Maybe the people with higher ratings already had high ratings (and vice versa).*

c *Yes, I would change the study design by assigning people randomly to groups. I would also want a bigger sample size, but my class isn't very big, so I might see if I could get people from a couple of other classes to participate. I would also want to know what their well-being rating was before they started, so I could see if there really is an effect.*

Level 3 response

a *The data does support their idea because the students who spent time being creative had a higher well-being.*

b *There are elements that could be improved in the study design to increase confidence in the results. The sample sizes were small in both studies, and they didn't assign people randomly. This means their results might not be accurate.*

c *I would assign people to groups randomly.*

Level 2 response

a *The data does support their idea that being creative improves well-being.*

b *The sample size was small, and they didn't assign people randomly.*

c *More people.*

Level 1 response

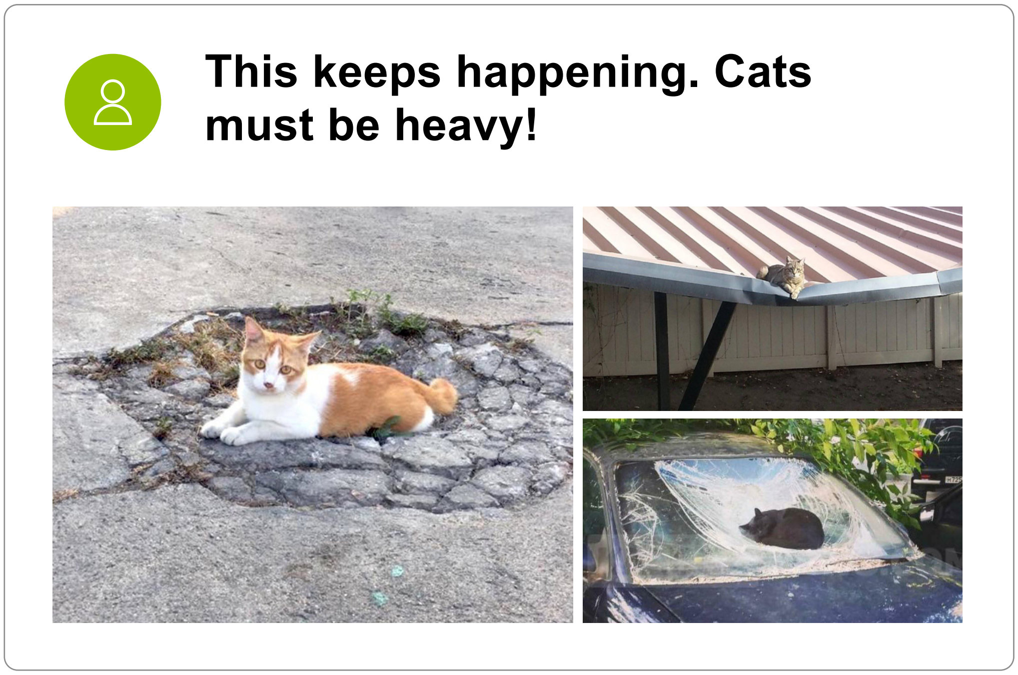
a *The data supports being creative.*

b *There was a small sample size.*

c *I would change it.*

CONNECTIONS TO EVERYDAY LIFE

④ **A friend claims, "Those cats are crushing things!" and sends you the following social media post:**



Explain why this post does not demonstrate causation. In your response, include sample size and at least two of the four key questions about causation.

This meme doesn't show causation because the sample size is too small, and there is an alternative explanation. There are only three cats, and that's not enough to prove an association exists. Just because the cats are settled into those crushed spots doesn't mean that they caused them. There isn't a likely mechanism for how the cats could have caused the damage. Also, we don't know the timing, so an alternative explanation is that the structures were already crushed before the cats sat there. So, we can't say the cats caused the damage based on just these three examples.

5. **Why is it sometimes difficult to get a really large sample size for an experiment with people as participants? Provide a few examples.**

It may be difficult if you have to take time to recruit and collect data from all the participants, and the researcher has limited time. Or if participants need to be paid, and there is limited money. Or if you want to study a specific demographic or disease, and there are not many people who fit the criteria for the study. Or if it requires an unpleasant treatment, and not many people are willing to try it even if it might ultimately help them.

REFERENCES

Fuller, S. [@Fuller_Si]. "This keeps happening. How heavy are cats?" X, April 30, 2021, https://x.com/fuller_si/status/1388106214324662272?mx=2

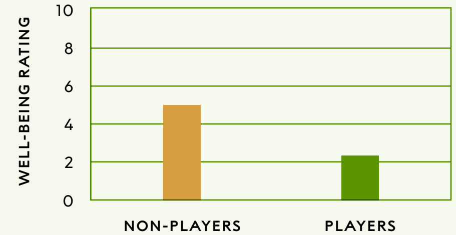
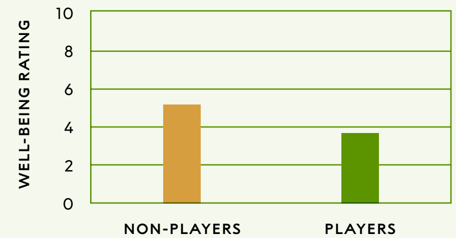
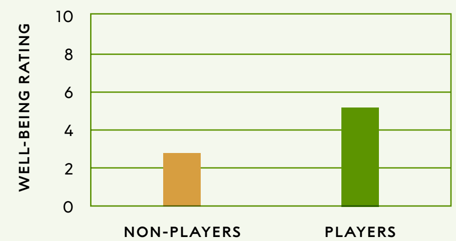
STUDENT SHEET 7.1 | COMPARISON OF STUDY DESIGNS | NAME

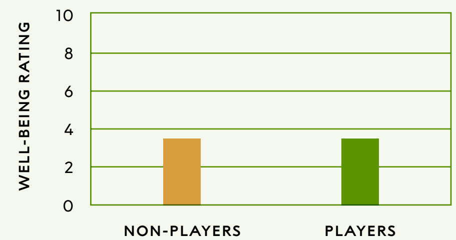
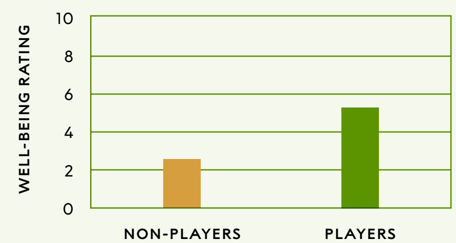
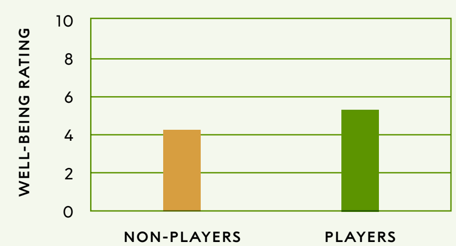
STUDY DESIGN	SAMPLE SIZE	EXPERIMENT GROUPS		ASSIGNMENT TO GROUP BY circle one	RESULTS draw bar chart	INTERPRETATION circle one of the bold words
		EXPERIMENTAL circle one	CONTROL circle one			
1		ONCE ONCE A DAY FOR A WEEK	NOT PLAYING SIDELINES: WATCHING	NEIGHBORHOOD COIN FLIP GRADE	WELL-BEING RATING: 10, 8, 6, 4, 2, 0 NON-PLAYERS PLAYERS	Experimental group had **LOWER** **EQUAL** **HIGHER** well-being compared to the control group.
2		ONCE ONCE A DAY FOR A WEEK	NOT PLAYING SIDELINES: WATCHING	NEIGHBORHOOD COIN FLIP GRADE	WELL-BEING RATING: 10, 8, 6, 4, 2, 0 NON-PLAYERS PLAYERS	Experimental group had **LOWER** **EQUAL** **HIGHER** well-being compared to the control group.
3		ONCE ONCE A DAY FOR A WEEK	NOT PLAYING SIDELINES: WATCHING	NEIGHBORHOOD COIN FLIP GRADE	WELL-BEING RATING: 10, 8, 6, 4, 2, 0 NON-PLAYERS PLAYERS	Experimental group had **LOWER** **EQUAL** **HIGHER** well-being compared to the control group.

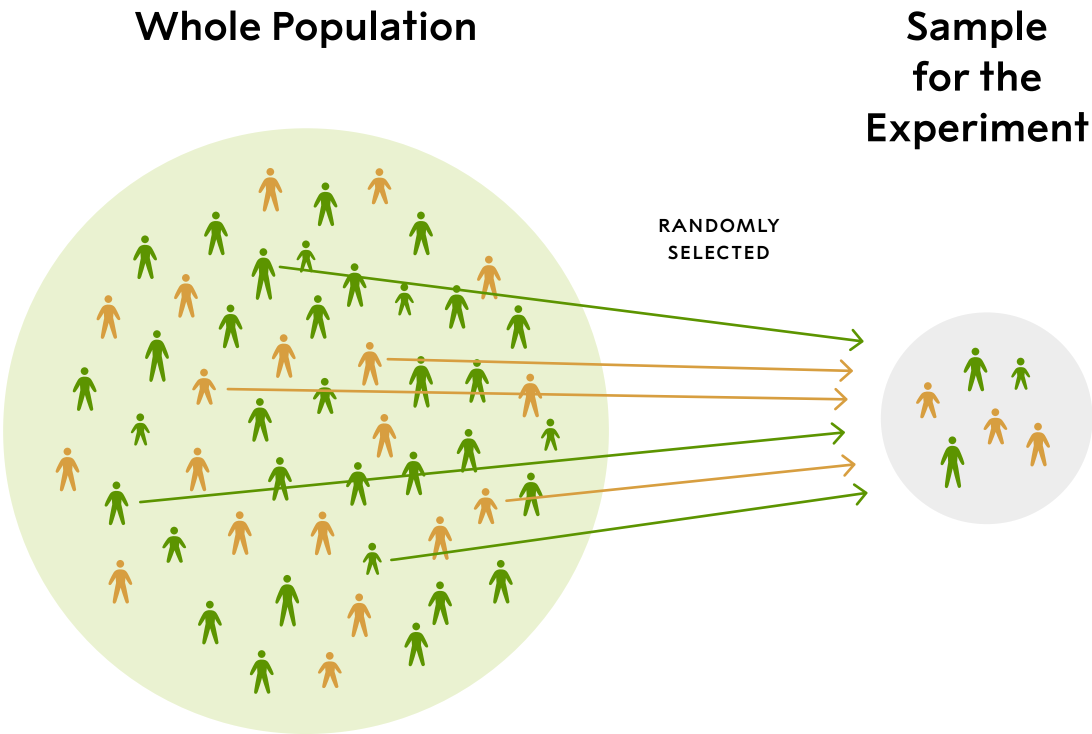
STUDENT SHEET 7.1 | COMPARISON OF STUDY DESIGNS CONTINUED | NAME

STUDY DESIGN	SAMPLE SIZE	EXPERIMENT GROUPS: EXPERIMENTAL circle one	EXPERIMENT GROUPS: CONTROL circle one	ASSIGNMENT TO GROUP BY circle one	RESULTS draw bar chart	INTERPRETATION circle one of the bold words
4		ONCE ONCE A DAY FOR A WEEK	NOT PLAYING SIDELINES: WATCHING	NEIGHBORHOOD COIN FLIP GRADE	WELL-BEING RATING 0 2 4 6 8 10 NON-PLAYERS PLAYERS	Experimental group had **LOWER** **EQUAL** **HIGHER** well-being compared to the control group.
5		ONCE ONCE A DAY FOR A WEEK	NOT PLAYING SIDELINES: WATCHING	NEIGHBORHOOD COIN FLIP GRADE	WELL-BEING RATING 0 2 4 6 8 10 NON-PLAYERS PLAYERS	Experimental group had **LOWER** **EQUAL** **HIGHER** well-being compared to the control group.
6		ONCE ONCE A DAY FOR A WEEK	NOT PLAYING SIDELINES: WATCHING	NEIGHBORHOOD COIN FLIP GRADE	WELL-BEING RATING 0 2 4 6 8 10 NON-PLAYERS PLAYERS	Experimental group had **LOWER** **EQUAL** **HIGHER** well-being compared to the control group.

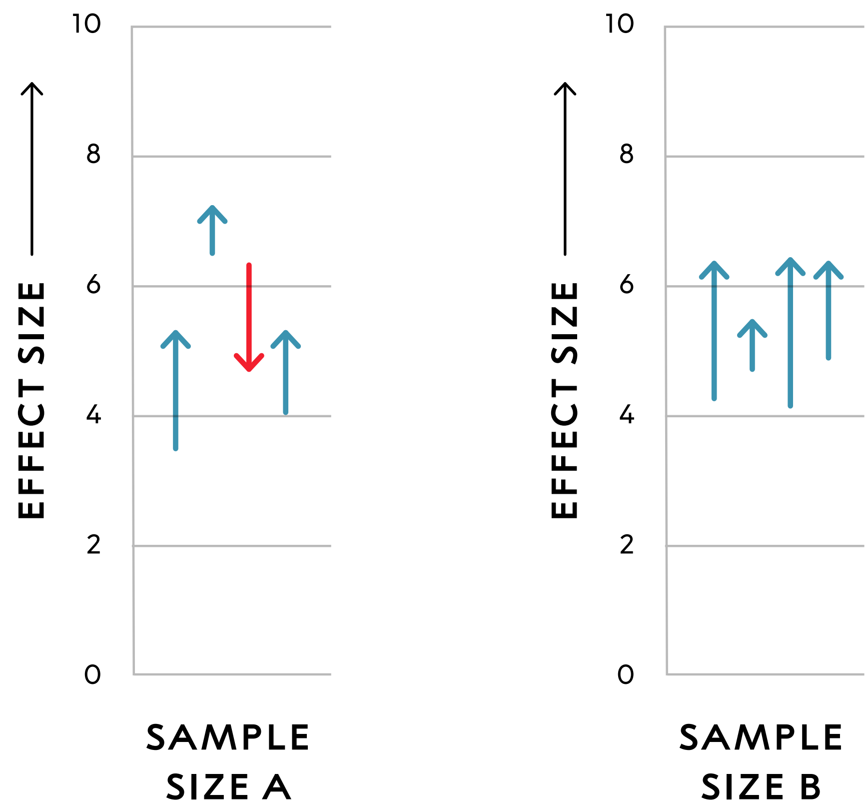
STUDENT SHEET 7.1 | COMPARISON OF STUDY DESIGNS | NAME *SAMPLE STUDENT RESPONSE*

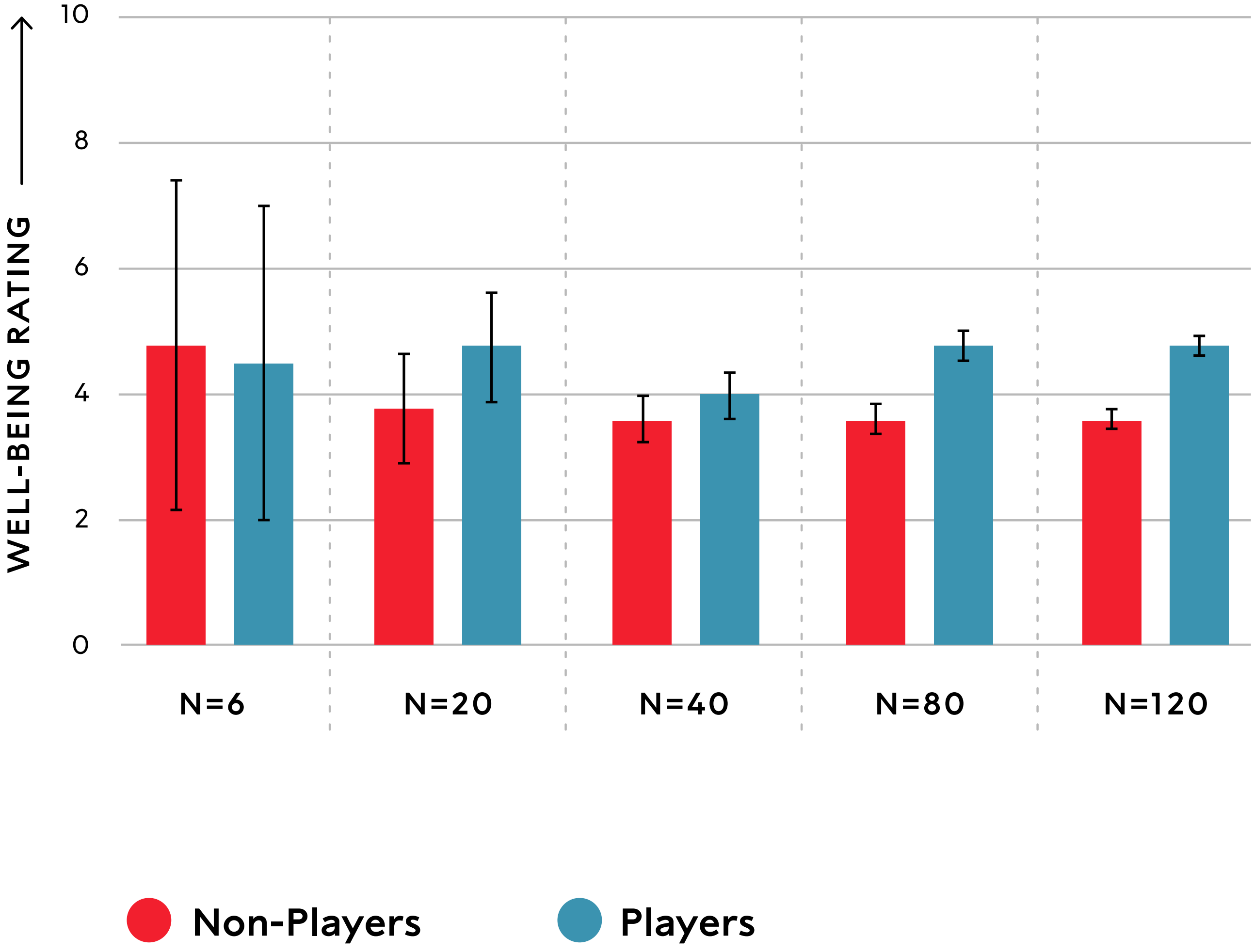
STUDY DESIGN	SAMPLE SIZE	EXPERIMENT GROUPS: EXPERIMENTAL circle one	EXPERIMENT GROUPS: CONTROL circle one	ASSIGNMENT TO GROUP BY circle one	RESULTS draw bar chart	INTERPRETATION circle one of the bold words
1	*50*	**ONCE** ONCE A DAY FOR A WEEK	**NOT PLAYING** SIDELINES: WATCHING	*confound* ↓ **NEIGHBORHOOD** COIN FLIP GRADE		Experimental group had **LOWER** EQUAL HIGHER well-being compared to the control group.
2	*120*	ONCE **ONCE A DAY FOR A WEEK**	**NOT PLAYING** SIDELINES: WATCHING	*confound* ↓ **NEIGHBORHOOD** COIN FLIP GRADE		Experimental group had **LOWER** EQUAL HIGHER well-being compared to the control group.
3	*80*	ONCE **ONCE A DAY FOR A WEEK**	NOT PLAYING **SIDELINES: WATCHING** ↑ *confound*	NEIGHBORHOOD **COIN FLIP** GRADE		Experimental group had LOWER EQUAL **HIGHER** well-being compared to the control group.

STUDY DESIGN	SAMPLE SIZE	EXPERIMENT GROUPS: EXPERIMENTAL (circle one)	EXPERIMENT GROUPS: CONTROL (circle one)	ASSIGNMENT TO GROUP BY (circle one)	RESULTS (draw bar chart)	INTERPRETATION (circle one of the bold words)
4	*8*	**ONCE** ONCE A DAY FOR A WEEK	**NOT PLAYING** SIDELINES: WATCHING	NEIGHBORHOOD **COIN FLIP** GRADE		Experimental group had LOWER **EQUAL** HIGHER well-being compared to the control group.
5	*120*	ONCE **ONCE A DAY FOR A WEEK**	NOT PLAYING **SIDELINES: WATCHING** ↑ *confound*	NEIGHBORHOOD COIN FLIP **GRADE** ↑ *confound*		Experimental group had LOWER EQUAL **HIGHER** well-being compared to the control group.
★ 6	*120*	ONCE **ONCE A DAY FOR A WEEK**	**NOT PLAYING** SIDELINES: WATCHING	NEIGHBORHOOD **COIN FLIP** GRADE		Experimental group had LOWER EQUAL **HIGHER** well-being compared to the control group.


Whole Population
Sample for the Experiment
RANDOMLY SELECTED

Which sample size, A or B, shows more consistent results?




WELL-BEING RATING
10
8
6
4
2
0
N=6
N=20
N=40
N=80
N=120
Non-Players
Players

SCORING GUIDE: ANALYZING AND INTERPRETING DATA (AID)

WHEN TO USE THIS SCORING GUIDE:

This Scoring Guide is used when students analyze and interpret data that they have collected or that has been provided to them.

WHAT TO LOOK FOR:

- Response describes patterns and trends in data.
- Response interprets patterns and trends to describe possible causal relationships.

LEVEL	GENERAL DESCRIPTION
Level 4 **Complete and correct**	The student analyzes the data with appropriate tools, techniques, and reasoning. The student identifies and describes patterns in the data and interprets them completely and correctly to identify and describe relationships. When appropriate, the student: • makes distinctions between causation and correlation. • states how biases and errors may affect interpretation of the data. • states how study design impacts data interpretation.
Level 3 **Almost there**	The student analyzes the data with appropriate tools, techniques, and reasoning. The student identifies and describes patterns in the data BUT incorrectly and/or incompletely interprets them to identify and describe relationships.

LEVEL	GENERAL DESCRIPTION
Level 2 **On the way**	The student analyzes the data with appropriate tools, techniques, and reasoning. The student identifies and describes, BUT does not interpret, patterns and relationships.
Level 1 **Getting started**	The student attempts to analyze the data BUT does not use appropriate tools, techniques and/or reasoning to identify and describe patterns and relationships.
Level 0 **Missing or off task**	The student's analysis is missing, illegible, or irrelevant to the goal of the investigation.
X	The student had no opportunity to respond.

WHEN TO USE THIS SCORING GUIDE:

This Scoring Guide is used when students analyze and interpret data that they have collected or that has been provided to them.

WHAT TO LOOK FOR:

- Response describes patterns and trends in data.
- Response interprets patterns and trends to describe possible causal relationships.

LEVEL	GENERAL DESCRIPTION	ITEM-SPECIFIC DESCRIPTION
Level 4 **Complete and correct**	The student analyzes the data with appropriate tools, techniques, and reasoning. The student identifies and describes patterns in the data and interprets them completely and correctly to identify and describe relationships. When appropriate, the student: • makes distinctions between causation and correlation. • states how biases and errors may affect interpretation of the data. • states how study design impacts data interpretation.	The student response: • describes the data as supportive and provides thorough reasoning. • lists and explains at least three limitations for the study. • lists and explains at least two changes they would make to the study design.

LEVEL	GENERAL DESCRIPTION	ITEM-SPECIFIC DESCRIPTION
Level 3 **Almost there**	The student analyzes the data with appropriate tools, techniques, and reasoning. The student identifies and describes patterns in the data BUT incorrectly and/or incompletely interprets them to identify and describe relationships.	The student response: • describes the data as supportive and provides reasoning. • lists and explains at least two limitations for the study. • lists and explains at least one change they would make to the study design. The student response may have minor errors or limited responses related to: • reasoning. • limitations for studies.
Level 2 **On the way**	The student analyzes the data with appropriate tools, techniques, and reasoning. The student identifies and describes, BUT does not interpret, patterns and relationships.	The student response: • describes the data as supportive but provides limited or no reasoning. • lists at least two limitations but may not explain limitations. • lists at least one change they would make to the study design but may not explain. The student response may have several minor errors.

LEVEL	GENERAL DESCRIPTION	ITEM-SPECIFIC DESCRIPTION
Level 1 **Getting started**	The student attempts to analyze the data BUT does not use appropriate tools, techniques, and/or reasoning to identify and describe patterns and relationships.	The student response: • describes the data as supportive but provides illogical or no reasoning. • lists at least one limitation but may not explain. • states if they would change the study or not but may not list or explain changes. The student response may have several errors.
Level 0 **Missing or off task**	The student's analysis is missing, illegible, or irrelevant to the goal of the investigation.	
X	The student had no opportunity to respond.	