

## **Instinct and Second-guessing Oneself on Tests**

**Bethany Bennett and Nicole Lafser**

**Lindenwood University**

*When taking exams, students may find themselves deciding between the answer that first comes to mind, and a second logical answer. When given the option to record both first instinct and second-guess answers, are students able to answer correctly more on their first or second try? Furthermore, do those students who tend to answer correctly on their second try have lower grade point averages? If so, this would indicate that when taking tests, they have not been able to fully represent their knowledge. In the present study, 16 participants read an informative passage and completed a series of ten open-ended questions regarding the passage. Students tended to answer correctly more on their first instinct. This implies that students should either stick to their first instinct, or that the study was not representative due to the small sample size. Overall, students answered incorrectly most of the time, indicating the difficulty of the test. There were no significant findings when correlating GPA with performance on the test.*

How often have students taken a multiple-choice test and found themselves stuck deciding between two plausible answers? A particular answer may seem correct at first, but upon further inspection, a second one seems just as likely. Test-taking tips and strategies have often declared that students should stick to their instinctive answers. Who came up with this commonly taught strategy, and what research has been done? The purpose of this study is to shed light on this idea of first instinct on exams. Should students stick with their gut, or put some extra thought into their answers? Furthermore,

this study will compare performance on an experimental test and GPA in order to determine if students with a low GPA tend to answer correctly with further thought rather than instinct. Searching for previous studies has yielded few dependable results, although countless sources proclaimed that students should go by their first instinct.

The concept of sticking to instinct is the dominant belief among college students and faculty (Benjamin, Cavell & Shallenberger, 1984). Though this belief is common, no scientific sources were found defending or supporting it. Empirical evidence suggests that students should not stick to their first instinct. Studies indicate that in the long run, changing answers is beneficial. Benjamin and his colleagues reviewed 20 studies on this issue, and found that answer changes that went from the wrong answer to the right answer outnumbered changes that went from the right answers to the wrong answer by a sizable margin.

Furthermore, three points are typically gained for every one point lost by changing an answer (Geiger, 1991). Knowing the truth regarding this matter is important because it has been found that “test-wise” students outperform those who are not “test-wise” (Rogers & Bateson, 1994; Towns & Robinson, 1993). Test-wiseness is defined as the ability to use knowledge of the characteristics of tests and the testing process to improve one’s performance (Millman, 1966).

Although there is a significant amount of hearsay that test-takers should go by their first instinct, empirical evidence continues to state otherwise. The current research is not being conducted to determine if students do better when they change their answers, because research already supports such a hypothesis. The purpose of this research is to determine how often students answer correctly when given the opportunity to state their

second answer without penalty, and if the frequency of this occurrence correlates with their GPA. Thus, it was predicted that not only will participants correctly answer open-ended questions on their second try; participants who have a high amount of correct second try answers will also have a low GPA. This hypothesis was conceived because previous studies support the idea that test-takers perform better when they change their answers, and go with a second instinct. Furthermore, if participants have a high amount of correct second try answers; it would make sense that they would have a low GPA. It may be the case that on tests they tend to stick with their first instinct, therefore scoring lower by not changing their answers. To test this, participants will read a short informative passage and then be given a series of ten open-ended questions. Participants will first write their instinctive answer. Secondly, participants will consider the question a second time, assuming that their first answer is wrong. They will then write in their second-choice answer.

## Method

### *Participants*

Twenty-five Lindenwood University undergraduates participated in this study, but the data from only 16 participants could be used. Nine participants did not follow the directions given to them for the experiment, so their data was not useful. Participants were recruited in Butler library by use of the verbal script (see Appendix A).

### *Materials*

In order to execute the research as designed, the participants were asked to sit at tables in Butler Library. Butler Library was chosen because of the similar atmosphere to that of a classroom. Basic materials used included a desk for the participants to sit at, a

chair for the participants to sit on, a pencil for the participants to write with, and appropriate documents for each individual participant. The packet of appropriate documents included the informed consent, a GPA release form (see Appendix B) instructions on how to perform the experiment (C), the informative passage (see Appendix D), and a question and answer sheet (see Appendix E). These documents were stapled in this order. A feedback letter was also required, and was given to the participants after they finished the experiment. Materials for the experimenter included an expanding file folder to file participant documents into, a pen to code the papers with, a verbal script that was used to recruit participants (see Appendix A), and an acceptable answer sheet (see Appendix F).

### *Procedure*

After hearing the verbal script and volunteering to participate in the research, participants were asked to take a seat at a designated table in the library. After being seated, they were given the packet of documents, and were verbally informed that any questions would be welcomed. First, they signed the GPA release and consent forms. Any questions regarding these documents were answered. Most frequently, participants needed verbal confirmation that their GPA would not be identified with their person. Next, participants read the instructions. The instructions stated that the next page of the packet would include an informative passage that they would need to read. It informed them that they would be taking a ten question test regarding the passage, and the passage could not be referred to after the test began. The instructions also explained the format of the test. It was clearly stated that participants should answer with their first instinct on the first blank, and then assume their answer was wrong and re-answer on the second blank.

If they had no questions, they informed the researcher that they were ready to begin. Once the participants read the informative passage, they turned to the series of ten questions and answered accordingly. After the questions were complete, the packet was returned to the experimenter. At this time, the GPA release form and the answer sheet were coded with a matching number. Furthermore, the consent forms were detached from the packet and filed into an expanding file folder. GPA release forms, consent forms, and answer sheets each had their own respective areas.

In order to score the “tests” we considered each answer. If the answer was correct, we placed an “R” next to it. If the answer was incorrect, we placed a “W” next to it. The following three combinations were possible: “RW,” “WR,” and “WW.” “RW” would indicate that the participant answered correctly by their first instinct. “WR” would indicate that the participant answered incorrectly on instinct, but with further consideration, they answered correctly. Clearly, a “WW” indicates that they could not recall the correct answer.

### Results

It was hypothesized that participants would achieve more “WR” results than “RW” results. Furthermore, it was hypothesized that students with higher GPAs would have a higher amount of “RW” answers. Contrastingly, students with lower GPAs would have a higher amount of “WR” answers. On average, 3.87 questions were answered correctly on the first try. On the second try, 1.56 questions were answered correctly. The maximum number of questions remembered correctly on the second try was six, and this circumstance occurred twice. On average, 4.56 questions were answered incorrectly

(“wrongness”). When observing “rightness”, or the occurrence of achieving a correct answer on either the first or second try, 5.44 questions were answered correctly.

In order to analyze the findings, Pearson correlations were done, comparing GPA with scores. The Pearson correlation ( $r$ ) for “WR” and GPA was .209, with a significance of .219. When correlating GPA with total wrongness,  $r$  was found to be .448, with significance of .035. For GPA and the number of “RW”,  $r$  was -.404 and significance was .060. The GPA and adjusted score correlation was -.196 and significance level was .233. Adjusted score was found by grading tests in such a way that 10 points were gained for each “RW” and .5 points were gained for each “WR”. Graphical representations of findings can be found in Figures 1 – 4.

**Figure 1. GPA and number of “right-wrong” responses**

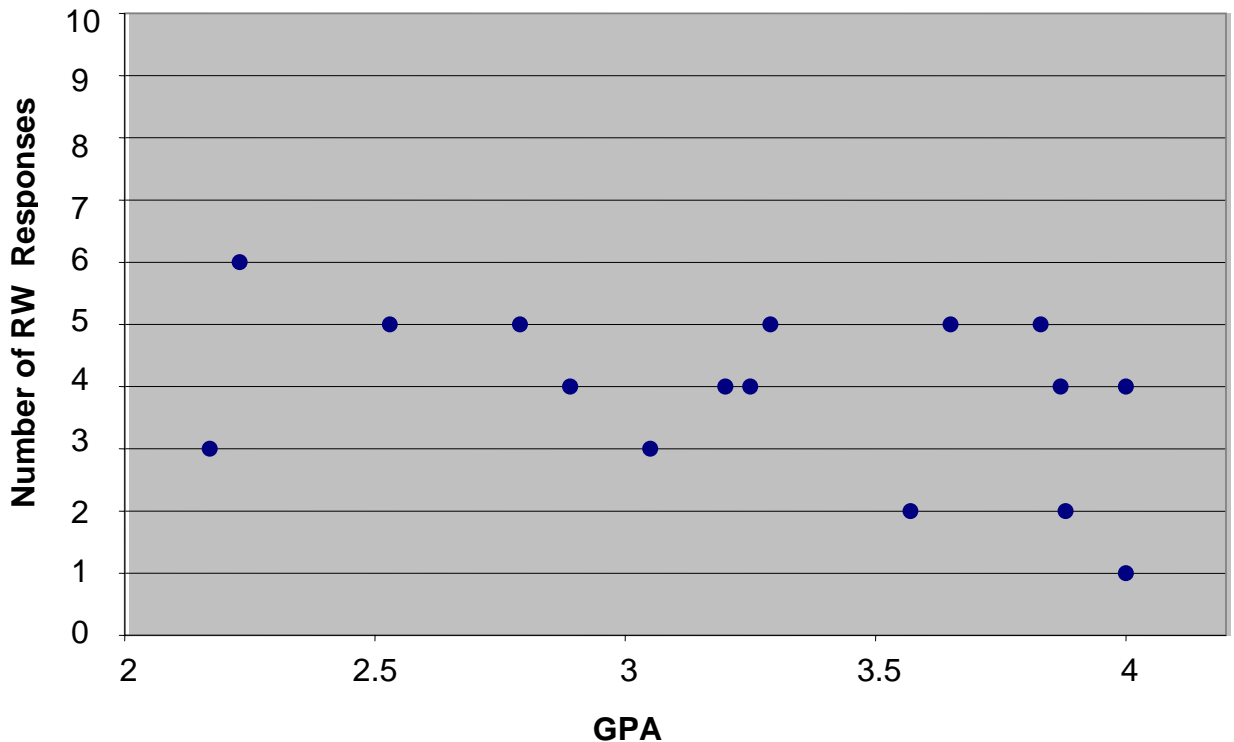


Figure 2. GPA and number of “wrong-right” responses

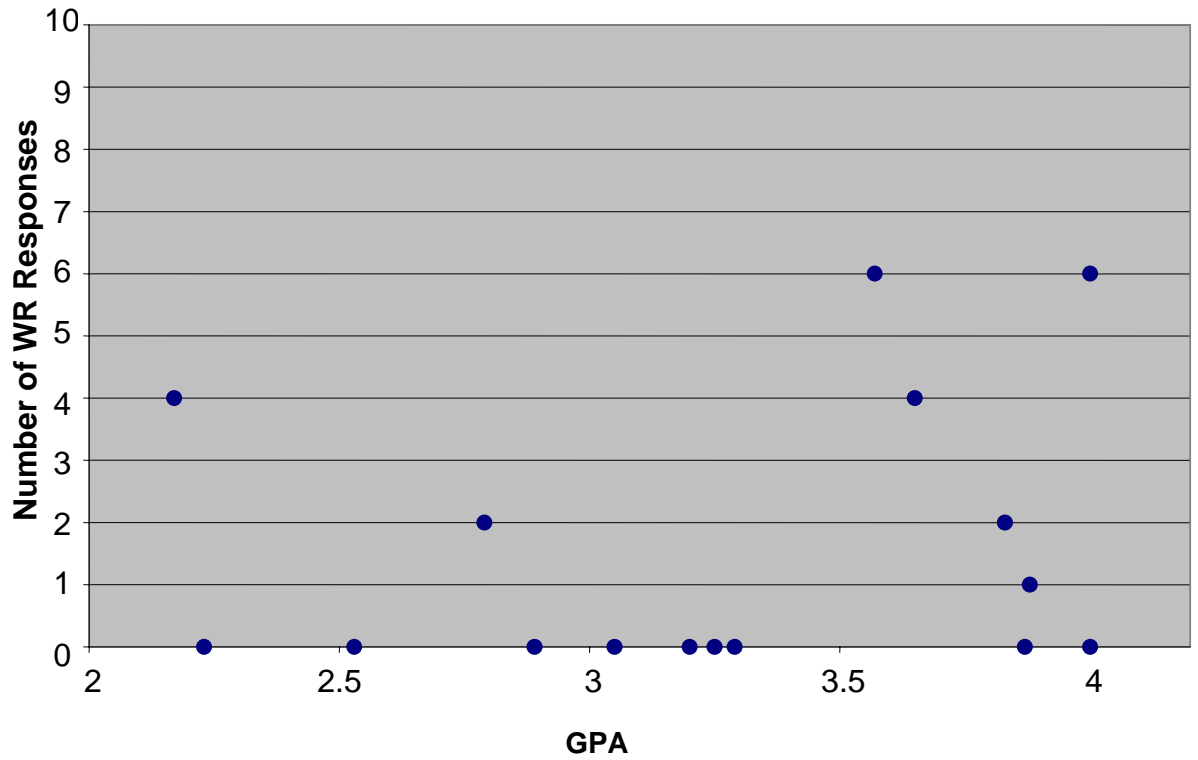


Figure 3. GPA and number of wrong responses

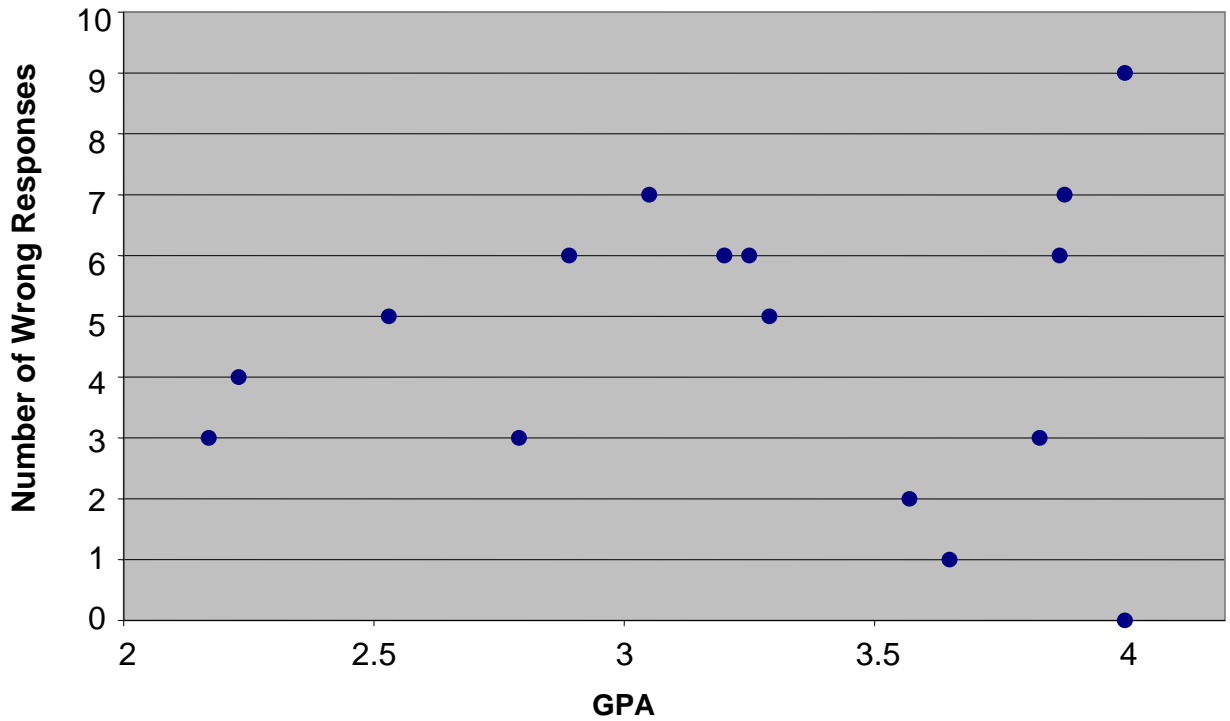
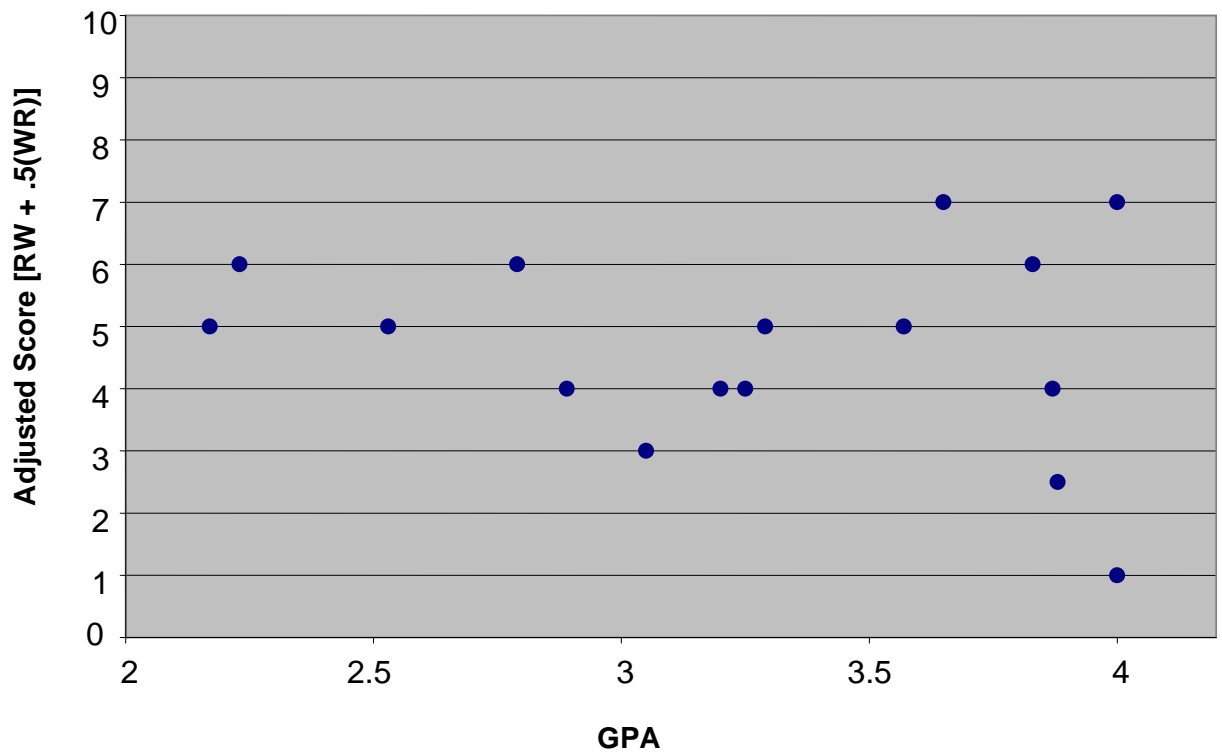


Figure 4. GPA and number of “wrong-right” responses



#### Discussion

Statistically, not enough data was gathered in order for this finding to be conclusive, but for the purposes of this research, the findings were accepted. Both of our hypotheses were found to be incorrect. We really think that this is because we did not get enough participants; especially because of the odd findings when looking at GPA. Analysis of our data indicated no statistically significant findings when correlating GPA. The most significant correlation was found between the number of wrong answers and GPA. The higher a persons GPA, the more questions they answered incorrectly.

One of the most troubling circumstances that we ran into was due to the instructions. We concluded that most participants did not actually read the instructions thoroughly, perhaps because they thought that when they did the experimental test, it



would be self-explanatory, as more tests are the same. Participants may not have realized that our test was to be set up differently. We came to this conclusion because of the way that participants answered the questions with the same answer twice, although our instructions clearly indicated that the answer should be different from the first. In order to correct this, we should have read the instructions aloud, and asked for questions before the person began in order to resolve these issues. This way, there would be no chance of participants skimming over the instructions, and we would know that they had been exposed to the instructions in their entirety.

A notable occurrence occurred in participants who did not follow the directions. Those who repeated their instinct answer, though instructed to assume their first answer was wrong, obviously may have strongly felt that their first instinct answer was correct. We can draw this conclusion because they did not repeat their answer on all of the questions, indicating that when they chose two different answers they were utilizing the ability to guess twice, and have a higher probability of being correct. Though they strongly believed that their answer was correct, a noticeable amount of the time they were actually wrong.

This leads to the next issue, the way the data was analyzed. We did not take into account that the answers that were wrong both times, or right both times, because of participants answering the same answer twice. Instead, we dropped that data, which is why we had such an insignificant amount of participant data to analyze. The insignificance of our results probably stems from the fact that in real life people actually find their instincts and logic to be either right or wrong at the same time. Not accounting for this is crucial because it forces us see only one part of a much greater picture.

Furthermore, many studies have been conducted that support our hypothesis, but our results indicated the opposite. We are accepting our results, but attributing this to the low number of participant data.

During and after the experiment, we got several complains that the questions were much too hard, and that they should have been changed. Moreover, some participants found even the idea of our experiment too difficult, and decided not to participate after merely reading the instructions. It may have been too difficult. However, our participants were also not provided with any incentives, which would explain for their lack of trying. An interesting finding was that we had to students with cumulative GPA's of 4.0. One student got all of the answers right, either on his or her first or second-try. The other student got nine wrong out of ten. We assume that these participants are typically hard working; therefore, the difference in their score may be due to the lack of incentive.

In future experiments, we might consider giving participants a timed test first. This investigation would involve testing quickly answered questions (instincts) versus answers achieved with an unlimited amount of time (logic). These times would be a significant asset to us because they would help show the nature of the answers. In addition to changing the answers, we would also change the entire test. Rather than giving them a single test we would give them two tests. Though different, they would be similar in order to measure both instinct and logic on separate levels.

The final and probably least controversial problem of the entire study was the environment in which the participants had to take the test. All the participants were tested within Butler Library. While the library is quiet, as a testing atmosphere would be,

there is also an immense amount of traffic most of the time, specifically between class periods. Those who did poorly may have experienced interference from their friends being in the same vicinity, or other traffic in general. This distraction could have been as simple as glancing up and looking at someone during the reading. Furthermore, they may have been in a hurry and may have wanted to get back to their studies in the library.

Though briefly mentioned already, in order to achieve a more significant and critical value, we would definitely need more participants. Of course, having more people typically leads to more reliable results.

If it could be done without breaking confidence, future experimenters may want to try checking test scores in an actual classroom setting. This could show a more realistic account of what happened in the student's mind while taking a test. For example, if the student drew an 'x' over an answer then chose another, experimenters know that they had an issue with their first instinct and then second-guessed themselves. If they additionally put an 'x' over their new answer, and decided to revert to their first answer by making a note to the instructor, this shows even more consideration of the answers. Finally, we would like to see a person take a test for what they believe to be a grade in their class. This way, a true testing environment would be created. The conflict that arises when taking tests when students second-guess their instinctive answers may be more likely to occur if this testing environment is created.

References

- Benjamin, L.T., Cavell, T.A., & Shallenberger, W.R. (1984). Staying with initial answers on objective tests: Is it a myth? *Teaching of Psychology, 11*, 133-141.
- Geiger, M.A. (1991). Changing multiple-choice answers: Do students accurately perceive their performance? *Journal of Experimental Education, 59*(1), 119-127.
- Millman, J., Bishop, C. H., & Ebel, R. (1965). An analysis of testwiseness. *Educational and Psychological Measurement, 25*, 707-726.
- Rogers, W.T., & Bateson, D.J. (1994). Verification of a model of test-taking behavior of high school seniors. [Special issue: Cognition and assessment.] *Alberta Journal of Educational Research, 40*, 195-211.

## **Appendix A**

### **Verbal Script**

My name is (Bethany/Nicole) and I am here to recruit participants for our study. The study should not take more than 10-15 minutes of your time. In fact, it should typically take less depending on how fast you read. You will be asked to answer a series of questions after reading a short story. We are not interested in your individual results, but instead the results of a large group of participants. No identifying information about you will be associated with our findings. Our study is not only evaluating the idea of instinct versus second-guessing on exams, but the findings of our study also have the potential to elicit reason for teachers to consider altering their testing methods, so it should be an exciting study to contribute to. Would you like to participate?



## Appendix C

### Instructions

*Please do not turn to the next page of the packet until instructed to do so.*

The next page of this packet is an informative passage for you to read. After reading the passage, you will need to turn to the next page. **Once you have turned the page you may not turn back and refer to the informative passage.** The page following the informative passage includes a list of 10 open-ended questions. Each question will have two blanks underneath it, with the first blank labeled 1 and the second blank labeled 2. You should read each question and immediately answer in blank 1. Answer on your first instinct, placing the first answer that comes to mind in this 1<sup>st</sup> blank. Remember, FIRST blank = FIRST instinct.

Next, consider the question again, assuming that your first choice answer was wrong, and answer the question again in blank 2. Please answer seriously, avoiding nonsense answers. You should write educated, possible answers in blank 2, even if you think that your first choice answer is correct. Remember; ASSUME that your first choice answer is incorrect.

Our intentions in this study are to test the accuracy of participant's first instinct compared to their second guesses or educated guesses. Please keep in mind that once you have read a question you must provide your first instinct answer. *An example question is provided below.* If you have any questions, please ask the researcher now. If you have no questions, inform the researcher that you are ready to begin.

What is the name of the main character?

1) \_\_\_\_\_

2) \_\_\_\_\_

## **Appendix D**

### **Informative Passage**

In 1892 the Sierra Club was formed. In 1908 an area of coastal redwood trees north of San Francisco was established as Muir Woods National Monument. In the Sierra Nevada mountains, a walking trail from Yosemite Valley to Mount Whitney was dedicated in 1938. It is called John Muir Trail.

John Muir was born in 1838 in Scotland. His family name means “moor,” which is a meadow full of flowers and animals. John loved nature from the time he was small. He also liked to climb rocky cliffs and walls. When John was eleven, his family moved to the United States and settled in Wisconsin. John was good with tools and soon became an inventor. He first invented a model of a sawmill. Later he invented an alarm clock that would cause the sleeping person to be tipped out of bed when the timer sounded.

Muir left home at an early age. He took a thousand-mile walk south to the Gulf of Mexico in 1867 and 1868. Then he sailed for San Francisco. The city was too noisy and crowded for Muir, so he headed inland for the Sierra Nevadas.

When Muir discovered the Yosemite Valley in the Sierra Nevadas, it was as if he had come home. He loved the mountains, the wildlife, and the trees. He climbed the mountains and even climbed trees during thunderstorms in order to get closer to the wind. He put forth the theory in the late 1860's that the Yosemite Valley had been formed through the action of glaciers. People ridiculed him. Not until 1930 was Muir's theory proven correct.

Muir began to write articles about the Yosemite Valley to tell readers about its beauty. His writing also warned people that Yosemite was in danger from timber mining and sheep ranching interests. In 1901 Theodore Roosevelt became president of the United States. He was interested in conservation. Muir took the president through Yosemite, and Roosevelt helped get legislation passed to create Yosemite National Park in 1906. Although Muir won many conservation battles, he lost a major one. He fought to save the Hetch Valley, which people wanted to dam in order to provide water for San Francisco. In the late 1913 a bill was signed to dam the valley. Muir died in 1914. Some people say losing the fight to protect the valley killed Muir.



## Appendix E

### Question and Answer sheet

When was the Sierra Club formed?

- 1) \_\_\_\_\_
- 2) \_\_\_\_\_

What is the name of the trail dedicated to John Muir?

- 1) \_\_\_\_\_
- 2) \_\_\_\_\_

Where was John Muir born?

- 1) \_\_\_\_\_
- 2) \_\_\_\_\_

How did Muir's alarm clock wake sleeping persons up?

- 1) \_\_\_\_\_
- 2) \_\_\_\_\_

How long was walk south of the Gulf of Mexico that Muir took?

- 1) \_\_\_\_\_
- 2) \_\_\_\_\_

Why did Muir climb mountains and trees during thunderstorms?

- 1) \_\_\_\_\_
- 2) \_\_\_\_\_

What was Muir's theory pertaining to how the Yosemite Valley had been formed?

- 1) \_\_\_\_\_
- 2) \_\_\_\_\_

What year was Muir's theory about the formation of Yosemite Valley proven correct?

- 1) \_\_\_\_\_
- 2) \_\_\_\_\_

Muir wrote that Yosemite Valley was in danger of timber mining and what one other interest?

- 1) \_\_\_\_\_
- 2) \_\_\_\_\_

When people said, "losing the fight to protect the valley killed Muir", which valley were they referring to?

- 1) \_\_\_\_\_
- 2) \_\_\_\_\_

## **Appendix F**

### **Acceptable Answer Sheet**

When was the Sierra Club formed?

Acceptable Answer(s): 1892

What is the name of the trail dedicated to John Muir?

Acceptable Answer(s): John Muir Trail, John Muir

Where was John Muir born?

Acceptable Answer(s): Scotland

How did Muir's alarm clock wake sleeping persons up?

Acceptable Answer(s): Tipped people out of bed, tipped them

How long was walk south of the Gulf of Mexico that Muir took?

Acceptable Answer(s): One thousand miles, 1,000 miles

Why did Muir climb mountains and trees during thunderstorms?

Acceptable Answer(s): To get closer to the wind

What was Muir's theory pertaining to how the Yosemite Valley had been formed?

Acceptable Answer(s): Glaciers

What year was Muir's theory about the formation of Yosemite Valley proven correct?

Acceptable Answer(s): 1930

Muir wrote that Yosemite Valley was in danger of timber mining and what one other interest?

Acceptable Answer(s): Sheep ranching, sheep

When people said, "losing the fight to protect the valley killed Muir", which valley were they referring to?

Acceptable Answer(s): Hetch valley, Hetch