

# **CONFIDENTIAL - FOR PEER-REVIEW ONLY**

## **The Role of Images for Knowledge Understanding (#68156)**

Created: 06/10/2021 11:29 AM (PT)

This is an anonymized copy (without author names) of the pre-registration. It was created by the author(s) to use during peer-review.  
A non-anonymized version (containing author names) should be made available by the authors when the work it supports is made public.

### **1) Have any data been collected for this study already?**

No, no data have been collected for this study yet.

### **2) What's the main question being asked or hypothesis being tested in this study?**

Are participants going to answer more visual questions correctly after seeing Wikipedia articles with images in comparison to without images?

### **3) Describe the key dependent variable(s) specifying how they will be measured.**

The key dependent variable is the number of correct answers.

Number of correct answers is defined by how many questions are correct out of 9 questions for each participant. We will generate the multiple choice question for each Wikipedia article.

### **4) How many and which conditions will participants be assigned to?**

This experiment will be a within-subjects design. It comprises of the following factors and levels:

1. Type of multiple choice questions. For each article the survey taker will be asked 3 questions (image, visual, non-visual). For image questions the survey taker will be asked to pick the correct image depicting the article topic. (e.g. Which is a Monet painting?) For visual questions the survey taker will be asked about a visual aspect of the article topic. (e.g. How many minarets does the Taj Mahal have?) For non-visual questions the survey taker will be asked about a non-visual aspect of the article topic. (e.g. In what city is the Taj Mahal located?)
2. Text length. The number of words in all three articles combined.
3. Familiarity (1, 2, 3, 4, 5). The understanding of the topic of Wikipedia article topic (1, no experience - 5, expert)
4. Participant id.
5. Article id.
6. Native Language. Native Language of the survey taker.
7. Country. Country that the survey taker is from.
8. Education Level.
9. Image presence. Each participant will randomly see/not see the main image of the article topic for each Wikipedia article that they read.

### **5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.**

We will be using a mixed linear model (lmer in the lme4 package in R). The number of correct answers will be predicted by the participant id (Random), article id (Random), condition (with|without images), time, education level, native language, country, familiarity with topic, and text length.

### **6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.**

We would exclude participants who indicate that they cheated on the survey if they indicate so.

We would exclude participants who don't finish the survey.

We would exclude participants who spend more than 1 hour on the survey.

### **7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.**

We would like to recruit at least a few hundred participants.

### **8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)**

1. We will choose Wikipedia articles for the survey based on the categories, such as biology, architecture, food, etc. Each participant will randomly see 3 articles of these articles during the experiment.
2. We are also generating multiple choice questions for these articles.
3. To answer whether images make all questions better, we are going to conduct an additional linear regression analysis.
4. To answer whether images make all questions better, we are going to conduct an additional linear regression analysis.
5. To answer whether images make non-visual questions better, we are going to conduct an additional linear regression analysis.
6. To answer whether images make visual questions better, we are going to conduct an additional linear regression analysis.
7. To answer whether images will change how many questions someone answers correctly, we are going to conduct an additional linear regression analysis.
8. To answer whether images that are isolated are better/worse or images that are colorful are better/worse in helping people answer more questions correctly, we are going to conduct an additional linear regression analysis.

9. To answer whether text complexity will positively influence the helpfulness of images for answering questions, we are going to conduct an additional linear regression analysis.

10. To answer whether elderly participants will be better at answering questions in the text only condition than younger participants, we are going to conduct an additional linear regression analysis.