Dataset Requirements: We are looking for datasets wherein participants see multiple cognitive or social stimuli (loosely defined), and the data is preserved with each original item-answer. For example, you may have participants rate 25 items on their pleasantness. If the data contains each rated item for each participant (i.e., not averaged across items), this data would be an appropriate dataset for our project. Note that it does not have to be your data, but you may know an appropriate dataset that is open source that we can use.

Project/Data Title: Overconfidence for picture cues in foreign language learning

Project/Data Description: (200-500 words brief description of the theory/background for the data)

Previous research shows that participants are overconfident in their ability to learn foreign language vocabulary from pictures compared with English translations. The current study explored whether this tendency is due to processing fluency or beliefs about learning. Using self-paced study of Swahili words paired with either picture cues or English

translation cues, picture cues garnered higher confidence judgements but not faster study times, and this was true whether judgements of learning were made after a delay (Experiment 1).

Methods Description: (brief description of how the data was collected)

Forty-two English-Swahili word pairs from Carpenter and Olson’s (2012) Experiment 2 were used. The English words were one-syllable nouns, ranging between three and six letters, with an average concreteness rating of 4.86 on a 5-point scale (SD = .16) (Brysbaert, Warriner, & Kuperman, 2014), and an average frequency of 106.52 per million (SD = 113.40) (Brysbaert & New, 2009).

Participants began the experiment with instructions informing them that they would be learning Swahili words paired with either pictures or English translations as cues. To illustrate each type of cue, they were given an example of an item (Train: Reli) that was not included among the 42 experimental items. They were informed that each pair of items (English-Swahili pairs or picture-Swahili pairs) would be presented one at a time, and they would have as much time as they needed to study it. Participants were encouraged to do their best to learn each pair, and to encourage full and meaningful processing of each, they were instructed to press the spacebar once they felt they had fully “digested” it. For each participant, 21 items were randomly selected to be presented as English-Swahili pairs, and 21 as picture-Swahili pairs. Participants saw each stimulus pair one at a time, in a unique random order with English-Swahili pairs and picture-Swahili pairs intermixed. Each pair was presented in the center of the computer screen and remained on screen until participants pressed the spacebar to move on to the next pair. After each of the 42 pairs was presented for self-paced study in this way, the same pairs were presented again for JOLs. During a JOL trial, each cue-target pair was presented on the screen and participants were asked to estimate—using a scale from 0% (definitely will NOT recall) to 100% (definitely will recall)—the likelihood of recalling the Swahili word from its cue (either the picture or English translation) after about 5 minutes. Participants entered a value between 0 and 100 and pressed the ENTER key to advance to the next item.

Data Location: https://osf.io/2byt9/

Date Published: October 10, 2019

Dataset Citation:

Carpenter, S. K., & Geller, J. (2020). Is a picture really worth a thousand words? Evaluating contributions of fluency and analytic processing in metacognitive judgements for pictures in foreign language vocabulary learning. Quarterly Journal of Experimental Psychology, 73(2), 211–224. https://doi.org/10.1177/1747021819879416

Keywords: [Overconfidence](https://journals.sagepub.com/keyword/Overconfidence), [metacognition](https://journals.sagepub.com/keyword/Metacognition), [processing fluency](https://journals.sagepub.com/keyword/Processing+Fluency), [analytic processing](https://journals.sagepub.com/keyword/Analytic+Processing), [foreign language learning](https://journals.sagepub.com/keyword/Foreign+Language+Learning)

Use License:

Geographic Description - City/State/Country of Participants:

Ames, IA, USA

Column Metadata: Fill in the chart below for each column of data in the dataset. Please note you can filter out columns that are not useful for this project.

Use data from the Exp\_1\_2\_raw\_data.xlsx file

| Variable Name | Variable Description | Type (numeric, character, logical, etc.) |
| --- | --- | --- |
| Experiment | Experiment 1 (1) or 2 (2) ONLY USE 1 |  |
| Subject | Subject ID | Numeric |
| CueType | Whether participant was presented with word translation (1) or word with picture (2) | Numeric |
| Stimulus | Swahili words presented on each trail | Character |
| EncodeJOL | JOL (1-100) 1=not likely to recall 100=very likely to recall | Numeric |

What columns should we use to simulate the data?

| Variable Name | Variable Description | Type (numeric, character, logical, etc.) |
| --- | --- | --- |
| Subject | Subject ID | Numeric |
| CueType | Whether participant was presented with word translation (1) or word with picture (2) | Numeric |
| Stimulus | Swahili words presented on each trail | Character |
| EncodeJOL | JOL (1-100) 1=not likely to recall 100=very likely to recall | Numeric |

Goals: we will use this data to provide examples of our simulation process on how to determine sample size for a project based on item rather than participant. You can read about this idea here: <https://github.com/SemanticPriming/SPAML/blob/master/02_Power/power_aipe.pdf> We will use the example provided in this link as the main portion of the paper and then add your data as a vignette example to supplement the paper. You will be considered an author for completing this template worksheet (no coding skills necessary, we will do that part), and reviewing/commenting on the draft of the paper. Please email [007spaml@gmail.com](mailto:007spaml@gmail.com) if you have questions.