UK Biobank

Picture Vocabulary Test:

Test adaptation, data collection and data recalibration

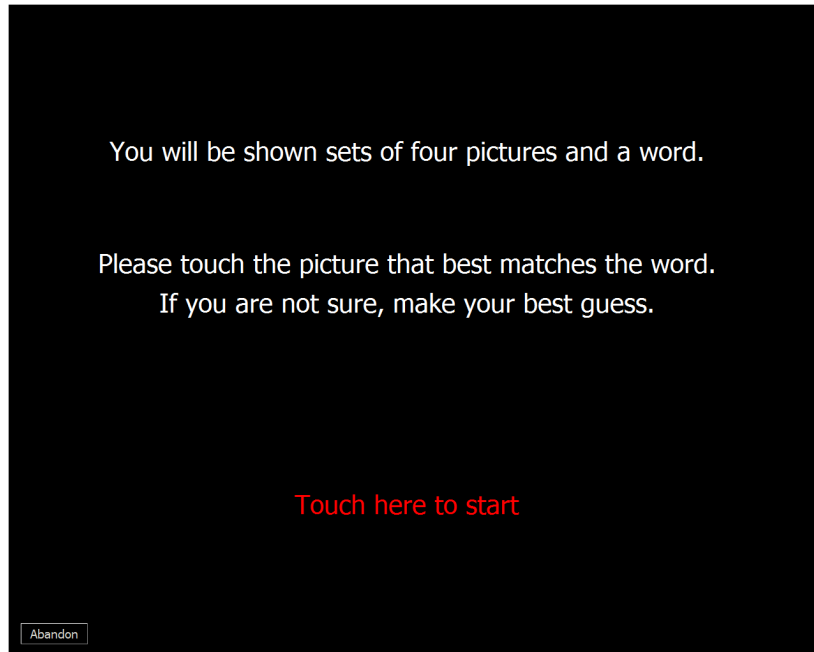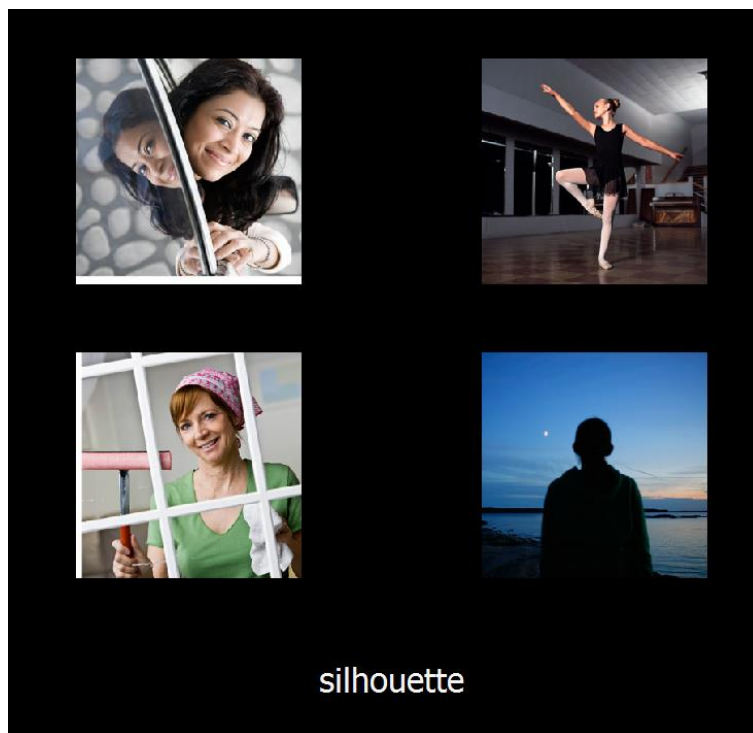Version 1.0

Picture Vocabulary Test

The Picture Vocabulary Test is administered using a touchscreen. The initial instructions appearing on the screen are:

The participant is shown a sequence of at least 20, and at most 30, screens displaying a word and four pictures, for example:

The process may be abandoned at any time using the 'Abandon' button. If the participant presses 'Abandon' they are presented with the message 'Are you sure?'. If they select 'No', they are returned back to the current word with accompanying pictures.

Testing Model

The Picture Vocabulary test is a computer-adaptive test (CAT), using Item Response Theory (IRT) methodology.

At each round, the choice of word to show a participant is based on their responses up to that point. Each word has a difficulty level (within the range -11 to +11), and a participant's vocabulary level is assessed using the same range of difficulty. An estimate is made of a participant's vocabulary level by using a maximum likelihood estimate (MLE), based on their answers to all the words they have been presented with so far.

The algorithmic adaptation made during the test can be summarised as follows:

> 1) *If the participant has got all the words right so far, let the next word be significantly more difficult.*
>
> 2) *Otherwise, calculate (with MLE) the participant's estimated vocabulary level, and select next the word closest to it in difficulty, out of those words not yet attempted.*

The jump in 1) is necessary to quickly narrow in on the participant's vocabulary level.

Each participant is shown at least 20 words to attempt. The test ends after the estimate of the participant's vocabulary is sufficiently accurate (within a standard error of <0.5), or at a maximum of 30 words.

Data Collection

The data collected for a participant as a result of a run of the test is as follows:

- the word presented at each round of the test
- the picture that the participant selected for each round
- how long the participant took to select the picture
- the final estimate (MLE) of the participant's vocabulary level, as calculated at the end of the test
- the standard error (from the MLE calculation) giving the uncertainty of the participant's vocabulary level, as calculated at the end of the test

The vocabulary levels as originally calculated during the test were obtained from difficulty levels calibrated to a US audience (see below), and following initial collection of the data, recalibration was carried out by Adam Hampshire's team at Imperial College, London [2], resulting in the following derived data:

- an estimate of cognitive ability (specific ability) when performing the picture vocabulary task
- estimate of the basic visuomotor processing times of the participant

It is strongly recommended that the recalibrated estimate is used in preference to the original MLE calculation directly from the test.

Source

The UK Biobank Picture Vocabulary Test was adapted from the NIH Toolbox Picture Vocabulary Test [1]. Please note that any publications, journal articles or presentations using data from this Picture Vocabulary Test should acknowledge that this test was a modified version of the NIH Toolbox Picture Vocabulary Tests, amended with permission from NIH Toolbox for use in UK Biobank.

The following outlines the similarities and differences for the UK Biobank version of the test.

The NIH test, for a US audience, was administered using a touchscreen by a member of staff, who also pronounced the words presented on the screen. Thus persons taking the test were presented with both an audio and visual depiction of the word. In contrast, the UK Biobank version of the test was delivered using a touchscreen only, so participants did not experience listening to the words.

The dataset of words with their associated pictures were the same as those used in the NIH test, but with a small number of modifications, due to the differences between US English and UK English:
- Some words were altered to match their UK counterparts, for example "intersection" was replaced with "crossroads".
- Some words were removed; for example "tote" is associated with betting in UK English, not a bag as in US English.
- Some words were removed due to potential ambiguity, due to there being no audio version of the word to disambiguate. For example, the word "minute" was removed, as the picture of a clock could have confused from its intended meaning of "extremely small".

This resulted in a dataset of 340 words in total. The difficulty levels of the words as used in the UK Biobank version were the same as used in the NIH test, and the same algorithm for adaptive testing was used for both tests.

[NB: one of the words, "matron", was included, but subsequently attracted a surprisingly low proportion of correct answers; in US English, "matron" is associated more with the definition referring to an older married woman, unlike UK English which associates more with the definition involving medical professionals; one of the wrong answers was a picture of a nurse.]

However the start and end conditions differed. In the US version, person taking the test starts off with some practice words, before the test scoring commences. In the UK Biobank version, instead of practice questions, the first word presented (the same for all participants) is a very easy word, resulting in the first few words being relatively easy. The maximum number of rounds, where the test ends even if the estimate is not within the target margin for the standard error, is different between the US and UK versions.

References

1. Gershon, Richard C., et al. "Language measures of the NIH toolbox cognition battery." Journal of the International Neuropsychological Society 2014 Jul 20(6): 642-651.
2. Giunchiglia, Valentina. Et al. "Neural correlates of cognitive ability and visuo-motor speed: validation of IdoCT on UKBiobank Data" [forthcoming]