**Behavioral and neural investigation of the testing effect in declarative memory**

**1 Main aim**

The testing effect refers to the finding that testing, relative to mere studying, dramatically improves declarative memory. Although the testing effect has been empirically demonstrated in many studies and holds major educational implications, its origin has remained unclear. In line with a recent theory developed by my promotor (Ergo, De Loof, & Verguts, 2020), I propose that the testing effect might originate from reward prediction error (RPE). I aim to examine whether the nature of the testing effect is a kind of RPE effect by a modified variable-choice paradigm, in which the different factors can be systematically manipulated. I will investigate their effects at both behavioral and neural (functional magnetic resonance imaging; fMRI) levels. The outcome of this study should clarify the nature of the ubiquitous testing effect, and thus should have clear implications for educational practice.

**Keywords:** Testing effect; Reward Prediction Error effect; declarative memory; functional magnetic resonance imaging (fMRI)

**2 Design**

**2.1 Main aim**

The main aim of Experiment 1 is to make tentative exploration for our project.

**2.2 Participants**

Based on the previous study using the variable-choice task (Calderon et al., 2021; De Loof et al., 2018), we recruited 76 participants (mean age = 30.35, SD = 6.07, range = 18 - 41; 37 female) via an online experimental platform named Prolific (https://prolific.co/). All the participants’ first language was English.

**2.3 Materials**

A total of 450 words (90 English and 360 Swahili) were used. Participants memorized 90 English-Swahili word pairs. The experiment was programmed with JavaScript.

**2.4 Procedure**

The procedure is presented in Figure 1.

**2.4.1 Acquisition phase**

In the acquisition phase, participants studied 90 English-Swahili word pairs. All word pairs were sequentially presented on the screen for 3 seconds. Presentation time was deliberately kept short to avoid participants learning all word pairs after a single presentation. As such, we can ensure that the bulk of learning takes place in the subsequent testing phase.

**2.4.2 First Test: pretest**

The first test was a pretest with no feedback. The main aim of this test was to remove the learning from Acquisition phase. To be specific, the correct recalled pairs in Test 1 would be excluded from the formal analysis.

Each English word was presented together with the four foreign translations. Participants need to choose the correct Swahili translation (press “f”, “v”, “j”, or “n”) from the four options (no time limit). After making a choice, participants need to indicate on a scale ranging from 1 (very uncertain) to 5 (very certain) how certain they were of their choice (no time limit).

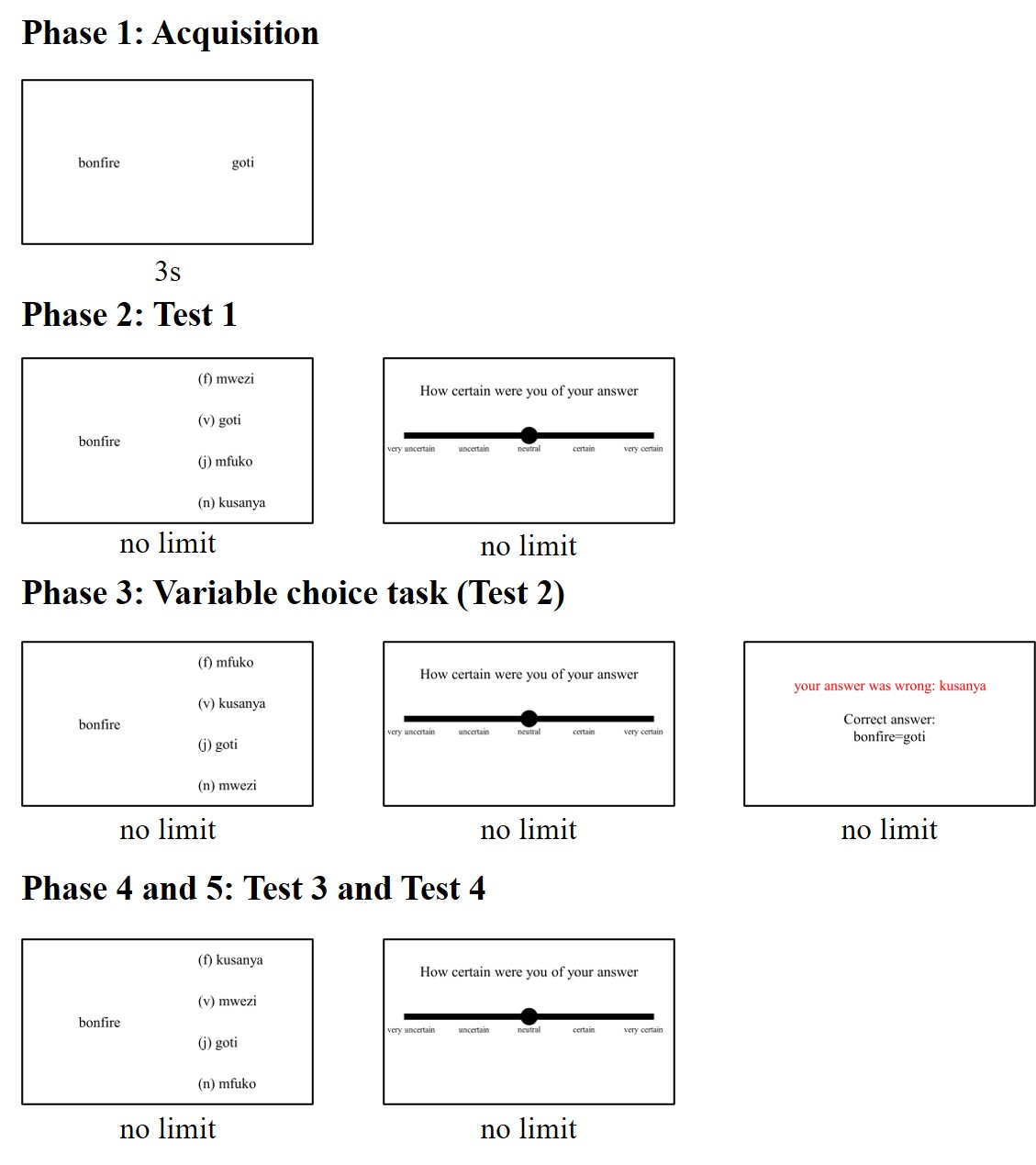
**2.4.3 Second Test: variable choice task**

The second testing was the variable choice task, which defined the studying and testing conditions and evoke the RPE. Its procedure was similar to the first test (but pairs were presented in a different random order). One or four options with the correct answer were framed, which differentiated the studying (20%) from the testing (80%) conditions. After indicating the certainty, feedback was presented (“Wrong” in red or “Correct” in green, no time limit), together with the participant's choice and the correct choice.

**2.4.4 Third and Fourth tests: recalls**

Before the recalls, there was a filler test. On each trial, five arrows were presented, and participants need to indicate the direction of the middle arrow (press the left or right arrow key on the keyboard).

The two recalls separately aimed to measure the final memory performance and to filter the flukes from real knowledge. To be specific, only when a word pair is recalled successfully in both recall tests, it would be seen as correct. Their procedures were as same as the first test, except for the order of the word pairs and options.



**Figure 1. The procedure of Experiment 1**

**3 The main logic of Data analysis**

We predicted that if RPE causes a testing effect, the inclusion of both RPE (i.e., reward, prediction, and prediction error) and testing regressors in the same statistical model would render the testing regressor statistically non-significant and the RPE regressors statistically significant.