

Replication of

Retrieval Practice Produces More Learning than Elaborative Studying with Concept Mapping

By Karpicke, J.D. / Blunt, J.R. (2011).

In: *Science*, 331, pp. 772-775.

Replication Author(s)

Nick Buttrick, Anup Gampa, Lilian Hummer, & Brian A. Nosek

In a lab study with Purdue undergraduates, Karpicke and Blunt (2011) gave participants 5 minutes to read a short article about sea otters. Participants were then randomly assigned to one of four conditions: to either do no additional studying (the Study-once condition); to study the passage for 15 additional minutes in 5 minute blocks (the Repeated-Study condition); to spend 25 minutes concept-mapping the passage (the Concept-Mapping condition); or to spend 10 minutes writing what they could remember from the passage, then taking 5 more minutes to study the passage, then another 10 to write what they could recall (the Retrieval Practice condition). Compared to the other three conditions, especially the Concept-Mapping condition, those in the Retrieval Practice condition remembered more about the passage when retested one week later. This paper consisted of two studies, both in-lab studies. Following the protocol for this replication project, Study 1 is the study being replicated; Study 2 extended Study 1 by using different texts, a different assessment of learning, and a within-subjects design (Study 1 was between-subjects).

Hypothesis to replicate and bet on:

In a memory test one week after learning, Retrieval Practice leads to participants recalling more correct information than Concept-Mapping. A t-test, $p < .05$ using a two-tailed test, comparing the Retrieval Practice and Concept Mapping conditions.

Original test statistics: $N=40$ (20 in each condition); Mean performance = 0.67 in the Retrieval Practice condition and 0.45 in the Concept Mapping condition. The comparison between Retrieval Practice and Concept Mapping was reported as $F(1, 38)=21.63$; $p=.000039$

Power Analysis and Criteria for Replication: First Data Collection

The original sample size was 40 observations for the two conditions to be replicated (20 per condition). The standardized effect size between the two conditions of interest, measured as an r was 0.602. To have 90% power to detect 75% of

the original effect size a sample size of 47 is required. We will collect 48 to have 24 in each condition. The criteria for replication is an effect in the same direction as the original study and a $p\text{-value} < 0.05$ (two-sided test).

Power Analysis and Criteria for Replication: Second Data Collection

If the original result is not replicated in the first data collection, a second data collection of 68 additional individuals will be carried out so that the total sample size is 116. If a second data collection is carried out, it will be tested if the original result replicates in the pooled sample of the first and second data collection.

To have 90% power to detect 50% of the original effect size a sample size of 114 is required; i.e. a sample size of 68 in the second data collection to have a sample size of 114 in total for the first and second data collection pooled. To have equal numbers of participants per condition, we will collect 48 (24 per condition) in the first data collection and 68 (34 per condition) in the second data collection for a total of 116. The criteria for replication is an effect in the same direction as the original and a $p\text{-value} < 0.05$ (in a two-sided test) in the pooled data.

Sample

The sample size in the first data collection will consist of 48 individuals from the Charlottesville, Virginia community. Participants will either be recruited using the UVA research participant pool, or they will be recruited through in-person/flyer recruitment on grounds at the University of Virginia and from popular community locations in Charlottesville, such as the downtown pedestrian mall. Participants recruited through the participant pool will be compensated with research credit for both experimental sessions. Participants recruited in-person will be compensated \$20 for participation in each experimental session (\$40 total for participation in both sessions), directly by the experimenter.

If the original result is not replicated in the first data collection (two-sided $p\text{-value} < 0.05$ in the original direction), a second data collection of 68 additional individuals from the same population will be carried out so that the total sample size is 116.

Materials

We will use the materials from the original study, provided by the original authors; including the target passage about sea otters, instructions to participants, the example concept map, and the assessment-of-learning and judgement-of-learning tools, as described on pages 1-3 of the Supplementary Information. The experiment will be in English as in the original study.

Procedure

We will follow the procedure described in the original article and have incorporated recommendations provided by the original authors. The following summary of the experimental procedure is based on pages 772-773 of the main article and pages 1-3 of the Supplementary Information, in combination with direct feedback provided by the original authors.

Subjects will be run through the learning session of the study with 1 to 4 participants at a time run in individual carrels. After giving informed consent, all students will be given a short passage on sea otters and told to study the passage for 5 minutes. Those in the Concept Mapping condition will be given instructions about concept mapping, shown an example concept map, and given 25 additional minutes to map the passage, during which they can refer to the passage freely. They will then be asked if they had any prior knowledge of, or experience with a concept map. Those in the Retrieval

Practice condition will be given 10 minutes to write down as much as they could remember about the passage, then given 5 additional minutes to study the passage, then given another 10 minutes to recall what they could remember. Finally, participants in both conditions will be asked to predict how much they'd remember about the passage in one week on an 11-point scale (0%, 10%...100%).

One week later, all participants will return to the lab and are given an untimed follow-up learning test featuring 14 direct-recall based short-answer questions (e.g. "What does sea otter fur consist of?") and 2 inference questions (e.g. "What would be the consequences of removing sea otters from their environment?").

Analysis

There were no reported exclusion criteria, so we will include all participants that complete the key dependent measure in both sessions. In the primary test, scores on the follow-up learning test will be subjected to a two-tailed t-test comparing scores in the Concept-Mapping and Retrieval Practice conditions.

As secondary analyses, we will compare the learning-phase concept maps and free recall from participants in the Concept Mapping and Retrieval Practice conditions.

The test results for all participants will be scored as follows: 1 point for every idea unit correctly recalled. Two raters will independently do the scoring for every participant. Differences in concept map scores and scores from the second retrieval practice session will be assessed with a t-test. Judgments of Learning from the learning phase will also be assessed in the same manner as the follow-up learning test, using a two-tailed t-test comparing Concept Mapping with Retrieval Practice.

Differences from Original Study

The replication procedure is largely the same as that of the original study, with some unavoidable deviations. Where the original study used 4 conditions: Study Once, Repeated Study, Concept Mapping, and Retrieval Practice, the current replication only includes Concept Mapping and Retrieval Practice. As such, instead of using a one-way ANOVA with follow-up pairwise comparisons to test the difference of interest, we will use a t-test instead.

The E-Prime computer program used in the original study was unable to be retrieved by the original authors. Specifically, in the original study, the sea otters text was presented on the computer screen, as was the answer box for the retrieval practice condition. However, all other tasks were completed using pencil and paper in the original study. For this reason, our replication will use hard copies of all of the original materials used in the original study (including the sea otters text and the retrieval practice answer box originally run through the E-Prime program). Original authors agreed that the change from computer to paper responses should not be a consequential difference for this effect.

The replication will be performed in Charlottesville in the fall of 2016 or spring of 2017, whereas the data in the original study was carried out at Purdue, Fall 2008. This replication will either pay participants or provide them course credit, while the original study compensated participants with course credit only. As such, as in all replications, the sample, recruiting, and setting are different from the original study. There are no claims in the original article that suggest that these deviations are material for the tested effects. Nevertheless, we sought review from the original authors

before conducting the replication to confirm.

The original paper contains two studies: for the replication the focus is only on study 1 following the project protocol to select the first study in the paper reporting treatment effects.

Replication Results for the First Data Collection (90% power to detect 75% of the original effect size)

Participants 52 students from the University of Virginia participated in return for course credit. All were recruited from the Department of Psychology's Participant Pool, and none were paid for their participation. Fifty-six percent of participants were female, with an average age of 19.23 (SD=1.26). 3 participants were dropped from the sample, due to lack of attendance for the follow-up learning test held one week later, leaving a final sample of 49 participants, one more than originally planned.

The hypothesis that retrieval practice would lead to better recall one week later than concept-mapping was supported by the data. Those assigned to retrieval practice (n=23) recalled more in the one-week follow-up test than those assigned to concept-mapping (n=26): Retrieval M=62.3% recalled, SD=19%; Concept-Mapping M=46.9% recalled, SD=19%; $t(48)=2.88$, $p = 0.006$, $r = 0.39$ [95% CI: 0.11, 0.71].

As in the original study, participants did not differ in their initial learning. A comparison of the number of concepts in the concept maps for those assigned to concept mapping versus the number of concepts recalled immediately after the second recall phase were not reliably different in the numbers of learned concepts: Recall M=71.7% recalled, SD=18%; Concept-Mapping M=63.2% recalled, SD=16%;

$t(48)=1.69$, $p = 0.097$.

Also, as in the original study, students predicted that they would learn more in the concept-mapping than they would in the retrieval condition (even though the opposite turned out to be true). Consistent with the original result, in our replication, those in the concept-mapping condition predicted that they would remember 69% of the material on average (SD=14%), while those in the retrieval condition only predicted that they would remember 57.8% of the material (SD=19%), $t(48)=2.33$, $p = 0.025$. Participants' predicted memory of the material did not predict their resulting memory, $r(48)=-0.08$ [-0.36, 0.20], $p=0.570$, and there was no difference in the strength of the correlation between the two conditions, $z=-0.629$ [-0.729 0.391], $p=0.530$.

Unplanned Protocol Deviations

There were no unplanned protocol deviations during the study.

Discussion

Our replication observed results consistent with the original study. Although the effect size of our replication was smaller than the original ($r=.60$ for original and $r=.39$ for replication), the original effect size fell within the confidence interval of the replication. This successful replication adds confidence and precision to the original observation, and suggests generalizability to a new sample and setting.

References

Karpicke, J.D. / Blunt, J.R. (2011): *Retrieval practice produces more learning than elaborative studying with concept mapping*, Science, 331, pp. 772-775.

