Journal of Memory and Language 83 (2015) 53–61

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Testing enhances learning across a range of episodic memory abilities

Steven C. Pan, Harold Pashler, Zachary E. Potter, Timothy C. Rickard *

Department of Psychology, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0109, United States

ARTICLE INFO

Article history:
Received 15 December 2014
revision received 7 April 2015
Available online 29 April 2015

Keywords:
Individual differences
Testing effect
Episodic memory
Retrieval practice

ABSTRACT

Brewer and Unsworth (2012) reported that individuals with low episodic memory ability exhibit a larger testing effect, a finding with potentially important educational implications. We conducted two replication attempts of that study. Exp 1 ($n = 120$) drew from a broad demographic sample and was conducted online, while Exp 2 ($n = 122$) was conducted in the lab with undergraduate students. Both experiments demonstrated a large testing effect across the range of episodic ability in our sample, and with no trend suggesting a larger testing effect for lower ability subjects. We show that apparent differences in the distribution of episodic ability levels between our samples and that of Brewer and Unsworth provide a plausible account of the contrasting correlation results, and that, more generally, sampling from a restricted ability range can yield positive, negative, or no correlation even if there is no difference in the effectiveness of testing for low vs. high ability subjects in the broader population. We discuss methodological and theoretical issues that complicate interpretation of individual differences effects in this domain, individual difference predictions of testing effect models, and educational implications.

Introduction

A large body of empirical research has established that retrieval from memory during a test enhances subsequent memory for that information more than does an equivalent period of time spent restudying the same materials. This phenomenon has frequently been referred to as the *testing effect* or *retrieval practice effect*. In recent years, the testing effect has been repeatedly demonstrated using a wide variety of materials ranging from word pairs to lecture content (for reviews see Carpenter, Pashler, Wixted, & Vul, 2008; McDaniel, Roediger, & McDermott, 2007; Roediger & Karpicke, 2006). While there has been a great deal of research into the cognitive mechanisms underlying the testing effect in recent years, the role of individual differences in cognitive abilities has only recently begun to receive attention (Bouwmeester & Verkoeijen, 2011; Brewer & Unsworth, 2012).

Much of the widespread interest in the testing effect reflects its potential for enhancing learning in applied contexts. Naturally, a conclusive finding that such enhancements are confined to a subset of individuals would be of great import. Brewer and Unsworth (2012) reported evidence suggesting just that. They had subjects complete a battery of assessments designed to measure working memory, attention control, episodic memory, and general-fluid intelligence (Unsworth & Spillers, 2010), along with a paired-associate task that served as a measure of the testing effect (study/test was compared to restudy, in a design roughly modeled after Carpenter, Pashler, & Vul, 2006). Brewer and Unsworth observed no correlation between working memory or attention control abilities and the magnitude of the testing effect. However, both

* Corresponding author at: Department of Psychology, #0109, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0109, United States. Fax: +1 858 534 7190.

E-mail address: trickard@ucsd.edu (T.C. Rickard).

http://dx.doi.org/10.1016/j.jml.2015.04.001
0749-596X/

the episodic memory and general-fluid intelligence constructs were negatively correlated with the testing effect; that is, low episodic memory and general-fluid intelligence scores were associated with a *larger* testing effect. Based on their results, Brewer and Unsworth concluded that test-enhanced learning is most effectively targeted at lower-ability students.

Brewer and Unsworth (2012) were circumspect in proffering explanations for the correlation between general-fluid intelligence and the testing effect. With regard to episodic memory, though, they advanced two potential accounts of the negative correlation with the testing effect. The first was that higher-ability subjects may be better able to use elaborative encoding in both the study/test and restudy conditions (relating to the elaborative retrieval hypothesis of Carpenter, 2009), thus reducing the size of the testing effect. The second was that lower-ability subjects may be forced to use more efficient retrieval strategies during initial testing.

The work described here focused on determining whether Brewer and Unsworth's (2012) episodic memory results can be independently replicated and confirmed. The same methodologies and materials (provided by the original authors) were used. We completed two replication attempts, the first online, sampling from a general population of online experimental subjects, and the second in the laboratory, sampling from university students.

Experiment 1

In Experiment 1, we administered the four episodic memory measures (cued recall, picture source, gender source, and delayed free recall) used by Brewer and Unsworth (2012), along with the same paired-associate testing task (detailed in Carpenter et al., 2006), in the same overall order of presentation, and with the same delay interval between sessions (24 h). Aside from the online data collection (which we did not expect to cause differences in outcome; see Buhrmester, Kwang, & Gosling, 2011; Crump, McDonnell, & Gureckis, 2013), the primary difference between this experiment and that of Brewer and Unsworth's design is that we dropped their ability measures for working memory, attention control, and general-fluid intelligence.

Method

Subjects

Sample size was selected based on a priori power analyses using G*Power 3.13 (Heinrich Heine University Düsseldorf, Germany). Given $\alpha = 0.05$ and a desired power of 0.95 to detect a correlation of -0.29 (as observed by Brewer & Unsworth, 2012) or larger (one-tailed test), the required sample size is 120. One-hundred twenty subjects were thus recruited from the Amazon Mechanical Turk worker pool using online advertisement (at https://www.mturk.com). Each subject was compensated $1.50 for their participation. Access to the study was limited to subjects from the United States that had an approval rate of 80% or greater on prior Mechanical Turk Human Intelligence Tasks (HITs). Payment was contingent on completion of both sessions of the experiment and the submission of a valid completion code. The minimum age requirement for participation was 18 years, and there was no upper age limit. Descriptive statistics for subject ages were as follows: $M = 36.74$, $SD = 12.64$; range = 18–65 yrs of age. Over half of the sample (58%) was female.

Materials

As in the Brewer and Unsworth (2012) study, the paired-associate testing task involved 40 word pairs. These pairs were originally published in Carpenter et al. (2006). The four episodic memory measures also used the same word lists, picture stimuli, and audio clips as in original study and were provided by the original authors.

Design and procedure

The experimental design followed that of Brewer and Unsworth (2012), with modifications as follows. Due to our specific interest in episodic memory ability, and the lack of any significant correlations of working memory and attention control abilities with the testing effect in the original study, we only included the episodic memory measures from the prior work. Across the two sessions, subjects completed the episodic memory measures and the paired-associate testing task in the same order as in the original study (session 1 beginning with the cued recall episodic memory measure followed by the study and training phases of the paired-associate testing task; session 2 featured the image source, gender source, and delayed free recall episodic memory measures, followed by the final test of the paired-associate testing task). Sessions 1 and 2 lasted approximately 15 and 25 min, respectively (in contrast, each session of the original study was two hrs long, which was necessary to accommodate a total of 13 cognitive ability assessments as well as the paired-associate testing task).

To enable online participation, the experiment was programmed using Adobe Flash Professional CS6 (Adobe Systems, San Jose, CA) and subjects were able to access the study using any Adobe Flash plugin-equipped web browser and with any computer featuring functional audio output capabilities. Subjects were required to create a username that was used to log-in to both sessions. At the end of session 1, subjects were reminded to return at the same time the following day to complete the experiment. They were also given the opportunity to enter an e-mail address in order to receive an automated reminder of their session 2 appointment. Session 2 became available for log-in at exactly 24 h after the server-recorded start time of session 1. Subjects had a completion window of two hrs to complete session 2 and finish the experiment.

Tasks

The paired-associate testing task and the four episodic memory measures are described below (following Brewer & Unsworth, 2012).

Paired-associate testing task. Subjects first studied 40 word pairs for 6 s each, followed by a training phase in which half of the word pairs were again restudied for 6 s each,