Journal of Applied Research in Memory and Cognition 3 (2014) 171–176

Contents lists available at ScienceDirect

Journal of Applied Research in Memory and Cognition

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

Does testing with feedback help grade-school children learn key concepts in science?

Amanda Lipko-Speed[a,*], John Dunlosky[b,**], Katherine A. Rawson[b]

[a] The College at Brockport, State University of New York, United States
[b] Kent State University, United States

ARTICLE INFO

Article history:
Received 27 November 2013
Received in revised form 2 April 2014
Accepted 6 April 2014
Available online 18 April 2014

Keywords:
Testing effect
Retrieval practice
Science concepts
Grade-school children

ABSTRACT

Testing effects have been well-established across a variety of studies involving school-age children. Specifically, children's test performance improves when they are given the opportunity to practice retrieval prior to the final test as compared to when practice involves only study. The current investigation focused on the influence of testing with feedback on fifth graders' learning of science concepts. Across two experiments, 65 students studied twenty key concept definitions. After study, the definitions were randomly assigned to one of four conditions: control, test only, study only, and test plus feedback. Concepts assigned to the latter three conditions were presented again (either for test only, study only, or test plus feedback) during two separate sessions, and a final test was administered during a separate session. At the final test, students were asked to provide the definition for each of the twenty key concepts. Final test performance was best for definitions that had been tested with feedback, followed by recall for those that had only been tested or only restudied, with lowest performance for those that did not receive additional practice. Despite multiple sessions spaced across several days, however, performance for all conditions was low, which may have resulted from the ineffective use of feedback. In summary, testing followed by feedback can boost younger students learning of science concepts, but mastery will require the use of other strategies as well.

1. Introduction

Tests over to-be-learned materials do more than assess what students have learned, because testing itself requires retrieval practice that has been shown to improve learning, comprehension, and retention (for reviews, see Carpenter, 2012; Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013; Karpicke, 2012; Roediger & Butler, 2011). Although the promise of practice tests for improving student learning has been demonstrated across numerous domains over the past 100 years, the degree to which testing can help younger students learn more complex science concepts has not yet been investigated. In the present research, we first consider the available evidence about the benefits of testing for young children, and we then report two experiments investigating the degree to which testing (with feedback) can help grade-school children learn science concepts.

Whereas the vast majority of testing effect research has involved adult learners, the number of studies involving upper-elementary school students is relatively small but growing (see Table 1 for published studies involving fifth and sixth graders). In this research, students usually began by studying target materials. After initial study, students typically either took tests followed by feedback (Test + FB in the far-right column of Table 1) or restudied the material (Study Only). For instance, Rohrer, Taylor, and Sholar (2010) had 4th and 5th graders learn the names of regions on fictional maps. Each student learned two maps, one via test-plus-feedback practice and the other via study only. In the test-plus-feedback condition, students were tested on their memory of region names and were then immediately shown the correct answer to restudy. In the study-only condition, students were presented with the region names for restudy. One day later, students took a final test in which they had to label the regions on a blank map. Final test performance was better for region names that had been tested with feedback than for region names that had only been restudied. Across the published studies listed in Table 1 (with one exception, Metcalfe

* Corresponding author at: The College at Brockport, Department of Psychology, 350 New Campus Drive, Brockport, NY 14420, United States. Tel.: +1 585 395 5082.
** Corresponding author at: Kent State University, Department of Psychology, Kent, OH 44241, United States.
E-mail address: alipko@brockport.edu (A. Lipko-Speed).

http://dx.doi.org/10.1016/j.jarmac.2014.04.002

Table 1
Summary of current testing effect literature with fifth and sixth grade students.

Authors	Age	Materials	Conditions (after initial study)
Spitzer (1939)	6th graders	577-word text	Test only > no study
Metcalfe and Kornell (2007)	6th graders	Definition-Word Pairs	Test + FB = study only > test only
Metcalfe, Kornell, and Son (2007)	6th–7th graders	Definition-Word Pairs, Spanish Vocabulary	Test + FB > study > no study
Metcalfe, Kornell, and Finn (2009)	6th graders	Definition-Word Pairs	Test + FB > study only
Roediger, Agarwal, McDaniel, and McDermott (2011)	6th graders	Social study facts	Test + FB > study only
Rohrer, D., Taylor, K., & Sholar, B. (2010)	4th–5th graders	Map Location-Names	Test + FB > study only
Bourwmeester and Verkoeijen (2011)	Ages 8–12	DRM Lists	Test only > study only[a]

Note. FB, feedback
[a] Not all subgroups of participants showed this effect.

& Kornell, 2007), criterion test performance was greater when the students practiced retrieval followed by feedback than when they only restudied. Moreover, even younger children demonstrate testing effects (e.g., Fritz, Morris, Nolan, & Singleton, 2007; Lipowski, Pyc, Dunlosky, & Rawson, in press; Marsh, Fazio, & Goswick, 2012) as well as older middle-school students (e.g., Carpenter, Pashler, & Cepeda, 2009; McDaniel, Thomas, Agarwal, McDermott, & Roediger, 2013; Nungester & Duchastel, 1982). Thus, the fifth-grade students who participated in the present experiments were expected to demonstrate a testing effect.

In the present experiments, we contribute to this small but growing literature in two key ways: (a) the to-be-recalled material—full definitions of key science concepts—are more complex than used in prior research with this age group, and (b) we included multiple conditions to assess the degree to which the effects of testing are direct or indirect (Roediger & Karpicke, 2006). These conditions were defined by whether the initial study phase was followed by practice tests and/or study: (a) testing only, (b) testing with feedback,[1] (c) study only, or (d) neither (control). The direct effects of testing occur when testing itself boosts performance, which largely arises from correctly recalling targets during practice (but see Kornell, Hays, & Bjork, 2009; Vaughn & Rawson, 2012). These direct effects are evident when criterion performance is greater after testing only than in the control condition.

Indirect effects of testing can occur when learners use the outcome from testing to enhance the effectiveness of subsequent study when feedback (i.e., the correct definition) is presented. In the present context in which students can study feedback as long as they want, these indirect effects could arise from effective self-regulated learning. According to general theories of self-regulated study (for a review, see Dunlosky & Ariel, 2011), accurate monitoring of one's learning can allow students to effectively allocate their study time. Namely, during test trials, students may evaluate how well they can recall each definition, which may help them to use the self-paced feedback trials to focus on materials that are not yet well learned. In this manner, testing could potentiate learning during restudy. These indirect effects of testing would result in superadditive effects exceeding the combined direct effects of testing alone and studying alone during practice. Considering that prior work with these age groups has typically compared test plus feedback to study alone (Table 1), whether the benefits of testing were due to direct or indirect effects remains an open question, because both may have contributed to the observed effects (although evidence from Bourwmeester & Verkoeijen, 2011, indicates that testing can directly influence performance for learning word lists). Thus, we included all four conditions in the present experiments.

Concerning the materials used in the current research, we had students learn key concept definitions. The materials were provided by fifth-grade teachers and represent science concepts that the students would be expected to learn in units on light and sound (Experiment 1) and geography (Experiment 2). For instance, one concept for light and sound is "What is amplitude? Height of a wave from its resting position to its highest or lowest point" (for other examples, see Appendix). Practice tests in the present experiments involved presenting a concept term (e.g., "amplitude") as the cue and having students attempt to retrieve the correct definition from memory. By contrast, other research using definitions to explore testing effects in children (see Table 1) has prompted them to retrieve key terms when presented with the definitions as a cue. For example, Metcalfe and colleagues (Table 1) presented children with definitions of vocabulary terms (e.g., "words or letters written, printed, or engraved on a surface") and were asked to type the corresponding target word (e.g., "inscription"). Retrieving a single key term may lead to fewer commission errors than attempting to retrieve a target definition, and the latter may undermine the benefits of testing. For instance, when presented with a definition to a newly learned concept, the key term often will either be correctly recalled or nothing will be recalled (i.e., an omission), and if a commission error does occur, then the feedback (i.e., presenting the correct key term) will allow the students to easily identify the error. However, when the task is to recall the correct definition (as in the present experiments), students often make commission errors, and they also often have difficulties evaluating the quality of their recall even when they receive the correct definition for feedback (Lipko et al., 2009). Thus, the feedback trials may not be as useful for helping them to correct errors, and if so, one might expect that the testing effect will arise more from its direct effects on memory than from its indirect effects.

To evaluate such possibilities, we conducted two experiments that used identical procedures but different stimuli, to evaluate the generalizability of the effects across different materials. In Experiment 1, we used definitions about light and sound, and in Experiment 2, we used geography definitions. In both experiments, children began by studying a set of twenty definitions. These definitions were then randomly assigned to one of four conditions: test only, study only, test plus feedback, or control. On the final day, children were tested on all twenty definitions. Given the similarity of the two experiments, we present them together for brevity.

2. Experiments 1 and 2

2.1. Method

2.1.1. Participants

Thirty fifth graders participated in Experiment 1 and 35 fifth graders participated in Experiment 2. Due to class absences, 8 participants did not complete all sessions (3 and 5 in Experiments 1 and 2, respectively), and their data were excluded from analyses. Thus, 27 and 30 participants contributed to analyses in the two experiments. Students were recruited from northeast Ohio middle schools by the Research Center for Educational Technology at Kent

[1] Functionally, this condition involved testing followed by restudy, which we refer to here as "feedback" in keeping with terminology commonly used in the testing effect literature.