Learning and Instruction 39 (2015) 32–44

Contents lists available at ScienceDirect

Learning and Instruction

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

Do students use testing and feedback while learning? A focus on key concept definitions and learning to criterion☆

John Dunlosky*, Katherine A. Rawson

Department of Psychology, Kent State University, Kent, OH 44242, USA

ARTICLE INFO

Article history:
Received 2 April 2014
Received in revised form
30 April 2015
Accepted 17 May 2015
Available online 2 June 2015

Keywords:
Testing effect
Retrieval practice
Self-regulated learning
Feedback
Criterion learning

ABSTRACT

College students must regulate much of their learning and hence it is important to discover whether they use effective study techniques, such as testing with feedback. We conducted three experiments to evaluate the degree to which college students use testing with feedback as they are learning key concept definitions. The three main issues of interest concerned (a) whether students tested themselves during practice, (b) whether they followed tests with feedback, and (c) whether they continued to practice until they could correctly recall each definition one or more times. In each experiment, students in the *self-regulated group* could choose to engage in three activities (test, study, or evaluate test responses) in whatever order or dosage they wanted during practice. For purposes of comparison in each experiment, a *Criterion 3* group completed an experimenter-controlled schedule of test-study practice until each definition was correctly recalled three times. All participants completed a final test one week later. Among the key outcomes, students who self-regulated their learning used tests and often sought feedback soon after. When students incorrectly recalled a definition, they typically chose to study the definition, whereas when they recalled some of a definition correctly, they typically chose to evaluate the quality of their response. Although students on average continued to test themselves until they correctly recalled each definition once, they underperformed on the final test as compared to the Criterion 3 group. Thus, in the present context, students do use testing while learning, but results also suggest that they do not take full advantage of this effective study technique.

When students are attempting to learn almost any kind of material, their long-term retention is typically much better when they test themselves while learning than when they merely study the materials. The benefits of testing are even larger when it is combined with feedback (for recent reviews, see Dunlosky, Rawson, March, Nathan, & Willingham, 2013; Roediger & Butler, 2011). Testing here refers to trying to recall a sought after response from long-term memory, and feedback pertains to reviewing of the correct answer that had been previously tested. Testing with feedback during practice not only improves one's memory for the material, but it also produces transfer to tests that tap application and comprehension (Carpenter, 2012; Karpicke, 2012). Finally, tests are especially effective when students successfully retrieve the correct information on practice tests, with later retention increasing with the number of correct retrievals during practice (e.g., Rawson & Dunlosky, 2011; Vaughn & Rawson, 2011). Given the impressive effects of testing both in the laboratory and in the classroom (e.g., McDaniel, Agarwal, Huelser, McDermott, & Roediger, 2011; Roediger, Agarwal, McDaniel, & McDermott, 2011), it is apparent that students should be using testing with feedback as a study technique as they learn course materials.

These observations lead to three questions that are focal to our current investigation. Do students use testing (versus relying primarily on studying alone) while learning the definitions of key concepts? To the extent that students use self-testing, do they follow it with feedback? And, do they continue to test themselves until they can correctly recall each definition? Answers to these questions have implications for application and for theory. Concerning application, if students are not using testing or do not use it effectively to learn course materials, then interventions aimed at teaching its effective use could benefit student learning and performance. Concerning the latter, theories of self-regulated learning are directed at explaining why students make particular decisions during study, with one aim being to evaluate whether learners'

☆ This project was supported by a James S. McDonnell Foundation 21st Century Science Initiative in Bridging Brain, Mind and Behavior Collaborative Award.

* Corresponding author.
E-mail address: jdunlosk@kent.edu (J. Dunlosky).

http://dx.doi.org/10.1016/j.learninstruc.2015.05.003
0959-4752/

self-regulation is optimal (Metcalfe & Kornell, 2005; Son & Sethi, 2006). For instance, students often choose to space their practice rather than mass it (e.g., Pyc & Dunlosky, 2010; Son, 2004; Toppino & Cohen, 2010). Unfortunately, when given a choice to use a long versus short lag for spacing, they choose the short lag (Cohen, Yan, Halamish, & Bjork, 2013; Wissman, Rawson, & Pyc, 2012), which is less optimal for learning than the long lag. This and other evidence suggests that students often do not regulate their learning effectively, which may also be the case with testing (for a review, see Bjork, Dunlosky, & Kornell, 2013).

In the current study, we investigated students' use of testing and feedback while learning key concept definitions from Introductory Psychology (e.g., "What is the sensorimotor stage? The stage in which children know the world mostly in terms of their sensory experiences and motor activities"). We chose key concept definitions because they comprise a large portion of the material that students are expected to learn in most introductory courses and because they are foundational to more advanced studies. Introductory Psychology concepts were used because the participants were enrolled in this course, so learning the concepts would be relevant to their course work (cf. Rawson, Dunlosky, & Sciartelli, 2013).

During the learning phase of the present experiments, participants in the *self-regulation groups* could choose to test themselves (e.g., "What is the sensorimotor stage?"), study the concept term plus the definition, or make a judgment about the quality of their test response. Importantly, participants in self-regulation groups were free to engage in any or all of these activities (test, study, judge) in any order they wanted and for as long as they wanted. As described further below, this method represents a significant advance beyond prior research exploring students' spontaneous use of testing and will provide answers to the three questions outlined above. To revisit, the first two questions are straightforward and pertain to whether students use testing with feedback. The third question pertains to whether students use testing to reach a given *criterion*, which is the number of times an answer is correctly recalled during the practice phase. Reaching a criterion of one correct recall is essential for having a chance for correctly recalling it after a delay, and reaching even higher criterion levels (e.g., three correct recalls) yields even better retention (e.g., Nelson, Leonesio, Shimamura, Landwehr, & Narens, 1982; Pyc & Rawson, 2009; Rawson & Dunlosky, 2011). Thus, with respect to how *effectively* students use testing and feedback, we were interested in whether students continue to test themselves and use feedback until they reach a criterion of one (or more) correct recalls.

1. Prior literature on self-regulated use of testing and feedback

To put the present research in context, consider the prior research exploring students' use of testing and feedback. Two approaches have been used, including surveys to obtain students' self-reported use of testing (Carrier, 2003; Gurung, 2005; Hartwig & Dunlosky, 2012; Karpicke, Butler, & Roediger, 2009; Kornell & Bjork, 2007; Kornell & Son, 2009; McCabe, 2011; Wissman et al., 2012) and laboratory observation studies that examine how students spontaneously use testing (Bottiroli, Dunlosky, Guerini, Cavallini, & Hertzog, 2010; Karpicke, 2009; Kornell & Bjork, 2008; Kornell & Son, 2009; Murphy, Sanders, Gabriesheski, & Schmitt, 1981; Murphy, Schmitt, Caruso, & Sanders, 1987; Son, 2005). The present study uses the latter approach, but evidence from both is relevant.

Concerning survey data, depending on how college students are asked to report their use of testing, estimates are that from about 40% to 70% of them use testing when preparing for upcoming examinations. These estimates are lower than students' reports of rereading, which is a much less effective technique (for a recent review, see Dunlosky et al., 2013). For instance, Karpicke et al. (2009) had college students describe how they studied: 84% reported rereading as a study technique, whereas 40% reported using flashcards (which presumably involve self-testing) and only 11% explicitly reported using a testing technique. One recent survey on students' use of testing included more specific questions to assess (a) whether and how students used feedback after testing and (b) the degree to which students attempted to practice until information could be correctly recalled (Wissman et al., 2012). Of the several hundred students surveyed, nearly 70% reported using flashcards. However, only 49% of these students indicated that they would use external feedback to check their answers, and 27% indicated that they would check their answer in their minds, which would not provide effective feedback if in fact they were incorrectly recalling a response. Concerning criterion learning, 83% endorsed that they would continue practicing an item on a flashcard until they correctly recalled it, although only 26% claimed that they would continue until they could correctly recall an item more than once. As noted above, correctly recalling an item multiple times on a practice test often yields the biggest benefit to later performance, so students to some degree report using flashcards effectively.

A few laboratory observational studies have investigated students' use of testing during study. One method introduced by Son (2005) involves presenting a list of to-be-learned items for study, and then having students decide (for each item or for the entire list) either to receive a test on the next trial or to study the item on the next trial (see also Karpicke, 2009; Kornell & Bjork, 2008; Kornell & Son, 2009). These studies converge on the conclusion that students do use tests to learn items, but they are largely mute with respect to (a) whether students spontaneously use tests *with* feedback and (b) whether they continue to test until they reach a given criterion level. Specifically, in these studies, many aspects of the learning phase remained under experimenter control, and students were given relatively limited control over the amount, timing, or duration of practice activities. Thus, these methods constrained the degree to which students could control their use of testing and feedback during the learning phase, and these constraints would likely curtail how much students used feedback and/or the extent to which they attempted to learn to a particular criterion. For instance, in one study (Karpicke, 2009; Experiment 2), all students first learned each item to criterion on an experimenter-controlled schedule of test-study practice and only then were asked to choose from one of three options for each item, either test two more times, study again two more times, or drop items from further learning (which were selected 25, 15, and 60% of the time, respectively). This method did not give students the option to choose feedback after testing and is silent on the extent to which students would have spontaneously engaged in learning to criterion (because all students were required to do so).

Similarly, in other studies, students had to choose either test or study and were given a fixed number of trials (e.g., Kornell & Son, 2009; Son, 2005). Hence, they were not given the option to test with feedback and may not have been able to reach a criterion level of learning even if they had planned to do so. In a study by Kornell and Bjork (2008), students were presented with paired associates for practice trials always involving testing followed by study. Students could decide to drop an item from further practice during any trial, so whether they met a criterion could be assessed. Most items were dropped from study after being correctly recalled once. Note, however, that students were given only 10 min to learn the items, and strict time limits can influence students' decision making (Thiede & Dunlosky, 1999).

Another method involves using a recall readiness task in which