



Published in final edited form as:

Memory. 2012 ; 20(8): 899–906. doi:10.1080/09658211.2012.708757.

## Memorial consequences of testing school-aged children

Elizabeth J. Marsh<sup>1</sup>, Lisa K. Fazio<sup>2</sup>, and Anna E. Goswick<sup>1</sup>

<sup>1</sup>Department of Psychology & Neuroscience, Duke University, Durham, NC, USA

<sup>2</sup>Department of Psychology, Carnegie Mellon University, Pittsburgh, PA, USA

### Abstract

A large literature shows that retrieval practice is a powerful tool for enhancing learning and memory in undergraduates (Roediger & Karpicke, 2006a). Much less work has examined the memorial consequences of testing school-aged children. Our focus is on multiple-choice tests, which are potentially problematic since they minimise retrieval practice and also expose students to errors (the multiple-choice lures). To examine this issue, second graders took a multiple-choice general knowledge test (e.g., *What country did the Pilgrims come from: England, Germany, Ireland, or Spain?*) and later answered a series of short answer questions, some of which corresponded to questions on the earlier multiple-choice test. Without feedback, the benefits of prior testing outweighed the costs for easy questions. However, for hard questions, the large increase in multiple-choice lure answers on the final test meant that the cost of prior testing outweighed the benefits when no feedback was provided. This negative testing effect was eliminated when children received immediate feedback (consisting of the correct answer) after each multiple-choice selection. Implications for educational practice are discussed.

### Keywords

Suggestibility; Education; Testing

---

Recent research has emphasised how testing helps students to learn and remember information, both in the laboratory and in the classroom (for reviews see Bangert-Drowns, Kulik, & Kulik, 1991; Roediger & Karpicke, 2006a). Studies with adults have shown that practice retrieving information (e.g., by recalling it) can benefit memory more than other popular learning techniques, including re-reading (Roediger & Karpicke, 2006b), creating a concept map (Karpicke & Blunt, 2011), and note-taking (McDaniel, Howard, & Einstein, 2009). Testing is a powerful learning strategy, in part because it requires effort on the part of the learner and also because retrieval practice matches the task the learner will ultimately face (namely, retrieval). These benefits are not limited to undergraduates; for example, early work by Gates (1917) demonstrated that testing helps children as young as third graders to learn biographical facts. Although much of the pertinent work with children has not involved educational materials, the basic principle is clear: Recalling information from memory helps children to remember a wide range of materials, including stories (second graders; Petros & Hoving, 1980), names of toys (preschoolers; Fritz, Morris, Nolan, & Singleton, 2007), games (first and fourth graders; Baker-Ward, Hess, & Flannagan, 1990), and medical events (children aged 3 to 7 years; Goodman, Bottoms, & Schwartz-Kenney, 1991).

Less clear is whether children benefit from multiple-choice testing, even though instructors frequently prefer multiple-choice tests since they are easier to grade and are often perceived as more objective. The first issue is that selecting an answer from a list of multiple-choice options requires less retrieval effort than does generating a response from memory (Kang, McDermott, & Roediger, 2007). However, a second issue is that multiple-choice tests (by design) expose students to plausible incorrect answers, meaning that such tests have the potential to teach students errors. Undergraduates judge multiple-choice lures as truer after seeing them on an earlier multiple-choice test (Toppino & Brochin, 1989; Toppino & Luipersbeck, 1993), and testing increases the likelihood that undergraduates will answer later general knowledge questions with multiple-choice lures (Roediger & Marsh, 2005). This negative testing effect persists over a 1-week delay and occurs for different types of questions (Fazio, Agarwal, Marsh, & Roediger, 2010; Marsh, Agarwal, & Roediger, 2009; Marsh, Roediger, Bjork, & Bjork, 2007). The present work focuses on whether school-aged children show a similar negative testing effect, and how the size of any cost compares to the size of any benefit received from answering multiple-choice questions.

In his classic work with children, Spitzer (1939) argued that educational tests allow rehearsal of errors as well as correct answers, although he did not test this empirically. To our knowledge this question about children's memories has not been examined using educationally relevant materials and procedures, although there have been explorations of the negative consequences of testing episodic memories. For example, it is well documented that suggestive questioning techniques influence child eyewitnesses (Ceci & Bruck, 1993). Perhaps most relevant is Brainerd and colleagues' demonstration that children sometimes mistakenly misattribute lures encountered during a recognition test to an earlier study phase (*the mere testing effect*, Brainerd & Mojardin, 1998; Brainerd & Reyna, 1996). Source confusions play a large role in these kinds of episodic memory errors (e.g., Bright-Paul, Jarrold, & Wright, 2005; Giles, Gopnik, & Heyman, 2002), with the child misattributing information from a recent source (the suggestive interview, the test) to an earlier time and place (the original witnessed event, the study list). To link to this literature on suggestibility in episodic memory, we tested children aged 6 to 8 years old (second graders). However, it is not clear that learning errors from general knowledge tests will work the same way as suggestibility in these episodic memory paradigms. General knowledge tests do not require the learner to think back to a particular time and place; instead, general knowledge tests encourage the learner to rely on the fluency with which answers come to mind (Kelley & Lindsay, 1993).

Because knowledge is often unassociated with its original context of learning (e.g., Conway, Gardiner, Perfect, Anderson, & Cohen, 1997), an additional goal was to explore children's awareness of any learning from multiple-choice testing. Past work indicates that 4- and 5-year-olds often struggle to judge what they just learned, overestimating what they knew prior to learning facts from a storybook (Esbensen, Taylor, & Stoess, 1997; Taylor, Esbensen, & Bennett, 1994). Of particular interest was whether an illusion of prior knowledge would accompany any negative testing effect observed; that is, if testing increases children's likelihood of answering later questions with multiple-choice lures, will children believe they knew these answers before the experiment or will they acknowledge learning them in the experiment? To the extent that children are aware of learning from the multiple-choice test, it would suggest that source monitoring instructions might help children avoid the negative testing effect (see Poole & Lindsay, 2002 for an example of source instructions reducing suggestibility in episodic memory).

A related goal was to explore whether any negative testing effect would be eliminated if children were told the correct answer immediately after answering each question, as occurs with undergraduates (Butler, Karpicke, & Roediger, 2008). Several issues made it unclear

whether children would benefit from such feedback. First, not a lot is known about how second graders use feedback, as most studies have involved children who are slightly older (for demonstrations with fourth to sixth graders see Bardwell, 1981; Kippel, 1975; Peeck & Tillema, 1978; Peeck, Van Den Bosch, & Kreupeling, 1985; Travers, Van Wagenen, Haygood, & McCormick, 1964). Kindergarteners benefit from feedback when learning to pluralise words, but the effects are small (Bryant & Anisfeld, 1969). Furthermore, none of these studies separated feedback's role in error correction (Pashler, Cepeda, & Wixted, 2005) from whether it helped learners to maintain correct answers (Butler & Roediger, 2008). Finally, in adults, the benefits of feedback depend on the confidence with which responses are made (Butler & Roediger, 2008; Butterfield & Metcalfe, 2001), and sometimes children are not as skilled as adults when judging confidence (e.g., Roebbers, 2002). To address this concern, children in the present study indicated the confidence with which they made each answer, allowing us to examine the relationship between confidence and accuracy, and whether confidence in individual answers affected children's ability to benefit from feedback.

To preview, we conducted an experiment with second graders to observe the positive and negative memorial consequences of multiple-choice testing. Children answered multiple-choice questions about a subset of facts that were tested on a later general knowledge test. Some children received immediate feedback on their multiple-choice answers, to allow us to examine its consequences for later performance. In both conditions, the multiple-choice test included easy and hard questions, to ensure that responses would be made with a range of confidence, and to examine whether children were more likely to endorse and repeat multiple-choice lures for unfamiliar topics. After the children took the final general knowledge test, they made source judgements about where they believed they had learned their answers. Overall, the goal was to document the negative testing effect in children, and to understand situations that might reduce or eliminate it.

## METHOD

### Participants

Thirty-six second graders were recruited from the community to participate in a laboratory experiment (mean age = 7.71,  $SD = .55$ , 64% female, ranging from 6.73 to 8.74 years). Of these, 14 children were assigned to a control no-feedback condition and 22 were assigned to a feedback condition; mean age did not differ across conditions,  $t(34) = 1.16$ ,  $SE = .19$ ,  $p > .25$ . Additional children were tested in the feedback condition to allow for more detailed analyses of error corrections. Parents received monetary compensation and children received prizes for participating.

### Design

The design was a 2 (Feedback: None vs Immediate)  $\times$  2 (Multiple-Choice Testing: Not Tested vs Tested)  $\times$  2 (Question Difficulty: Easy vs Hard) mixed design. Feedback was manipulated between-participants whereas multiple-choice testing and question difficulty were manipulated within-participants.

### Materials

Forty-eight questions on science, history, geography, and other educational topics were selected from Brain Quest® flashcards. Brain Quest® questions are categorised by grade; second grade questions (Feder, 2005a) were used for easy items and fourth grade questions (Feder, 2005b) were used for hard items. Parallel multiple-choice and short answer questions were created; the correct answer was paired with three plausible lures for each multiple-choice question, whereas each short answer question had exactly the same question

prompt, but was open-ended. The easy questions (in multiple-choice format) included, “*What country did the Pilgrims come from: England, Germany, Ireland, or Spain?*” and “*What continent has the most ice on it: Antarctica, Asia, Europe, or North America?*” Hard questions (in multiple-choice format) included “*What part of your body is covered with enamel: bones, nails, skin, or teeth?*” and “*What country has the largest population in the world: China, India, Japan, or Russia?*”. The multiple-choice test contained 24 questions (12 easy and 12 hard; half of the total set), which were asked in a single random order. Two versions of the test were created to counterbalance which questions were asked on the multiple-choice test across children. The final short answer test contained all 48 questions, which were asked in a single random order.

A source test was created to probe children’s beliefs about where they had learned various facts. For each item, the experimenter restated the child’s short answer response (e.g., “*You said that spring is another word for autumn?*”) and asked “*Where did you learn that from?*”. A picture board was created to cue children to the possible sources of their responses. The board contained four photographs: a family to represent *home*, a teacher in a classroom to represent *school*, the experimenter to represent *here* (the experiment), and a question mark to represent “*I don’t know*”. The source questions were presented in the same order as the short answer questions; however, source questions were not asked for “*I don’t know*” responses on the short answer test.

## Procedure

Each parent or guardian gave written consent for their child’s participation and for tape-recording of the session. After each child verbally assented to participate, the experimenter explained the procedure.

The child first completed the *initial multiple-choice test*; all testing was done aloud and the experimenter recorded the child’s responses. The experimenter explained that the child should answer each question, even if she had to guess. After the child answered a practice question (“*Where does a farmer work: a farm, a library, a restaurant, or a zoo?*”), she indicated her confidence in her answer using a 3-point scale, by pointing to one of three drawings on a picture board (taken from Figure 1 of Woolley, Boerger, & Markman, 2004). The picture board showed a smiling child for “*really sure*”, a thinking child for “*a little sure*”, and a confused child for “*not so sure*”. This procedure was repeated for a second practice question, “*What language do they speak in the Netherlands: Dutch, English, French, or German?*”. The two practice questions were constructed so that one would be easy to answer (*farm*) and one would be difficult (*Dutch*), so that the experimenter could ensure the child was using the confidence scale correctly. A second set of practice questions was available in case it was unclear whether the child understood the confidence scale, but was never needed.

Across conditions, the multiple-choice test differed in one important way: Children in the feedback condition received immediate feedback on their answers, whereas children in the control group never received feedback. Feedback was always delivered immediately after a child rated her confidence in her response, and took the form of a statement (e.g., “*The Pilgrims came from England*”). Children in the feedback condition received feedback for all answers, whereas the experimenter simply proceeded to the next question for children in the control no-feedback condition. All children were told that some of the questions might be hard, and that they should pick whichever choice they thought was best. Children in the feedback condition were also told “*If you get an answer wrong, just try to learn the correct answer when you hear it.*” Halfway through the multiple-choice test, all children received a sticker to maintain motivation and attention for the remainder of the task.

After completing the multiple-choice test, all children completed a *filler task*, which involving solving paper-and-pencil mazes for approximately 1.5 minutes.

In the third phase of the experiment all children completed the 48-question *short answer test*, which the experimenter read aloud. For this test, children were warned against guessing and told to respond “*I don’t know*” instead. All children received a sticker halfway through this test.

Lastly, the experimenter administered the *source test*, which required the child to attribute each of their answers (except for “*I don’t know*” responses) to one of four categories. The experimenter first restated each short answer response and then asked the source question (e.g., “*You said that a square has four corners. Where did you learn that from?*”). The child either answered verbally or pointed to one of the four board pictures (*home, school, here, I don’t know*). If the child responded with a source distinct from the four categories provided (e.g., “*I learned that at the beach*”), the experimenter recorded the answer verbatim for later coding.

Finally, the experimenter explained the study to the child, thanked her, and provided a prize. The parent received a written debriefing and his/her compensation. All children were tested individually and the experiment lasted approximately 30 minutes.

## RESULTS

All differences were significant at the  $p < .05$  level unless otherwise noted.

### Multiple-choice test performance

On average, children answered 55% of the multiple-choice questions correctly, and they performed better on easy questions ( $M = .66$ ,  $SD = .13$ ) than hard ones ( $M = .45$ ,  $SD = .15$ ),  $F(1, 34) = 51.95$ ,  $MSE = .01$ ,  $\eta_p^2 = .60$ . Feedback condition did not matter, which was expected since feedback was delivered after multiple-choice responses were made,  $F < 1$ . Overall, children showed high resolution when judging confidence, with children in both conditions correctly answering more questions rated as high confidence than low confidence ( $\gamma = .79$ ,  $SD = .45$  for the control and  $\gamma = .61$ ,  $SD = .61$  following feedback; resolution did not differ across conditions,  $t < 1$ ).

### Performance on the final short answer test

Two trained coders classified each short answer response as the correct answer, one of the multiple-choice lures, another wrong answer, or “*don’t know*”. They agreed on 99% of judgements and the third author resolved all discrepancies.

We begin with the correct answers, as shown in Table 1. Children showed a robust positive testing effect on the final short answer test, performing better on questions that had appeared on the prior multiple-choice test ( $M = .59$ ,  $SD = .21$ ) than on questions being tested for the first time ( $M = .25$ ,  $SD = .13$ ),  $F(1, 34) = 153.62$ ,  $MSE = .02$ ,  $\eta_p^2 = .82$ . The benefits of testing were similar for easy and hard questions,  $F < 1$ , and were greater for the participants who received feedback during the multiple-choice test,  $F(1, 34) = 29.41$ ,  $MSE = .02$ ,  $\eta_p^2 = .46$ .

However, as shown in Table 2, prior testing also increased the likelihood that children answered final short answer questions with multiple-choice lures ( $M = .10$ ,  $SD = .12$ ), as compared to the baseline for items that had not been tested previously ( $M = .03$ ,  $SD = .04$ ),  $F(1, 34) = 24.67$ ,  $MSE = .01$ ,  $\eta_p^2 = .42$ . Critically, this negative testing effect was limited to

the control no-feedback condition, where participants answered 18% of previously tested items with multiple-choice lures, as compared to only 3% of new questions,  $t(13) = 4.47$ ,  $SED = .04$ ,  $d = 1.59$ . The negative testing effect was eliminated in the feedback condition, 5% vs 4%,  $t < 1$ , resulting in a significant interaction between feedback condition and testing,  $F(1, 34) = 20.35$ ,  $MSE = .01$ ,  $\eta_p^2 = .37$ . Feedback eliminated the negative testing effect for both easy and hard items ( $t < 1$ ), whereas control no-feedback participants showed an even larger negative testing effect for hard items (increasing from .03 to .26,  $t(13) = 4.18$ ,  $SED = .05$ ,  $d = 1.66$ ), than for easy items (increasing from .02 to .11,  $t(13) = 2.79$ ,  $SED = .03$ ,  $d = 1.01$ ), leading to a significant three-way interaction between feedback condition, prior testing and question ease,  $F(1, 34) = 7.19$ ,  $MSE = .01$ ,  $\eta_p^2 = .18$ .

### Connecting multiple-choice and short answer performance

To better understand the effects of feedback, we examined how performance on the short answer test changed as a function of whether multiple-choice selections were correct or incorrect, as well as the confidence with which these selections were made.

Receiving feedback clearly aided children's error correction. Error correction was defined as producing the correct answer on the final test, after selecting an incorrect multiple-choice lure. In the control no-feedback condition, children only corrected 2% ( $SD = 5\%$ ) of errors made on the initial multiple-choice test. In contrast, 57% ( $SD = 21\%$ ) of multiple-choice errors were corrected following feedback, which was significantly better than the error correction observed in the control condition,  $t(34) = 9.43$ ,  $SED = .06$ ,  $d = 3.28$ . Furthermore, the data hint that children might have been more likely to correct high-confidence errors than erroneous guesses (a finding known as the *hypercorrection effect*; Butterfield & Metcalfe, 2001). Following feedback, children appeared more likely to correct their high-confidence errors ( $M = .65$ ,  $SD = .42$ ,  $n = 14$ ) than those made with medium ( $M = .54$ ,  $SD = .28$ ,  $n = 22$ ) or low confidence ( $M = .40$ ,  $SD = .29$ ,  $n = 14$ ). However, no analysis is reported as only 6 of the 22 children in the feedback condition made errors with all three levels of confidence.

A second function of feedback involves the maintenance of correct answers. Because work with adults indicates that this benefit is normally linked to low confidence correct answers (Butler & Roediger, 2008), we computed a 2 (Feedback Condition)  $\times$  3 (Confidence) ANOVA on the proportion of correct multiple-choice selections that were retained on the final test. Critically, the interaction between feedback condition and confidence level was significant,  $F(2, 32) = 14.65$ ,  $MSE = .05$ ,  $\eta_p^2 = .48$ . Children in the two conditions were equally likely to retain medium-confidence (control no-feedback  $M = .66$ ,  $SD = .28$ , feedback  $M = .66$ ,  $SD = .26$ ,  $t < 1$ ) and high-confidence correct answers (control no-feedback  $M = .93$ ,  $SD = .11$ , feedback  $M = .92$ ,  $SD = .17$ ,  $t < 1$ ); however, children who received feedback retained a greater proportion of low-confidence correct guesses ( $M = .81$ ,  $SD = .25$ ) than did those in the no-feedback condition ( $M = .19$ ,  $SD = .19$ ),  $t(17) = 6.04$ ,  $SED = .10$ ,  $d = 2.79$ .

In summary, feedback had two separate effects: It helped children to correct their errors and it helped them to retain correct guesses, both of which are consistent with the patterns normally observed in adults (Butler et al., 2008; Fazio, Huelser, Johnson, & Marsh, 2010).

### Source attributions

Children reported their beliefs about where they had learned their answers to the short answer questions. Although these source judgements cannot be objectively validated, and are open to item selection effects, they yield a sense of children's awareness of learning

during the experimental session. Our focus is on items that were previously tested on the multiple-choice test, since only these items should potentially be attributed to experimental learning. Table 3 shows the proportion of answers that were specifically attributed to learning during the experiment, attributed to learning from pre-experimental sources (the sources “home”, “school”, and other specific pre-experimental sources were collapsed to form this category), or not attributed to a specific source.

The left portion of Table 3 shows children’s source attributions for *correct answers*. Only feedback participants were aware that some of their correct answers had been learned in the experiment. Children who received feedback stated that they learned 36% of their correct answers in the experiment, compared to only 5% for children in the control group,  $t(34) = 5.55$ ,  $SED = .06$ ,  $d = 1.95$ .

Of particular interest was whether children were aware of learning *multiple-choice lure answers* in the experiment. The data in the right portion of Table 3 are limited to the children who produced at least one multiple-choice lure answer on the short answer test (12 in the no-feedback condition and 13 in the feedback condition). In the control no-feedback condition, almost none of the multiple-choice lures produced on the final test were attributed to the experimental setting. Although children in the immediate feedback condition produced few multiple-choice lures on the final test (see Table 2), these children attributed a greater proportion of their multiple-choice lure answers to the experiment ( $M = .32$ ) as compared to the no-feedback condition ( $M = .01$ ),  $t(23) = 2.69$ ,  $SED = .12$ ,  $d = 1.12$ . In contrast, children in the no-feedback condition experienced an illusion of prior knowledge, attributing a greater proportion of their multiple-choice lure answers to specific pre-experimental sources ( $M = .66$ ) as compared to the feedback condition ( $M = .30$ ),  $t(23) = 2.61$ ,  $SED = .14$ ,  $d = 1.10$ .

## DISCUSSION

Without feedback, children showed both positive and negative effects of testing: Answering multiple-choice questions helped them to answer later short answer questions, but it also increased the likelihood that they answered final questions with multiple-choice lures. For easy questions, the benefits of prior testing outweighed the costs. However, for hard questions prior multiple-choice testing yielded a larger increase in multiple-choice lure answers (+23%) than in correct answers (+16%). Although multiple-choice testing was sometimes beneficial, it hurt more than it helped when the questions were difficult and feedback was not provided.

Furthermore, children in the no-feedback condition were relatively unaware that they were reproducing lures from the earlier multiple-choice test, instead attributing these answers to pre-experimental sources such as learning at school. This illusion of prior knowledge means that source monitoring instructions would be unlikely to help children avoid the negative consequences of multiple-choice testing, in contrast to the benefits of source monitoring instructions in reducing child eyewitnesses’ errors (Poole & Lindsay, 2002). This illusion of prior knowledge fits with the small literature on children’s ability to judge the origins of their knowledge, where it has been demonstrated that young children are more likely to overestimate their prior knowledge for facts (the target of the present work) than for behaviours (Esbensen et al., 1997). It is not surprising that children were unaware of learning from tests, since even adults are often unaware of how tests might change what they know (and instead view tests as neutral assessment devices).

The children who received feedback showed very adult-like patterns of behaviour. First, feedback completely eliminated the negative testing effect, consistent with findings with

undergraduates (Butler & Roediger, 2008). Also consistent with the adult literature, feedback helped the children to maintain low-confidence correct answers (Butler et al., 2008; Fazio et al., 2010) as well as to correct errors (Pashler et al., 2005). Finally, children appeared to correct more of their high-confidence errors than their erroneous guesses (the hypercorrection effect; Butterfield & Metcalfe, 2001), although more research is needed on this point.

Overall, multiple-choice tests remain a viable option for educators, even with children as young as second graders. However, the present work highlights one way that errors can enter children's knowledge bases and yield an illusion of prior knowledge. Care should be taken to provide corrective feedback in order to prevent children from reproducing multiple-choice lures on later tests and integrating the errors into their knowledge bases.

## Acknowledgments

This research was supported by grant 1R03-HD-055683 from the National Institutes of Child Health and Human Development to EJM.

## References

- Baker-Ward L, Hess TM, Flanagan DA. The effects of involvement on children's memory for events. *Cognitive Development*. 1990; 5:55–69.
- Bangert-Drowns RL, Kulik JA, Kulik CLC. Effects of frequent classroom testing. *Journal of Educational Research*. 1991; 85:89–99.
- Bardwell R. Feedback: How does it function? *The Journal of Experimental Education*. 1981; 50:4–9.
- Brainerd CJ, Mojardin AH. Children's and adults' spontaneous false memories: Long-term persistence and mere-testing effects. *Child Development*. 1998; 69:1361–1377. [PubMed: 9839421]
- Brainerd CJ, Reyna VF. Mere memory testing creates false memories in children. *Developmental Psychology*. 1996; 32:467–478.
- Bright-Paul A, Jarrod C, Wright DB. Age-appropriate cues facilitate source-monitoring and reduce suggestibility in 3- to 7-year-olds. *Cognitive Development*. 2005; 20:1–18.
- Bryant B, Anisfeld M. Feedback versus no-feedback in testing children's knowledge of English pluralisation rules. *Journal of Experimental Child Psychology*. 1969; 8:250–255.
- Butler AC, Karpicke JD, Roediger HL III. Correcting a metacognitive error: Feedback increases retention of low-confidence correct responses. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2008; 34:918–928.
- Butler AC, Roediger HL III. Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. *Memory & Cognition*. 2008; 36:604–616.
- Butterfield B, Metcalfe J. Errors committed with high confidence are hypercorrected. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2001; 27:1491–1494.
- Ceci SJ, Bruck M. Suggestibility of the child witness: A historical review and synthesis. *Psychological Bulletin*. 1993; 113:403–439. [PubMed: 8316609]
- Conway MA, Gardiner JM, Perfect TJ, Anderson SJ, Cohen GM. Changes in memory awareness during learning: The acquisition of knowledge by psychology undergraduates. *Journal of Experimental Psychology: General*. 1997; 126:393–413. [PubMed: 9407649]
- Esbensen BM, Taylor M, Stoess C. Children's behavioral understanding of knowledge acquisition. *Cognitive Development*. 1997; 12:53–84.
- Fazio LK, Agarwal P, Marsh EJ, Roediger HL III. Memorial consequences of multiple-choice testing on immediate and delayed tests. *Memory & Cognition*. 2010; 38:407–418.
- Fazio LK, Huelser BJ, Johnson A, Marsh EJ. Receiving right/wrong feedback: Consequences for learning. *Memory*. 2010; 18:335–350. [PubMed: 20408043]
- Feder, CW. *Brain quest: 2nd grade*. 3. New York: Workman Publishing Company, Inc; 2005a. rev
- Feder, CW. *Brain quest: 4th grade*. 3. New York: Workman Publishing Company, Inc; 2005b. rev



- Fritz CO, Morris PE, Nolan D, Singleton J. Expanding retrieval practice: An effective aid to preschool children's learning. *The Quarterly Journal of Experimental Psychology*. 2007; 60:991–1004. [PubMed: 17616915]
- Gates AI. Recitation as a factor in memorising. *Archives of Psychology*. 1917; 6
- Giles JW, Gopnik A, Heyman GD. Source monitoring reduces the suggestibility of preschool children. *Psychological Science*. 2002; 13:288–291. [PubMed: 12009053]
- Goodman GS, Bottoms BL, Schwartz-Kenney BM. Children's testimony about a stressful event: Improving children's reports. *Journal of Narrative and Life History*. 1991; 1:69–99.
- Kang SHK, McDermott KB, Roediger HL III. Test format and corrective feedback modify the effect of testing on long-term retention. *European Journal of Cognitive Psychology*. 2007; 19:528–558.
- Karpicke JD, Blunt JR. Retrieval practice produces more learning than elaborate studying with concept mapping. *Science*. 2011; 331:772–775. [PubMed: 21252317]
- Kelley CM, Lindsay DS. Remembering mistaken for knowing: Ease of retrieval as a basis for confidence in answers to general knowledge questions. *Journal of Memory and Language*. 1993; 32:1–24.
- Kippel GM. Information feedback, need achievement, and retention. *The Journal of Educational Research*. 1975; 68:256–261.
- Marsh EJ, Agarwal P, Roediger HL III. Memorial consequences of answering SAT II questions. *Journal of Experimental Psychology: Applied*. 2009; 15:1–11. [PubMed: 19309212]
- Marsh EJ, Roediger HL III, Bjork RA, Bjork EL. Memorial consequences of multiple-choice testing. *Psychonomic Bulletin and Review*. 2007; 14:194–199. [PubMed: 17694900]
- McDaniel MA, Howard DC, Einstein GO. The read-recite-review study strategy: Effective and portable. *Psychological Science*. 2009; 20:516–522. [PubMed: 19320858]
- Pashler H, Cepeda NJ, Wixted JT. When does feedback facilitate learning of words? *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2005; 31:3–8.
- Peek J, Tillema HH. Delay of feedback and retention of correct and incorrect responses. *Journal of Experimental Education*. 1978; 47:171–178.
- Peek J, Van Den Bosch AB, Kreupeling WJ. Effects of informative feedback in relation to retention of initial responses. *Contemporary Educational Psychology*. 1985; 10:303–313.
- Petros T, Hoving K. The effects of review on young children's memory for prose. *Journal of Experimental Child Psychology*. 1980; 30:33–43.
- Poole DA, Lindsay DS. Reducing child witnesses' false reports of misinformation from parents. *Journal of Experimental Child Psychology*. 2002; 81:117–140. [PubMed: 11786006]
- Roebbers CM. Confidence judgements in children's and adults' event recall and suggestibility. *Developmental Psychology*. 2002; 38:1052–1067. [PubMed: 12428714]
- Roediger HL III, Karpicke JD. The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*. 2006a; 1:181–210.
- Roediger HL III, Karpicke JD. Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*. 2006b; 17:249–255. [PubMed: 16507066]
- Roediger HL III, Marsh EJ. The positive and negative consequences of multiple-choice testing. *Journal of Experimental Psychology: Learning, Memory, & Cognition*. 2005; 31:1155–1159.
- Spitzer HF. Studies in retention. *The Journal of Educational Psychology*. 1939; 9:641–656.
- Taylor M, Esbensen BM, Bennett RT. Children's understanding of knowledge acquisition: The tendency for children to report that they have always known what they have just learned. *Child Development*. 1994; 65:1581–1604. [PubMed: 7859544]
- Toppino TC, Brochin HA. Learning from tests: The case of true–false examinations. *The Journal of Educational Research*. 1989; 83:119–124.
- Toppino TC, Luipersbeck SM. Generality of the negative suggestion effect in objective tests. *The Journal of Educational Research*. 1993; 86:357–362.
- Travers RMW, Van Wagenen RK, Haygood DH, McCormick M. Learning as a consequence of the learner's task involvement under different conditions of feedback. *Journal of Educational Psychology*. 1964; 55:167–173.

Woolley JD, Boerger EA, Markman AB. A visit from the Candy Witch: Factors influencing young children's belief in a novel fantastical being. *Developmental Science*. 2004; 7:456–468. [PubMed: 15484594]

**TABLE 1**

Proportion of final short answer questions answered correctly, as a function of prior feedback, prior multiple-choice testing, and question difficulty

	<u>No feedback</u>		<u>Immediate feedback</u>	
	<b>Not tested</b>	<b>Tested</b>	<b>Not tested</b>	<b>Tested</b>
Easy	.38 (.15)	.57 (.16)	.37 (.21)	.78 (.17)
Hard	.11 (.08)	.27 (.14)	.12 (.12)	.63 (.20)
<i>M</i>	.24 (.10)	.42 (.13)	.25 (.15)	.70 (.18)

Standard deviations are shown in parentheses.

**TABLE 2**

Proportion of final short answer questions answered with multiple-choice lures, as a function of prior feedback, prior multiple-choice testing, and question difficulty

	<u>No feedback</u>		<u>Immediate feedback</u>	
	<u>Not tested</u>	<u>Tested</u>	<u>Not tested</u>	<u>Tested</u>
Easy	.02 (.04)	.11 (.12)	.03 (.06)	.04 (.06)
Hard	.03 (.05)	.26 (.19)	.05 (.06)	.05 (.07)
<i>M</i>	.03 (.03)	.18 (.13)	.04 (.04)	.05 (.06)

Standard deviations are shown in parentheses.

**TABLE 3**

Proportion of responses to previously tested short answer questions attributed to the experiment, to pre-experimental sources, or to no source, for correct answers and multiple-choice lure answers

	<b>Correct answers</b>		<b>MC lure answers</b>	
	<b>No feedback</b>	<b>Immediate feedback</b>	<b>No feedback</b>	<b>Immediate feedback</b>
Experiment	.05 (.11)	.36 (.19)	.01 (.03)	.32 (.40)
Pre-experiment	.76 (.26)	.50 (.20)	.66 (.31)	.30 (.37)
Sourceless	.20 (.25)	.14 (.14)	.33 (.31)	.38 (.41)

Standard deviations are shown in parentheses.