Learning and Instruction 22 (2012) 253–261

Contents lists available at SciVerse ScienceDirect

Learning and Instruction

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

Hypercorrection of high confidence errors in children

Janet Metcalfe [a,*], Bridgid Finn [b,1]

[a] Department of Psychology, Columbia University, Amsterdam Ave, New York, United States
[b] Department of Psychology, Washington University in St. Louis, United States

ARTICLE INFO

Article history:
Received 22 February 2011
Received in revised form
12 October 2011
Accepted 18 October 2011

Keywords:
Metacognition
Children
Hypercorrection
Error correction

ABSTRACT

Three experiments investigated whether the hypercorrection effect – the finding that errors committed with high confidence are easier, rather than more difficult, to correct than are errors committed with low confidence – occurs in grade school children as it does in young adults. All three experiments showed that Grade 3–6 children hypercorrected high confidence errors and the children also claimed that they 'knew those answers all along.' Experiment 2 included two second-guess tasks following error commission, one in which the children attempted to choose the correct answer from six options and the other in which they tried to generate a correct second response. Neither provided evidence that children actually knew high confidence corrections all along. Experiment 3, however, showed that the children had some preferential *partial* knowledge insofar as they needed fewer hints to guess the correct answers to high confidence than to low confidence errors.

A number of recent studies have investigated a phenomenon known as the hypercorrection effect (Butterfield & Metcalfe, 2001), in which errors endorsed with higher confidence are more likely to be corrected on a final test than are errors endorsed with lower confidence. This effect has been shown to occur when corrective feedback, in which the correct answer is provided, is given to all errors (Butler, Fazio and Marsh, 2010; Butler & Roediger, 2007; 2008; Butterfield & Mangels, 2003; Butterfield & Metcalfe, 2006; Eich, Stern, & Metcalfe, submitted for publication; Fazio & Marsh, 2009, 2010; Kulhavy & Stock, 1989; Kulhavy, Yekovich & Dyer, 1976; Marsh, 2011; Metcalfe, Butterfield, Habeck, & Stern, submitted for publication; Metcalfe & Finn, 2011; Sitzman & Rhodes, 2010). In the standard paradigm used to investigate this phenomenon (see, e.g., Butterfield & Metcalfe, 2001) participants are asked to generate the answers to general information questions and rate their confidence in the correctness of each answer they produced. They are then given the correct answer. In contrast to theoretical expectations, which suggest that responses in which one is highly confident should be particularly difficult to correct, the high confidence errors are more likely to be corrected on the retest than are the errors endorsed with lower confidence, even retest at a considerable delay (Butler et al., 2010; Butterfield & Mangels, 2003; Fazio, 2011). It appears, from these data, that – at least for young adults – a strong degree of belief in one's incorrect answers makes them more susceptible rather than less susceptible to being corrected.

The hypercorrection effect, itself, is surprising on theoretical grounds. Virtually all theories of memory and of the relation of memory to confidence (e.g., Gigerenzer, Hoffrage, & Kleinbolting, 1991; Hollingworth, 1913; Koriat, l997; Koriat, Goldsmith, & Pansky, 2000; Murdock, l974) indicate that responses that are made with high confidence are those in which the person believes most, and are the strongest in memory (e.g., Ebbesen & Rienick, 1998; Tulving & Thomson, l971). As such, they should be most easily accessible and most resistant to interference. Certainly, in all data presented to date on the hypercorrection effect (including in the present article), the correlation between confidence in one's first responses and the correctness of those responses is high. The responses in which people are highly confident are nearly always correct. But if such a high confidence response were in error, it, too, like correct high confident responses, should be strong, entrenched and difficult rather than easy to change.

The finding that young adults easily correct such high confidence errors (i.e., they hypercorrect) has now been replicated many times (Butler et al., 2010; Butler & Roediger, 2007; 2008; Butterfield & Mangels, 2003; Butterfield & Metcalfe, 2001, 2006; Fazio & Marsh, 2009, 2010; Kulhavy & Stock, 1989; Kulhavy et al., 1976; Metcalfe et al., submitted for publication; Metcalfe & Finn, 2011; Sitzman & Rhodes, 2010), and research has begun to focus on why the effect occurs. Two non-mutually exclusive factors have been isolated as

* Corresponding author. Tel.: +1 212 854 7971.
E-mail addresses: jm348@columbia.edu (J. Metcalfe), bridgid.finn@wustl.edu (B. Finn).
[1] Tel.: +1 646 942 0363.

0959-4752/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.learninstruc.2011.10.004

being central to the hypercorrection phenomenon in young adults. The first relates to their surprise at being wrong. Being surprised (and perhaps embarrassed) at having made a mistake on a response they strongly thought was correct, participants may rally their attentional resources to better remember the correct answer. Several lines of evidence support this surprise/attentional factor. For example, Butterfield and Metcalfe (2006) showed that when young adults were required to detect simultaneous soft tones while reading the corrective feedback, they were more likely to miss the tones during the presentation of high as compared to low confidence errors corrections. This result indicates that their attention was captured by the high confidence error feedback, leaving less capacity to detect the tones. Fazio and Marsh (2009) showed that the surrounding context that accompanied the high confidence error feedback was better remembered than was the context surrounding the low confidence error feedback. They interpreted this result as favoring the attentional explanation. And, Butterfield and Mangels (2003) showed that the p300 event related potential, thought to be related to surprise and the engagement of attention, occurred when corrective feedback was given. The magnitude of the component was graded by confidence. This finding, again, was consistent with increased attention being paid to high confidence error feedback.

The second factor that has been implicated in the hypercorrection effect is a greater familiarity with the corrections or greater semantic knowledge in the domain of the high than the low confidence errors. Butterfield and Mangels (2003) noted that there was differential participant-ascribed familiarity to high than low confidence responses, and Butterfield and Metcalfe (2006) showed that the a priori probability of a correct response was greater for high than for low confidence errors. Furthermore, Metcalfe and Finn (2011) have shown that young adults claim that they 'knew it all along' fairly frequently when they receive the corrective feedback to high confidence errors. In their study, young adults were more likely to produce a second guess that was correct following a high as compared to a low confidence error. They were also more likely to choose the correct alternative in a multiple choice test that excluded their original answer, if the error had been committed with high confidence. Additionally, they required fewer clues to guess the correct answer to questions on which they had made high as compared to low confidence errors. These results indicate that familiarity with the domain of the high confidence error, and plausibly, with the answer itself, plays a role in the hypercorrection effect.

Either or both of these two factors might be different in young adults as compared to children. Children might be less surprised or embarrassed by high confidence mistakes, and less likely to rally selective attention to correct these mistakes. In adults, high confidence errors are associated with both the p300 event related potential deflection (Butterfield & Mangels, 2006) and with anterior cingulate activation (see, Metcalfe et al., submitted for publication). While grade school children show p300 event related potential deflections, they are sometimes less pronounced than those of adults (Ladish & Polich, 1989; Polich, Ladish, & Burns, 1990). There are also age-related differences specific to the anterior cingulate (see Bush, Fan, & Posner, 2000 for review, as well as Casey et al., 1999). Given these brain differences, the surprise reaction of children might be different from that of adults. If so, they might show a difference in their responses to high confidence errors.

Children may also have a less well-developed or structurally different semantic network surrounding the high confidence errors, and hence not have the same kind of differential familiarity shown by adults to support the enhanced high confidence error correction. It is well-established that children's vocabulary itself is much less rich than that of adults. By the beginning of grade 3 children have a core vocabulary of only about 8000 words, at best, with 1000 words being added in each subsequent year (Biemiller & Slonim, 2001, and see Beck, McKeown & Kucan, 2002), depending on socio-economic status (Hart & Risley, 1995). By college age, adults have a vocabulary of approximately 17,000 words (D'Anna, Zechmeister, & Hall, 1991). Differences in the richness of the semantic network might (or might not – since the structure may still be similar even though the network is less rich) impact the processes related to familiarity with the correct target answer that supports hypercorrection in young adults (Metcalfe & Finn, 2011).

Finally, although the hypercorrection effect has been replicated many times with young adults, not all people show the effect. Eich et al. (submitted for publication) have shown that older adults, for example, fail to show the hypercorrection effect. Furthermore, Metcalfe et al. (submitted for publication) have shown, in a fMRI study, that prefrontal areas (as well as other areas, of course) of the brain are implicated in the hypercorrection effect in young adults. Both elders, because these areas are highly vulnerable to the negative effects of aging, and children, because these prefrontal areas may not yet be fully developed (see Bush et al, 2000), show processing impairments commonly associated with prefrontal cortex. It is plausible, then, to hypothesize that children, like older adults, might fail to show a hypercorrection effect to the extent that it is associated with frontal function. However, the reasons that older as compared to younger adults differ in hypercorrection are not yet fully understood. It is not known, for example, if the prefrontal cortex is implicated or if the difference is due to other factors. Insofar as young children differ from elders in many other ways, it is possible that children will not show the same effects as elders. Whether children do or do not correct high confidence errors with ease is, of course, educationally important. However, despite many replications of the effect in college-aged participants, it has never before been investigated in children.

Whether children show this effect or not is the primary question addressed here. Hypothesis 1 was that children, like young adults, would hypercorrect high confidence errors. The alternative hypothesis, a real possibility, was that they would not. In the experiment that follows we use the procedure of Metcalfe and Finn (2011) to investigate whether children show the hypercorrection effect, and also whether they think that they selectively knew the answers all along when they are given the answers to their high confidence errors. Hypothesis 2 was that children would claim to have known the answers to high confidence errors all along, once they were provided with corrective feedback.

1. Experiment 1

In Experiment 1, we tested Grade 3–6 children. The children were given general information questions and answered, giving their confidence in each response, until they had made 16 errors. After each error and confidence judgment, they were given corrective feedback followed immediately by the question of whether they had known the answer all along – a procedure previously used by Metcalfe and Finn (2011). Then a final cued recall test was given.

1.1. Method

1.1.1. Participants

The participants were 44 students enrolled in grades 3–6 in Bronx and Manhattan borough New York City public schools. They participated in this experiment as part of a research-related after school program investigating children's learning and metacognition. All participants were treated in accordance with APA ethical guidelines in this experiment and the experiments that follow.

1.1.2. Materials

Participants were asked general information questions from a pool of 93 questions, some of which had been taken from Nelson and Narens (1980) list of general information questions, although the most difficult or historically dated questions – questions that the children would have no chance of answering – were eliminated. We also included questions that had the same general format as those on Nelson and Narens' list, except that they were easier (such as "What animal is known as 'man's best friend'?"). In addition, some questions were added that related directly to the children's studies, such as "What famous civil rights leader made a speech that began with the words 'I have a dream'?"

1.1.3. Procedure

At the beginning of the experiment participants were instructed that they would be answering general information questions and indicating how sure they were of their answers, and that they would then be given the correct answers. They were encouraged to guess even if they did not know the answer. They were not told that there would be a retest at the end of the experiment. Each child was tested separately by an adult research assistant. The research assistant explained the procedures to the child, and gave them an example. During the initial test phase participants were presented, one at a time, with general information questions. The child gave their answer, and the research assistant entered it into the blank slot on the computer. The children were then asked to provide a confidence rating concerning the correctness of the answer, by pointing to the place on a horizontal slider on the computer screen that ranged from "very unsure" on the left end to "very sure" on the right end that indicated their confidence. The slider bar was set to the middle of the scale at the onset of each question, to not bias the decision. Confidence ratings were coded, by the computer, along a scale from 0 to 1.00, with 0 indicating a selection of the lowest limit of the slider, at the very unsure end, and 1.00 indicating a selection of the highest limit, at the very sure end. After participants had made the confidence rating by pointing to the screen (and this had been entered in to the computer by the research assistant) feedback was given. When the participant's answer was correct a chime sounded and the next general information question was presented. If, however, the answer was incorrect, the correct answer was presented on the screen and participants were asked to indicate whether they knew that answer all along. To so indicate, as before, participants pointed to a position on a second slider to indicate the extent to which they knew the answer all along, and the research assistant entered in this response to the computer. This slider, which was also initialized to the center position, ranged from "That's new to me" on the far left end to "I actually knew it all along" on the far right end. After they made their 'knew it all along' judgment the next general information question was presented. This process continued until the child had answered 16 items incorrectly. These 16 items became the items over which their original confidence in their errors, as well as their 'knew it all along' judgments were computed. Once 16 incorrect items had been accumulated, the program randomized those 16 and retested each in a final cued recall test.

At the end of the experiment, the children were thanked and told what the experiment was about and asked if they had any questions.

1.2. Results

1.2.1. Basic data

On average the children answered 27.57 ($SE = 1.00$) questions before they reached the 16 incorrect answer criterion. The children's initial confidence in their answers was .50 ($SE = .03$). To evaluate whether these confidence ratings were predictive of initial test performance a gamma correlation, which, is a non parametric correlation coefficient ranging from −1 to 1, which relates rankings (given on the confidence scale) to response correctness for each item (see Nelson & Narens, 1980) was computed for each child. The mean gamma correlation between confidence in their answers and initial recall performance was $G = .79$ ($SE = .03$), which was significantly greater than zero, $t(42) = 28.91$, $p < .01$. (In this and subsequent analyses we were sometimes unable to report a gamma correlation for some participants because some got everything right or everything wrong, or had too many ties and the statistic could not be computed. Thus, degrees of freedom listed for gamma correlations may differ from the total number of participants used in the experiment.) The children showed very good basic metacognition in this experiment as given by this resolution measure.

For the items that were answered incorrectly on the initial test, mean pre-feedback confidence in the incorrect responses was .31 ($SE = .03$), with .73 of the errors being assigned confidence ratings lower than .50, and .26 being assigned values between .51 and 1.0. Mean post feedback recall performance on the final test was .74 ($SE = .03$).

1.2.2. The hypercorrection effect

A hypercorrection effect would be in evidence if high confidence errors were more likely to be corrected on the final test than errors endorsed with lower confidence. The children showed a significant hypercorrection effect, providing support for Hypothesis 1: The mean gamma correlation between confidence in the original error and retest accuracy was $G = .22$, $SE = .08$, which was significantly greater than zero, $t(39) = 2.90$, $p < .01$, $d = .46$. This result established for the first time that children show a hypercorrection effect.

1.2.3. Knew it all along judgments

Children's mean 'knew it all along' judgment was .44 ($SE = .03$). The question of interest, concerning the possibility that high confidence errors were thought to be known all along, preferentially, was whether the corrections to the errors that had been endorsed with high confidence were given higher 'knew it all along' judgments than were the corrections to the errors endorsed with low confidence. This gamma was greater than zero, $G = .16$ ($SE = .05$), $t(43) = 3.59$, $p < .01$, $d = .54$ providing support for Hypothesis 2. A further assessment showed that the mean 'knew it all along' judgment to the corrective feedback was higher for high confidence errors, than for low confidence errors, when high and low confidence were classified as judgments of .50 and greater for high confidence, and judgments of .49 or lower, for low confidence (Low Confidence, Knew it all along mean: $M = .40$, $SE = .04$, High Confidence, Knew it all along mean: $M = .55$, $SE = .04$, $t(39) = 3.23$, $p < .01$). Thus, the children were more likely to claim that they knew the answer all along, post feedback, when they had made a high as compared to a low confidence error.

Finally, we computed the correlation between knew it all along judgments and final test performance. The gamma correlation was .40 ($SE = .08$), $t(39) = 5.39$, $p < .01$, $d = .79$. Thus, when they said they knew it all along, the children were more likely to get the answer correct later.

1.2.4. Mediation analyses

To examine the relationship among confidence, knew it all along judgments and final test performance further, we used a mediational model in which we assessed whether the effect of confidence on final test performance was mediated by the knew it all along judgments. Following the technique recommended by Baron and Kenny (1986), we found evidence of mediation of the impact of confidence on final test performance. There was a significant effect of confidence judgments on final test performance, $\beta = .18$, $t(659) = 4.66$, $p < .05$, and on knew it all along judgments, $\beta = .24$, $t(659) = 6.40$, $p < .001$. There was also a significant relationship between knew it all along judgments and final test performance $\beta = .32$, $t(659) = 8.53$, $p < .001$. As shown in Fig. 1, when both confidence judgments and knew it all along judgments were included as predictors in the regression equation, knew it all along judgments still predicted final test performance, $\beta = .29$, $t(659) = 7.63$, $p < .001$, as did confidence judgments, $\beta = .11$, $t(659) = 2.87$, $p < .01$. However, the decrease in the direct effect of confidence on final test performance was statistically significant, as measured by a Sobel test, $z = 4.90$, $p < .001$, again, indicating that the effect of confidence on final test performance was partially mediated by the knew it all along judgments.

1.3. Discussion

This experiment showed, for the first time, that, consistent with Hypothesis 1, children exhibit the high confidence error hypercorrection effect just as do young adults. Children were more likely to correct their high than low confidence errors, given feedback. Furthermore, they claimed upon viewing that feedback, that they knew the answers all along with a higher likelihood for the high than the low confidence errors. This experiment also showed that in nearly all other respects tested, children's metacognitive behavior was similar to that of adults. They had highly accurate confidence ratings: when they believed they had given the correct answer, they were likely to have given the correct answer, and when they believed they had not given the correct answer they were likely to have not given the correct answer. Furthermore, they took advantage of corrective feedback, with performance on questions that had been wrong soaring to over 74% correct.

These results suggest that children, at least by the middle grade levels, are metacognitively very competent, behaving much as do adults, and are not much set back by making errors, so long as those errors are accompanied by corrective feedback. Children's high confidence errors were not more resistant to correction than low confidence errors. In fact, as has also been shown by young adults, these high confidence errors were more readily corrected. The results also suggested that the children believed, very frequently, and consistent with Hypothesis 2, they knew the correct answers all along. Whether their claim that they knew the answers all along was a pure hindsight bias, as has previously been demonstrated in children (Bernstein, Atance, Meltzoff, & Loftus, 2007) or whether it was an indication that they really did know, is the issue that is investigated in the next two experiments.

2. Experiment 2

In the second experiment we sought to determine whether the children's knew it all along judgments were veridical, in the sense that they would be able to show that they did actually know the answers, selectively, to their high confidence errors, before they were given corrective feedback. It was not necessarily the case that they would be able to do this. Much research has shown that after people learn the outcome to a situation or the answer to a question, they tend to exaggerate their ability to have anticipated it: they often claim to have known the answer all along (see Hawkins & Hastie, 1990; Hoffrage & Pohl, 2003; Sanna & Schwartz, 2006). It was possible that the post feedback knew it all along judgments we obtained in Experiment 1 were simply a demonstration of such a classic hindsight bias (Fischhoff, 1975; Wood, 1978).

Experiment 2 could potentially reveal either that the children had knowledge of the correct answers all along when they said they did, or that the claim that they had made in the first experiment was purely a hindsight bias. Taking seriously the idea that they might have known the answers all along at time of the first test when asked a question like "What is the largest ocean in the world" to which they had given a highly confident response like "Atlantic", we thought they might have been able to come up with the correct answer if we had them slow down, think hard, and try again. We expected that the children would be more likely to generate the correct answer to questions that had evoked high rather than low confidence errors. But even if they were unable to generate the correct response themselves, they still might have been able to correctly select the answer, from a list of alternatives, especially when they had made a high rather than low confidence error. These two conditions: the Generate condition, and the Multiple Choice condition, were contrasted to the Standard Feedback condition (needed for comparison and to replicate the hypercorrection result) in Experiment 2. Hypothesis 3 was that both the Generate condition and the Multiple Choice condition would show that children did have some selective knowledge of the answers to high as compared to low confidence errors, in advance of being given the correct answers.

2.1. Method

2.1.1. Participants

Twenty-one children who were students in New York City public and private schools and who were in grades 3–5 were the participants. Data from three additional children were eliminated because they were unable to complete the task. None of the children in this experiment had participated in the earlier experiment.

2.1.2. Procedure

Participants answered general information questions until they had reached 27 errors. Some of the most difficult questions were eliminated from the pool used in Experiment 1, and some easier pop culture and general information questions were added after several children had been run to make the task easier for the

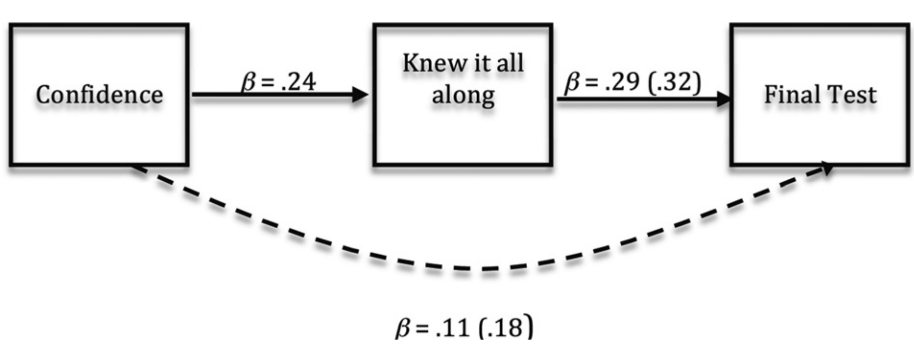

Fig. 1. Mediational model of direct and indirect effects of confidence and knew it all along judgments on final test performance. Values in parentheses indicate the direct effect before the mediator was included in the analysis. All values were significant at $p < .05$, from Experiment 1.

children. The data from children who completed the task with the more difficult questions, however, were not different from those with the slightly easier set, and, hence, were included in the analyses that follow. During this first test, questions were presented one at a time, in a random order. After each question, participants indicated a response and a confidence judgment about the correctness of their response. The answers were computer scored online. Immediately after making the confidence judgment participants heard a chime if the response was correct and moved on immediately to the next question. If the response was incorrect a lower pitched honk sounded and one of three within participants feedback conditions occurred, as detailed below. Immediately following the feedback treatment, participants made a knew it all along judgment in which they were asked 'Did you know that all along?' and had to hit either a 'yes' or 'no' button. We changed to two buttons, instead of using the continuous scale that had been used in Experiment 1 for this judgment, because the decision was binary. Participants moved to the next question following this judgment.

The three conditions were: (1) Standard Feedback, (2) Generation, and (3) Multiple Choice. Which feedback condition was given to which error was randomly determined with the constraint that all three conditions occurred before the next permutation of conditions began. There were 9 replications in each of the 3 treatment conditions, for each child, in the entire experiment. In the *Standard Feedback* condition, after indicating their confidence in their incorrect answer, the children were simply told, "Actually the correct answer is x", and the correct answer was presented in the response window on the computer screen, read aloud by the research assistant.

In the *Generation* condition, the participants were told, "Please choose another answer. If you do not know the answer, please guess." After they guessed, and the research assistant typed in their new response, they made a 'knew it all along' judgment and then moved on to the next question. They were not told whether their response was correct or not. This condition was directed at the possibility that high confidence errors were the result of impulsiveness, and that if the children were simply told that they were incorrect and asked to give another response they might be able to produce the correct answers.

In the *Multiple Choice* condition, a message read by the research assistant said, "Actually, the answer is one of these 6 options. Please choose one." A randomized array of six options, including the correct answer, was presented and read to the student who then chose a new response. The program randomly selected the six options from a set of nine potential options. If the participant's original error was included in the list of 6 options first selected by the computer, that option was replaced, randomly, with one of the remaining 3 options.

At the end of the 27th error, there was a short distractor task, and then there was a second test. Note that at the time of the second test for the Standard Feedback condition, the participants had previously been provided with the correct answer but for the Generation condition and the Multiple Choice condition, they had not yet been given corrective feedback. If the participant got the answer correct on the second test, the chime feedback sounded. If he or she got it wrong, regardless of condition, corrective standard feedback was given, that is, they were shown the correct answer. This was done, in part, to allow the children a better chance at learning the answers by the end of the experiment, and having a more rewarding (less frustrating) experience. It also allowed us to look at the effect of our manipulations in conjunction with corrective feedback, following those manipulations.

Once all 27 questions had been tested in the second test, and corrective feedback given, the participants were then given a third (and final) test on all 27 questions. Performance was much better on this third test, because the children had at this time received corrective feedback on all errors. Note, though, that in the Multiple Choice and the Generation condition they had received corrective feedback only once before getting this third test, whereas in the standard condition they had received corrective feedback twice. At the end of this third test, the computer told the children their overall score on the previously unknown general information facts, and they were generally pleased with how much they had learned.

A research assistant sat with each child participant throughout the experiment, and entered the answers and ratings into the computer, so that spelling mistakes did not alter the results. The research assistants also encouraged the children and kept them on task.

2.2. Results

2.2.1. Basic data

On average the children answered $M = 35.14$, $SE = 1.47$ questions on the first test before they reached the 27 incorrect answer criterion. Mean confidence was .30, $SE = .02$, and among the errors, .85 were between confidence ratings of 0 and .50, .14 were between .51 and 1.00. Confidence ratings were predictive of their first test performance. The mean gamma correlation between initial confidence ratings and first test recall performance was .81, $SE = .07$ and was significantly greater than zero, $t(20) = 11.38$, $p < .01$.

Using only the 27 items that were answered incorrectly on the initial test, we computed test performance on the *first post-corrective-feedback test*. This was the second test for the Standard Feedback condition and the third test for the Multiple Choice condition and the Generation condition. There was a significant effect of feedback condition on first post-corrective feedback test performance, $F(2, 40) = 3.67$, $MSE = .03$, $p < .05$, $\eta_p^2 = .16$, with the test performance being $M = .52$, $SE = .06$, in the Multiple Choice condition, $M = .42$, $SE = .05$ in the Generation condition, and $M = .40$, $SE = .04$, in the Standard Feedback condition. The Multiple Choice condition was significantly different from the Standard Feedback condition, $p < .05$, but was not different from the Generation condition, $p > .05$.

Performance on the second test, which was given after the participants had had a chance to take a second guess at the answer in the Generation condition, or had selected their choice in the Multiple Choice condition or after they had been given corrective feedback in the Standard Feedback condition, indicated that merely giving the children a second chance at the correct answer was no substitute for giving them corrective feedback, $F(2, 40) = 38.80$, $MSE = .02$, $p < .001$, $\eta_p^2 = .66$. Performance in the Standard Feedback condition, $M = .40$, $SE = .04$, was significantly better than performance in the Generation condition, $M = .07$, $SE = .03$, ($t(20) = 7.47$, $p < .001$) and in the Multiple Choice condition, $M = .19$, $SE = .03$, ($t(20) = 5.55$, $p < .001$). Performance was also significantly better when the answer could be selected from alternatives in the Multiple Choice condition than in the Generation condition where no additional information was provided, $t(20) = 4.01$, $p < .01$. The low score in the Generation condition reflects the fact that participants had nearly always omitted or gotten the answer wrong, when they were given a second chance to try to self generate the answer. The low performance in the Multiple Choice condition suggested that the children tended to remember the answers they had chosen on the multiple choice test but that these answers were usually wrong. Thus, having been given standard corrective feedback produced greatly superior test performance to either trying to generate the response (without feedback) or choosing a response from multiple alternatives (without feedback).

Performance on the third test in the Standard Feedback condition was $M = .56$, $SE = .05$. As noted above, performance on the third test in the Multiple Choice condition was .52, and performance in the Generation condition was .42. There was a main effect of condition $F(2, 40) = 4.26$, $MSE = .03$, $p < .05$, $\eta_p^2 = .18$. Performance in the Standard Feedback condition was significantly better than performance in the Generation condition, $t(20) = 3.22$, $p < .01$. No other comparisons were significant, smallest $t(20) = 1.64$. Recall, though, that by the time of the third and final test, the standard feedback condition had received two rounds of corrective feedback whereas the correct answer had only been given once in the Generation and the Multiple Choice conditions.

2.2.2. Hypercorrection

There was a hypercorrection effect, providing further support for Hypothesis 1. The mean gamma correlation between confidence in the original error and first post-corrective-feedback test accuracy was $G = .20$, $SE = .10$, overall, which was significantly greater than zero, $t(20) = 2.00$, $p = .05$. The gammas for the three separate conditions were $G = .25$, $SE = .16$, $G = .20$, $SE = .19$, and $G = .20$, $SE = .14$, for the Standard Feedback condition, the Generate condition and Multiple Choice condition, respectively, which were not different from one another, $F < 1$, power to detect the effect was $1 - \beta = .49$.

2.2.3. Did they know it all along?

As noted above, the hypercorrection effect obtained in the Standard Feedback condition. The claim that they ‘knew it all along’, though, was not significantly greater for high confidence errors than low confidence errors, Low Confidence: $M = .14$, $SE = .05$, High Confidence: $M = .25$, $SE = .12$, $t(11) = 1.08$, $p > .05$. Although the means were in the right direction, the effect could not be computed for many participants, because they had originally given only very low confidence ratings. The power to detect the effect was $1 - \beta = .05$. The gamma correlation between the knew it all along judgment and confidence was .02, $SE = .21$, and not significantly different from zero, $t < 1$. Even so, participants were more likely to be correct on the first post-corrective-feedback final test if they had claimed that they knew the answer all along ($G = .59$, $SE = .19$, $t(14) = 2.98$, $p = .01$).

There were only 8 correct second guess response produced in the Generation condition, over the entire experiment, $M = .05$, $SE = .02$. Only four of the children produced even a single correct second guess. Of these 8 correct second guess responses, 5 stemmed from high confidence errors and 3 were low confidence when we considered high confidence to be higher than .5 on the rating scale. With so few responses generated, we cannot interpret the results, and they provide no support for Hypothesis 3. The probability of producing the correct answer on the third test was, of course, higher when the answer had been produced on the second guess: Incorrect second guess: $M = .35$, $SE = .16$, Correct second guess: $M = .92$, $SE = .08$, $t(3) = 3.63$, $p < .05$. There was no significant difference between the second guess performance and test performance in this condition, and this was also not different a function of confidence.

Despite showing a hypercorrection effect in the Multiple Choice condition, the children showed no evidence of selectively knowing the answers all along, with the high, as compared to low confidence errors. While they were above chance on selecting the correct answer in the multiple choice task (chance was .167, and their overall mean was .30, $SE = .09$) there was no difference in the probability of choosing the correct response when an error had been produced with high as compared to low confidence. A lack of difference was observed when that confidence was measured by dividing at .5 on the scale (Low Confidence: $M_{correct} = .29$, $SE = .05$, High Confidence: $M_{correct} = .30$, $SE = .09$, $t < 1$) and when it was assessed by taking a gamma correlation between original confidence and whether the multiple choice selection was correct or incorrect, $G = .14$, $SE = .12$, $t(17) = 1.15$, $p > .05$, power to detect the effect was $1 - \beta = .05$. The probability of producing the correct answer on the second test was higher when the correct choice was made on the multiple choice test than when it had not (Incorrect: $M = .06$, $SE = .03$, Correct: $M = .41$, $SE = .07$, $t(19) = 4.38$, $p < .01$); but it was not higher when the original error had been made with high confidence. When the child had chosen a correct answer on the Multiple choice test, the probability of a correct answer on the third test was also higher than when the incorrect choice was made on the multiple choice test, Incorrect: $M = .40$, $SE = .07$, Correct: $M = .78$, $SE = .06$, $t(19) = 4.48$, $p < .01$.

2.3. Discussion

Experiment 2 showed in all three conditions that the post feedback hypercorrection effect was exhibited in children, providing additional support for Hypothesis 1. However, the experiment provided no support for Hypothesis 3, the ‘knew it all along’ claim. The children were able to generate almost nothing, in the Generation condition, and so this test provided no evidence that they actually did know the answers all along, selectively. The data in this condition do contravene the idea that children’s high confidence errors resulted because they had simply been impulsive and blurted out the first, wrong, thing that came to mind and could have exhibited greatly improved performance had they been less impulsive. When they were slowed down and asked to generate a second now-correct response, they were almost never able to do so. There are however, well-known age-related differences in verbal fluency (Regard, Stauss & Knapp, 1982; Welsh, Pennington, & Grossier, 1991). Difficulties in word generation, in general, might have made it difficult to detect whether the children might have known something about the answers all along by simply asking them to generate a second response.

Furthermore, while the children, when given a multiple choice second guess test picked the correct answers at a rate that was above chance, there was no difference in correct selection between the questions that had elicited high as compared to the low confidence errors. It is notable, however, that the power of the multiple choice second-guess test was low, and this test is less likely to show reliable ‘knew it all along’ differences than other measures, even when adults are tested (see, Metcalfe & Finn, 2011). Thus, while there was solid evidence from all three conditions of this experiment that children hypercorrect, there was little evidence favoring Hypothesis 3: that they selectively knew the answers all along for high confidence errors. But the results may not have been decisive.

3. Experiment 3

In this final experiment, we again sought to replicate the hypercorrection finding in children, and to see whether partial information might underlie children’s claim that they ‘knew it all along.’ The previous experiment had demonstrated the difficulties that children had in freely generating responses, or in demonstrating that they knew it all along under second guess multiple choice conditions. However, they still might have been able to demonstrate more partial knowledge for the items that they said they knew all along, if they were tested with a more sensitive method. To investigate the idea that the knew it all along assertion might have stemmed from partial knowledge, we used a scaffolding paradigm (Carpenter & DeLosh, 2006; Finn & Metcalfe, 2010). Adults had previously shown strong positive ‘knew it all along’ effects when tested with this method (Metcalfe & Finn, 2011). It seemed possible that children, like adults, might need fewer hints to guess the right answer when they had made high confidence errors and when they had said they knew it all along.

In this paradigm, after committing an error, the children were first asked to guess again. If they could not guess – which from the previous experiment we expected would usually be the case – then the computer would provide a hint in the form of the first letter of the answer, at which point the children were asked to guess again. If they could not get the answer with one letter, the computer would provide another, and another until either the child had correctly guessed the answer or the complete target was unveiled. Some preferential advance knowledge for the corrections to high confidence errors would be demonstrated if a smaller proportion of the word needed to be unveiled for high as compared to low confidence errors. This scaffolding paradigm is also of considerable pedagogical interest since Finn and Metcalfe (2010) showed that providing feedback in this way resulted in considerably better long term learning than did either standard feedback, answer until correct multiple choice feedback, or no feedback.

3.1. Method

Participants were 17 Grade 3, 4 and 5 children enrolled in a public school or a private school in New York City. One participant was removed for failing to comply with instructions leaving a total of 16 participants. In this experiment there was only one condition – the scaffolded feedback condition – allowing us to obtain as much relevant data as possible. Children answered general information questions, and gave their confidence about their answers, receiving immediate scaffolded feedback about the correct answer, after having committed an error. First they were asked to make a second guess. If that was not successful, they were given the first letter, and again asked to make a guess. Then they were given the second letter, third letter, and so on, with presentation of each successive letter intervened by an opportunity to guess the target, until they had correctly guessed the target. They answered questions and received scaffolded feedback until they had made a total of 16 errors. They then had a final test on these 16 questions.

3.2. Results

3.2.1. Basic data

On average the children answered 23.87 ($SE = 1.49$) questions before they reached the 16 incorrect answer criterion. Their confidence ratings were predictive of their initial test performance. The mean gamma correlation between initial confidence ratings and initial recall performance was .90 ($SE = 03$), and was significantly different from zero, $t(15) = 31.96$, $p < .001$. Their mean confidence was .50, $SE = .03$ and .69 of their errors had confidence ratings below .50 while .31 were between .51 and 1.0. Performance on the final test, which followed the scaffolded feedback procedure, was .54, $SE = .06$.

There was a hypercorrection effect. Errors that were committed with higher confidence were more likely to be corrected than errors committed with lower confidence, $G = .18$, $SE = .10$, $t(15) = 1.77$, p(one-tailed) $< .05$. There was also a 'knew it all along' belief effect, such that participants were more likely *to say* they knew it all along (once they had seen the correct answer) to high confidence errors than to low confidence errors, when confidence was split based on the center of the scale, (Low Confidence: $M = .33$, $SE = .04$, High Confidence: $M = .51$, $SE = .09$, $t(13) = 2.03$, p(one-tailed) $< .05$), although the gamma correlation computed between confidence and whether or not the child said that they knew it all along did not reach significance, $G = .16$, $SE = 12$, $t(15) = 1.40$, $p > .05$, power to detect the effect was $1 - \beta = .46$. Finally, participants were more likely to be correct on the final test if they had claimed that they knew the answer all along, $G = .33$, $SE = .14$, $t(15) = 2.27$, $p < .05$.

3.2.2. Did they know it all along?

When the children had committed a high confidence error the number of letters that they needed to produce the correct answer was significantly fewer than when they had produced an error with low confidence (Number of letters required to correct answer |High confidence error: $M = 3.86$, $SE = .49$; Number of letters required to correct answer |Low confidence error: $M = 5.01$, $SE = .30$, $t(13) = 2.85$, $p < .05$, with 50/50 split. These results provide support for Hypothesis 3.

The average length of the correct answer was not significantly longer for low confidence errors ($M = 6.79$ letters, $SE = .22$) than for high confidence errors ($M = 6.28$ letters, $SE = .26$, $t(15) = 1.60$, $p > .05$, with 50/50 split). Although the high and low confidence errors did not differ significantly in word length, we nevertheless computed for each response, the proportion of the target word that was necessary to generate a correct response. Splitting the data at .5 on the confidence scale, the analysis showed that the proportion of the word that needed to be revealed for correct responding given a high confidence error was .61, $SE = 07$, whereas the proportion of the word that needed to be revealed for correct responding given a low confidence error was .75, $SE = .03$, ($t(13) = 2.10$, p(one-tailed) $< .05$). These results, again, favor the idea that the children did have some partial information about the answers all along. Although these results indicate that the children had differential partial knowledge for the correct answer to high confidence errors, the result does not speak to whether the children subscribed to or 'knew' that the answer that they had guessed was, in fact, correct. It does appear that the children were neither confabulating, nor simply prone to a preferential hindsight bias, however, when they said they knew something extra about the answers all along. They did have preferential partial knowledge for high as compared to low confidence error corrections.

4. General discussion

These three experiments investigated whether children's subjective confidence in their errors plays a role in which errors are most likely to be amended. In young adults, errors that are endorsed with high confidence are hypercorrected following corrective feedback. The present results indicate that the hypercorrection effect obtains with grade school children as well as with young adults. In the experiments presented here, every condition in each of the three experiments showed a hypercorrection effect, for a total of 5 replications. Thus, this article establishes for the first time that children at the grade 3 to 6 level hypercorrect.

Since children show the hypercorrection effect, further discussion of differences between them and young adults concerning why they might not show the effect is unnecessary. However, children did show less facility, than adults, in exhibiting that they knew the answers all along. Did the children, then, really 'know all along' the answers to high confidence errors? The tests that we implemented in Experiment 2 to determine whether they actually knew those answers all along showed that the children were almost completely unable to generate the answers to any of the questions on which they had erred, either with high or low confidence. Neither did they choose the correct answers selectively on a second-guess multiple choice test. While the poor multiple choice test second guess performance might have been due to distracting interference that such tests evoke, the children's almost complete failure to generate anything at all as an alternative to their incorrect responses may stem from children's well-established lack of verbal fluency. Generating any kind of response is difficult for them. The full blown correct answers, then, do not seem to be available to the children despite their claim that upon hearing the answers they knew them all along.

But even though the children were almost completely unable to spontaneously generate or even select the corrections to high confidence erroneous responses, the third experiment provided supporting evidence that their claims that they knew something about the answers were not completely unfounded. When they were tested with a very sensitive method, the children did demonstrate differential partial information. The scaffolding technique (Finn & Metcalfe, 2010) in which successive letters were given and children were asked to guess the correct answer after each letter was provided, showed that fewer letter hints were needed to evoke correct guesses from the children for high as compared to low confidence errors –a result similar to that of adults. This is an important result, especially given the data of Finn and Metcalfe (2010) showing that this scaffolding procedure also resulted in enhancements in memory that were superior to any other method of feedback tested. Thus, while the children did not demonstrate that they could fluently produce the full blown correct responses, they did demonstrate that they had more partial information about the correct answers associated with high confidence errors.

Many researchers have demonstrated large memory benefits as a result of the processes involved in testing (Butler & Roediger, 2007; McDaniel, Anderson, Derbish, & Morrisette, 2007; McDaniel & Fisher, 1991; McDaniel, Roediger, & McDermott, 2007; Roediger & Karpicke, 2006a,b). Testing effects occur despite inaccuracies in confidence (Boekaerts & Rozendaal, 2010). The main caveat that might have qualified the use of such testing in educational settings as a means to enhance children's memory was that the errors that the children would necessarily commit when tested might have had highly detrimental memorial consequences. This could be particularly problematic for high confidence errors, which, theoretically, might be considered to be difficult to correct. The data presented here mitigate that concern, at least for normal children. Indeed, like young adults, the errors that children commit, and especially those committed with high confidence, are corrected very easily as long as corrective feedback is provided.

A final concern remains. While our data provide no cause for worry about errors committed by typical children, errors may have a more detrimental effect in special populations. People with memory disorders (Baddeley & Wilson, 1994; Glisky, Schacter, & Tulving, 1986) are particularly derailed when they make errors, and so caution is needed in applying the present findings to children with learning disorders, who may also respond differently than typical children, and who have not been tested in this paradigm. More research on these special populations is needed to determine the limitations and boundary conditions of the conclusions we reach here concerning the effects of errors. Even so, as we have shown here with typically developing middle childhood children, the avoidance of generation or testing procedures would not seem to be justified on the grounds that such procedures inevitably result in problematic, strong, high confidence errors that will be difficult to overcome. Indeed, with just a few moments of corrective feedback, the errors, and especially the errors in which the children expressed high confidence, are not recommitted, but rather are corrected at a very high rate.

Acknowledgments

This research was supported by grant R305H060161 from the Institute of Education Sciences of the Department of Education and by Grant 220020166 from the James S. McDonnell Foundation. The authors are solely responsible for the results reported herein. We thank Lisa Son and David Miele for their help and comments.

References

Baddeley, A., & Wilson, B. A. (1994). When implicit learning fails: amnesia and the problem of error elimination. *Neuropsychologia, 32*, 53–68.
Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: conceptual, strategic and statistical considerations. *Journal of Personality and Social Psychology, 51*, 1173–1182.
Beck, I. L., McKeown, M. G., Kucan, L., & , New York. (2002). *Bringing words to life: Robust vocabulary instruction*. The Guilford Press.
Bernstein, D. M., Atance, C., Meltzoff, A. N., & Loftus, G. M. (2007). Hindsight bias and developing theories of mind. *Child Development, 78*, 1374–1394.
Biemiller, A., & Slonim, N. (2001). Estimating root word vocabulary growth in normative and advantaged populations: evidence for a common sequence of vocabulary acquisition. *Journal of Educational Psychology, 93*, 498–520.
Boekaerts, M., & Rozendaal, J. S. (2010). Using multiple calibration indices in order to capture the complex picture of what affects students' accuracy of feeling of confidence. *Learning and Instruction, 20*, 372–382.
Bush, G., Luu, P., & Posner, M. I. (2000). Anterior cingulate development: cognitive and emotional influences in anterior cingulate cortex. *Trends in Cognitive Science, 4*, 215–223.
Butler, A. C., Fazio, L. K., & Marsh, E. J. (2010, Nov). The hypercorrection effect persists over a week, but high-confidence errors return. *Poster presented at the 50th annual meeting of the Psychonomic Society*, St. Louis, MO.
Butler, A. C., & Roediger, H. L., III (2007). Testing improves long-term retention in a simulated classroom setting. *European Journal of Cognitive Psychology, 19*, 514–527.
Butler, A. C., & Roediger, H. L., III (2008). Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. *Memory and Cognition, 36*, 604–616.
Butterfield, B., & Mangels, J. A. (2003). Neural correlates of error detection and correction in a semantic retrieval task. *Cognitive Brain Research, 17*, 793–817.
Butterfield, B., & Metcalfe, J. (2001). Errors committed with high confidence are hypercorrected. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 1491–1494.
Butterfield, B., & Metcalfe, J. (2006). The correction of errors committed with high confidence. *Metacognition and Learning, 1*, 1556–1623.
Carpenter, S., & DeLosh, E. L. (2006). Impoverished cue support enhances subsequent retention: support for the elaborate retrieval explanation of the testing effect. *Memory and Cognition, 34*, 268–276.
Casey, B. J., Trainor, R., Giedd, J., Vauss, Y., Vaituzis, C. K., Hamburger, S., et al. (1997). The role of the anterior cingulate in automatic and controlled processes: a developmental neuroanatomical study. *Developmental Psychobiology, 3*, 61–69.
D'Anna, C. A., Zechmeister, E. B., & Hall, J. W. (1995). Toward a meaningful definition of vocabulary size. *Journal of Literacy Research, 23*, 109–122.
Ebbesen, E. B., & Rienick, C. B. (1998). Retention interval and eyewitness memory for events and personal identifying attributes. *Journal of Applied Psychology, 83*, 745–762.
Eich, T. S., Stern, Y., & Metcalfe, J. (2011). Older adults do not hypercorrect. *Psychology and Aging*, submitted for publication.
Fazio, L. K. (2011). Using feedback to correct students' misconceptions. In *Association for psychological science annual meeting, invited symposium on applying learning research to education, May, 28, 2011*.
Fazio, L. K., & Marsh, E. J. (2009). Surprising feedback improves later memory. *Psychonomic Bulletin and Review, 16*, 88–92.
Fazio, L. K., & Marsh, E. J. (2010). Correcting false memories. *Psychological Science, 21*, 801–803.
Finn, B., & Metcalfe, J. (2010). Scaffolding feedback to maximize long-term error correction. *Memory and Cognition, 38*, 951–961.
Fischhoff, B. (1975). Hindsight/foresight: the effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance, 1*, 288–299.
Gigerenzer, G., Hoffrage, U., & Kleinbolting, H. (1991). Probabilistic mental models: a Brunswikian theory of confidence. *Psychological Review, 98*, 506–528.
Glisky, E. L., Schacter, D. L., & Tulving, E. (1986). Learning and retention of computer related vocabulary in memory-impaired patients: method of vanishing cues. *Journal of Clinical and Experimental Neuropsychology, 8*, 292–312.
Hart, B., & Risley, T. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore: Paul H. Brookes Publishing.
Hawkins, S. A., & Hastie, R. (1990). Hindsight: biased judgments of past events after the outcomes are known. *Psychological Bulletin, 107*, 311–327.
Hoffrage, U., & Pohl, R. (2003). Research on hindsight bias: a rich past, a productive present, and a challenging future. *Memory, 11*, 329–335.
Hollingworth, H. L. (1913). Experimental studies in judgments. *Archives of Psychology, 29*(Whole Issue).
Koriat, A. (1997). Monitoring one's knowledge during study: a cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General, 126*, 349–370.

Koriat, A., Goldsmith, M., & Pansky, A. (2000). Toward a psychology of memory accuracy. *Annual Review of Psychology, 51*, 481–537.
Kulhavy, R. W., & Stock, W. A. (1989). Feedback in written instruction: the place of response certitude. *Educational Psychology Review, 1*, 279–308.
Kulhavy, R. W., Yekovich, F. R., & Dyer, J. W. (1976). Feedback and response confidence. *Journal of Educational Psychology, 68*, 522–528.
Ladish, C., & Polich, J. (1989). P300 and probability in children. *Journal of Experimental Child Psychology, 48*, 212–223.
Marsh, E. J. (2011, May) Using feedback to correct students' misconceptions. *Paper presented at the 23rd annual convention of the American Psychological Society*, Washington, D.C.
McDaniel, M. A., Anderson, J. L., Derbish, M. H., & Morrisette, N. (2007). Testing the testing effect in the classroom. *European Journal of Cognitive Psychology, 19*, 494–513.
McDaniel, M. A., & Fisher, R. P. (1991). Test and test feedback as learning sources. *Contemporary Educational Psychology, 16*, 192–201.
McDaniel, M. A., Roediger, H. L., III, & McDermott, K. B. (2007). Generalizing test - enhanced learning from the laboratory to the classroom. *Psychonomic Bulletin & Review, 14*, 200–206.
Metcalfe, J., Butterfield, B., Habeck, C., & Stern, Y. (2011). Neural correlates of people's hypercorrection of their false beliefs. *Journal of Cognitive Neuroscience*, submitted for publication.
Metcalfe, J., & Finn, B. (2011). People's correction of high confidence errors: did they know it all along? *Journal of Experimental Psychology: Learning, Memory and Cognition, 37*, 437–448.
Murdock, B. B., Jr. (1974). *Human memory: Theory and data*. Potomac, MD: Erlbaum.
Nelson, T. O., & Narens, L. (1980). Norms of 300 general-information questions: accuracy of recall, latency of recall, and feeling-of-knowing ratings. *Journal of Verbal Learning & Verbal Behavior, 19*, 338–368.
Polich, J., Ladish, C., & Burns, T. (1990). Normal variation of P300 in children: age, memory span, and head size. *International Journal of Psychophysiology, 9*, 237–248.
Regard, M., Strauss, E., & Knapp, P. (1982). Children's production on verbal and non-verbal fluency tasks. *Perceptual and Motor Skills, 55*, 839–844, Bottom of Form.
Roediger, H. L., & Karpicke, J. D. (2006a). Test-enhanced learning: taking memory tests improves long-term retention. *Psychological Science, 17*, 249–255.
Roediger, H. L., & Karpicke, J. D. (2006b). The power of testing memory: basic research and implications for educational practice. *Perspectives on Psychological Science, 1*, 181–210.
Sanna, L. J., & Schwartz, N. (2006). Metacognitive experiences and human judgments: the case of hindsight bias and its debiasing. *Current Directions in Psychological Science, 15*, 172–176.
Sitzman, D. M., & Rhodes, M. G. (2010, Nov). Does the hypercorrection effect occur when feedback is delayed? *Paper presented at the 50th annual meeting of the Psychonomic Society*, St. Louis, MO.
Tulving, E., & Thomson, D. M. (1971). Retrieval processes in recognition memory: effects of associative context. *Journal of Experimental Psychology, 87*, 116–124.
Welsh, M. C., Pennington, B. F., & Groisser, D. B. (1991). A normative-developmental study of executive function: a window on prefrontal function in children. *Developmental Neuropsychology, 7*, 131–149.
Wood, G. (1978). The knew-it-all-along effect. *Journal of Experimental Psychology: Human Perception and Performance, 4*, 345–353.