Are There Hidden Costs to Teaching Mathematics with Incorrect Examples?

Min Kyung Hong (Min.Kyung.Hong@Vanderbilt.edu)
Department of Psychology and Human Development, Vanderbilt University, Peabody College #552, 230 Appleton Place
Nashville, TN 37203 USA

Darren J. Yeo (Darren.J.Yeo@Vanderbilt.edu)
Department of Psychology and Human Development, Vanderbilt University, Peabody College #552, 230 Appleton Place
Nashville, TN 37203 USA

Bethany Rittle-Johnson (Bethany.Rittle-Johnson@Vanderbilt.edu)
Department of Psychology and Human Development, Vanderbilt University, Peabody College #552, 230 Appleton Place
Nashville, TN 37203 USA

Lisa K. Fazio (Lisa.Fazio@Vanderbilt.edu)
Department of Psychology and Human Development, Vanderbilt University, Peabody College #552, 230 Appleton Place
Nashville, TN 37203 USA

Abstract

This study aims to address potential costs of using incorrect worked examples in teaching mathematics. While such practice has been shown to be effective in educational research, previous findings in the memory literature suggest that exposure to an incorrect solution may lead students to later believe that it is correct due to increased familiarity. We designed a two-session experiment with 1-week delay in which students studied correct and incorrect worked out examples. We found only small changes in students' ability to successfully distinguish between correct and incorrect solutions over time. Students did rate the previously studied incorrect examples as being more correct after the 1-wk delay, but this did not affect their correctness ratings of new correct and incorrect worked examples or their problem solving accuracy. We conclude that the unique nature of mathematical problem solving may protect students from the dangers of using incorrect worked examples.

Keywords: incorrect examples; worked examples; problem solving; mathematics learning; illusory truth; source memory

Introduction

A common challenge in schools is to help students learn novel problem-solving techniques. For example, in mathematics students need to learn how to solve quadratic equations and calculate probabilities. Numerous studies have found that students learn more when they alternate between studying worked out examples of the problem and solution and solving the problems themselves, as compared to simply solving double the number of problems (Sweller & Cooper, 1985; Ward & Sweller, 1990; Kirshner, Sweller, & Clark, 2006; Renkl & Atkinson, 2010).

According to *Cognitive Load Theory*, studying a worked-out example is less burdensome for the learners' working memory than solving a problem, which leaves more room for deeper cognitive processing such as understanding and learning the steps for a solution (Sweller, 1999; Sweller et al., 2011; Paas, Renkl, & Sweller, 2003; Zhu & Simon, 1987) . While most studies have focused on studying correct worked examples, recent research suggests that studying incorrect examples can optimize learning (Große & Renkl, 2007; Tsovaltzi et al., 2010; Durkin & Rittle-Johnson, 2012; Adams et al., 2014). Presenting students with common incorrect ways to solve a problem can help students to confront and eliminate those common errors.

However, teachers are often resistant to use incorrect worked examples in the classroom, as they believe that repeated exposure to incorrect procedures might cause students to confuse incorrect and correct solutions. Research from the memory literature suggests that these teachers' fears may be correct. People rate repeated statements as being more true than those they have not encountered before (i.e. illusory-truth effect, Dechêne, Stahl, Hansen & Wänke 2009; Hasher, Goldstein, & Toppino, 1977), even when they have relevant prior knowledge to suggest that the statements are true or false (Fazio, Brashier, Payne & Marsh, 2015). Repetition increases the familiarity and processing fluency of a statement, which is thought to lead to increased perceived credibility of the information (Unkelbach, 2007; Begg, Anas, & Farinacci, 1992). The illusory truth effect has been shown to occur for trivia statements (e.g., Bacon, 1979), information about consumer products (e.g., Johar & Roggeveen 2007), and political opinions (e.g., Arkes, Hackett & Boehm, 1989), and it occurs both in the laboratory and in naturalistic settings (Gigerenzer, 1984; Boehm, 1994).

In addition, research has shown that people often remember factual information while forgetting the source of that information (Barber, Rajaram, & Marsh, 2008; Conway et al 1997; Dewhurst et al. 2009). This forgetting may be a natural consequence of information moving from episodic to semantic memory. For example, you likely know that George Washington was the first president of the United States, but no longer remember where you first learned that information. In related phenomenon such as the sleeper effect (Kumkale & Albarracin, 2004), false fame (Topolinski & Strack, 2010), and unconscious plagiarism (Bink, Marsh, Hicks & Howard, 1999), participants remember previously presented information while forgetting its source or origin.

Taken together, these findings suggest that studying with incorrect worked examples may not yield benefits if students fail to accurately monitor the source of the information.

The present study was designed to examine if this forgetting of the source information occurs with correct and incorrect worked examples. That is, after a delay, would students continue to remember which worked examples were correct and which were incorrect or would they begin to confuse the two procedures? During the study phase, students saw both correct and incorrect worked examples and were asked to rate the correctness of the shown solution. They then received feedback about which examples were correct or incorrect and rated the examples again. Finally, they were asked to solve novel problems of the same type shown in the examples. One week later, the students returned and again solved problems and rated the correctness of the worked examples. Of interest was how students' ratings and problem solving accuracy would change over time.

If teachers' concerns are correct, then we would expect students to be less accurate at rating the correctness of the worked examples after the delay. In addition, we would expect students' problem solving accuracy to also decrease. If, however, students are able to remember which examples are correct and which are incorrect, then we would expect no changes in students' correctness ratings or problem solving accuracy over time.

Method

Participants

Thirty-four Vanderbilt University undergraduates participated in exchange for course credit (9 male; mean age 18.8 years). Three participants were excluded from the analysis because they failed to attend the second session of the experiment, leaving 31 participants in the analysis. Participants were tested individually or in small groups of up to four people.

Design

The experiment implemented a 3 (time: study, immediate, delay) × 2 (solution: correct, incorrect) within-subjects design (Figure 1).

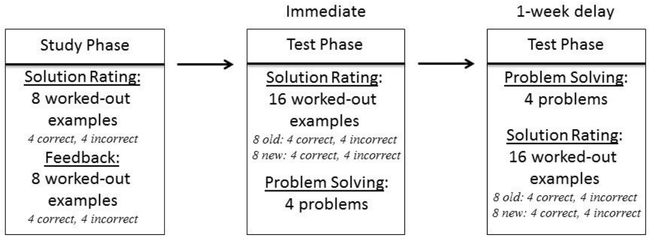


Figure 1. Study design.

Materials

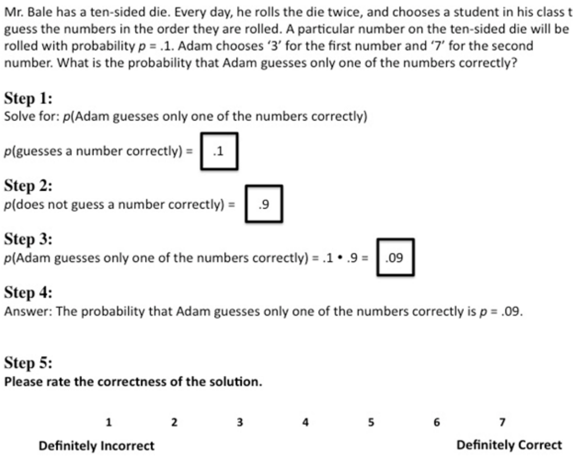
We adapted 10 pretest problems, eight worked-out examples, and four post-test probability problems from Experiment 2 of Große & Renkl (2007) with minor word changes (e.g. "skat cards" to "poker cards"). The problems all dealt with probabilities concerning the intersection between two different events or event order. Event intersection problems differed on whether two different events were mutually exclusive or if they could occur together; event order problems differed on whether the sequence of events were relevant or irrelevant. We then created an additional set of eight worked-out examples that were structurally similar to the eight worked-out examples from Große & Renkl (2007). Therefore, participants were shown a total of 16 worked-out examples, half of which were presented with correct solutions and half with incorrect solutions. The incorrect worked examples contained accurate calculations, but the procedure used was not appropriate for the problem (e.g. failing to subtract the intersection between event A and event B when calculating the probability of event A or event B happening). Stimuli for the pretest, study phase, filler task, and solution rating task were presented on a computer using MediaLab and DirectRT software (Empirisoft Corporation, New York, NY). The problem solving task was administered through a paper-and-pencil test.

Procedure

Participants gave informed consent and solved 10 basic probability problems (pretest), which were used to assess their prior domain knowledge.

Participants then studied eight worked-out examples, four correct and four incorrect. Participants were told that the solutions were provided by other students and they would be asked to rate the accuracy of the student's solution. Each worked example was presented one at a time, and participants keyed in the intermediate answers when they were prompted with fill-in-the-blank boxes (shown in Figure 2). This ensured that participants paid attention to each of the steps in the solution and increased their engagement and depth of processing. Each step was revealed progressively, regardless of whether the participant's input for each blank was correct or incorrect. Thus, if a participant answered incorrectly, their answer would disappear and the correct answer would be presented along with the next step. At the end of each worked example, the full solution was presented and participants rated the correctness of the solution using a 7-point Likert scale with 1 labeled as "Definitely incorrect" and 7 indicating "Definitely correct".

Participants then solved visuo-spatial puzzles as a filler task for two minutes. After the filler task, participants were presented with the same set of fully completed worked-out examples, in the same order that they had seen previously. For each problem, they were presented with feedback about whether the worked example was correct or incorrect and were given at least 20 seconds to review the solution and the feedback. After 20 seconds, their earlier correctness rating appeared at the bottom of the solution (e.g. "You rated the correctness of the solution of this worked-out example: 2"), and they were prompted to move onto the next worked example when they were ready. Participants then completed another set of visuo-spatial puzzles for two minutes before moving onto the test phase.



Mr. Bale has a ten-sided die. Every day, he rolls the die twice, and chooses a student in his class to guess the numbers in the order they are rolled. A particular number on the ten-sided die will be rolled with probability $p = .1$. Adam chooses '3' for the first number and '7' for the second number. What is the probability that Adam guesses only one of the numbers correctly?

Step 1:
Solve for: *p*(Adam guesses only one of the numbers correctly)

p(guesses a number correctly) = .1

Step 2:
p(does not guess a number correctly) = .9

Step 3:
p(Adam guesses only one of the numbers correctly) = .1 • .9 = .09

Step 4:
Answer: The probability that Adam guesses only one of the numbers correctly is $p = .09$.

Step 5:
Please rate the correctness of the solution.

1	2	3	4	5	6	7
Definitely Incorrect						Definitely Correct

Figure 2. Illustration of a worked-out example. Each step was revealed progressively, and participants were prompted to type their answers in each fill-in-the-blank box, represented by the black boxes above.

During the test phase, participants again rated the correctness of the worked examples and they solved novel probability problems. Task order within the test phase was counterbalanced across participants. During the solution rating task, participants were given 16 worked-out examples, eight of which were *old* problem*s* shown during the study phase, and eight of which were *new* problems that were isomorphic to the *old* problems but with different cover stories. Participants rated each solution using the 7-point response scale. During the problem solving task, participants were given four post-test probability problems. Participants' performance on the probability problems was used to examine their ability to apply any gained knowledge from the study phase to novel problems. The order of the problems in the booklet was counterbalanced across participants.

At the end of the test phase, participants were asked if they had learned how to solve these specific types of probability problems prior the experiment (if so, what did they learn and where) and to list the math and statistics classes that they had taken in high school (advanced placement courses only) and college.

A week after the first session, participants came back to the lab and completed the solution rating and problem solving tasks again. The order of the tasks was reversed from the first session. At the end of the second session, participants were asked to give demographic information about their age, gender, and ethnicity.

Results

Solution ratings for old items across time

Recall that students rated the examples on a scale of 1 to 7. Ratings above 4 indicated believing that the solution was more correct than incorrect, and ratings below 4 indicated believing that the solution was more incorrect than correct. As shown in Figure 3, during the study phase, prior to feedback, students rated both correct and incorrect solutions as being more correct than incorrect, although their ratings for correct solutions were higher than for incorrect solutions. Thus, participants were having some difficulty identifying incorrect solutions as incorrect. On the immediate test, as expected, feedback allowed participants to more clearly distinguish between correct and incorrect solutions, and they rated the incorrect solutions as incorrect. Of primary interest were ratings on the delayed test. Did participants forget which solutions were incorrect, in line with the illusory truth effect and source forgetting? The answer seems to be somewhat; students rated incorrect solutions more highly on the delayed posttest than the immediate posttest, although not as highly as before feedback (during the study phase).

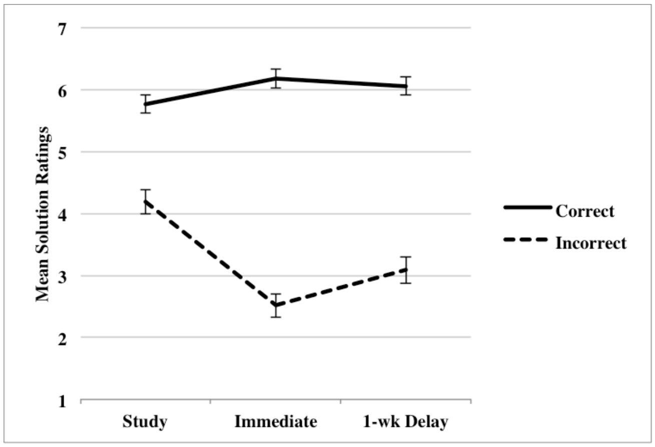


Figure 3. Mean solution ratings for correct and incorrect examples as a function of time. Error bars reflect the standard error of the mean.

To examine these patterns statistically, we conducted a 3 (time: study, immediate, 1-wk delay) × 2 (solution: correct, incorrect) repeated measures analysis of variance (ANOVA) on participants' solution ratings. This analysis was restricted to the eight problems that participants saw during all three phases of the experiment: prior to being told which solutions were correct and incorrect, immediately after the feedback, and one week later.

As shown in Figure 3, participants rated the correct solutions (M = 6.00) as being more correct than the incorrect solutions (M = 3.27), $F(1, 30) = 87.59$, MSE = 3.97, $p < .001$, $\eta_p^2 = 0.75$. This was true during the study phase before the students received feedback, $t(30) = 5.38$, $p < .001$, $d = 3.19$, and on the immediate, $t(30) = 10.39$, $p <$

.001, *d* = 4.90, and delayed tests, *t*(30) = 7.57, *p* < .001, *d* = 4.64. In addition, participants' ratings changed over time, *F*(2, 60) = 6.46, *MSE* = 0.99, *p* = .003, η_p^2 = 0.18, but this was qualified by a significant interaction between time and solution type, *F*(2, 60) = 21.05, *MSE* = 0.83, *p* < .001, η_p^2 = 0.41. While the ratings for correct solutions remained relatively constant over time, the ratings for incorrect examples changed dramatically. Follow-up *t*-tests showed no significant differences across the time points for correct solutions (study vs. immediate, *t*(30) = 1.70, *p* = .10; study vs. delay, *t*(30) < 1; immediate vs. delay, *t*(30) = 1.31, *p* = .20). However, there was a large decrease in the ratings for incorrect solutions after the participants received feedback (study *M* = 4.19; immediate *M* = 2.52; *t*(30) = 5.99, *p* < .001, *d* = 2.58). This change remained one-week later (study *M* = 4.19; delay *M* = 3.09; *t*(30) = 3.98, *p* < .001, *d* = 2.20), but the incorrect examples were rated as being slightly more correct as compared to immediately following the feedback (*t*(30) = 2.99, *p* = .005, *d* = 0.38). Overall, participants were able to successfully distinguish correct items from incorrect items, even after the delay. This pattern of results did not change when we included participants' pretest scores as a covariate.

Solution ratings for old versus new items

Given that students were able to successfully distinguish between correct and incorrect items, we were interested in whether they were able to transfer that knowledge to the new solutions. To more easily compare participants' knowledge over time, we conducted the analysis on the difference between participants' ratings for correct and incorrect items. Higher difference scores represent greater ability to distinguish correct solutions from incorrect solutions. The results are shown in Figure 4.

As noted above, for the old items, participants' ability to distinguish between correct and incorrect solutions was high immediately following the feedback, but this ability decreased after the one week delay. However, we did not see a similar pattern with the new examples. For the novel examples, the difference in rating of new examples was similar to the difference during the study phase for the old examples (mean differences = 1.96 vs. 1.58) and was lower than for the old examples on the immediate test. The difference in ratings for the new examples did not change from the immediate test to the 1-week delay. These findings suggest that participants were remembering particular examples of solution methods tied to individual problems and not a more general solution method that they used to rate the accuracy of new examples.

To confirm these observations, we conducted a 2 (time: immediate, 1-wk delay) × 2 (novelty: old, new) ANOVA on the difference scores. As shown in Figure 4, participants were better able to discriminate correct and incorrect solutions for old items (*M* = 3.31) than new items (*M* = 2.14), *F*(1, 30) = 21.49, *p* < .001, *SEM* = 2.00, η_p^2 = 0.42. There was no main effect of time, *F* < 1, but there was a significant interaction between time and novelty, *F*(1,30) = 9.85, *p* = .004, *SEM* = 0.87, η_p^2 = 0.25. The difference in knowledge between the old and new items was larger on the immediate test than one-week later. Difference scores for the old items decreased over time (immediate *M* = 3.66; delay *M* = 2.97; *t*(30) = 2.37, *p* = .024, *d* = 1.84), while difference scores for new items did not change significantly (immediate *M* = 1.96; delay *M* = 2.32; *t*(30) = 1.54, *p* = .13, *d* = 0.15). These results suggest that participants did not generalize what they learned from the feedback to the new examples.

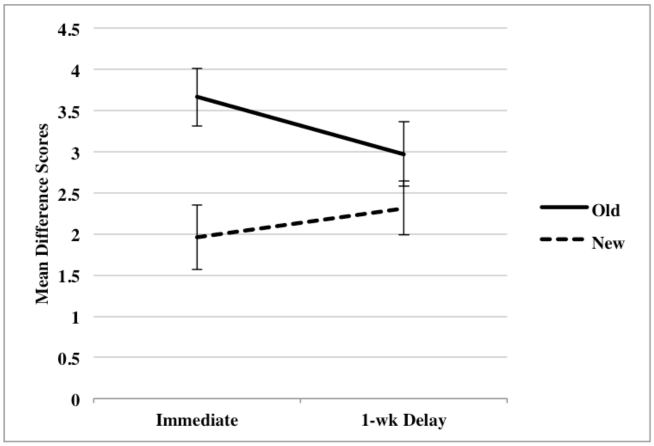


Figure 4. Solution rating differences across time for old and new problems. Error bars reflect the standard error of the mean.

Problem solving accuracy

Next, we examined how students' ability to solve the four booklet problems changed after a 1-week delay. There was no significant difference in the proportion of questions answered correctly on the immediate (*M* = .60) and delayed (*M* = .54) tests, *t*(30) = 1.16, *p* = .26, *d* = 0.21.

Problem solving strategies

We also examined the types of strategies used for each of four probability problems (shown in Figure 5) to see how often participants used the demonstrated strategies. Strategies were coded as *correct, incorrect,* or *other*. Incorrect strategies were the strategies presented in the incorrect worked-out examples, and other strategies were incorrect strategies that were not introduced through the worked-out examples. All the participants who used correct strategies solved the problems correctly.

Overall, students' strategy choice was consistent over time, and they used correct strategies more often than incorrect or other strategies. Students were equally likely to use correct strategies on the immediate and delayed tests, *t*(30) = 1.23, *p* = .23, *d* = 0.22. And there were also no changes in their use of incorrect, *t*(30) = -1.36, *p* = .18, *d* = -0.25, and other strategies, *t*(30) = -0.15, *p* = .88, *d* = -0.03.

Students did not seem to frequently adopt or revert to the incorrect strategies demonstrated in the incorrect examples.

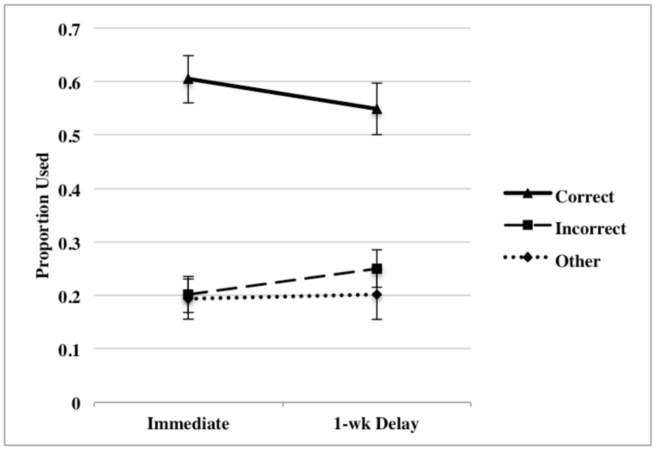


Figure 5. Proportion of correct, incorrect and other strategies used on the probability problems on the immediate and delayed test. Error bars reflect the standard error of the mean.

Discussion

The present study examined whether findings from the memory literature suggesting potential costs of learning through incorrect worked examples extended to mathematics problem solving. Overall, students showed little forgetting over time and were able to distinguish between the correct and incorrect worked examples, even after a one-week delay. For the repeated items, there was an increase in participants' ratings for the incorrect examples after a one-week delay, suggesting that students may have started to forget the "incorrect" tag. The fact that the studied incorrect examples were rated as more correct over time, while there was no corresponding decrease in the ratings for correct examples, suggests that the illusory truth effect may have played a role. Rather than simply forgetting which examples were correct or incorrect, participants may have believed all of the repeated examples to be more correct after the delay. The correct examples were already close to ceiling and thus could show no changes.

However, there was little evidence that participants generalized information from the studied examples. First, their ratings of new examples did not seem to be influenced by their improved knowledge of old examples on the immediate test or their decline in knowledge of old examples on the delayed test. Second, participants were not that likely to use the demonstrated incorrect strategies on the immediate test nor increase their use of it on the delayed test. So while the old incorrect examples gained a little bit of truth over the delay, this did not affect the students' ability to rate the new examples or to accurately solve the problems. Thus, our results provide some preliminary evidence for educators that their worries about the negative effects of presenting incorrect worked examples may be unfounded.

Worked examples differ from the stimuli used in previous memory studies in many ways. One key difference is that the correct and incorrect worked examples both support the learning of the same correct procedure. Thus, one does not need to remember both that this specific example was correct and that this specific example was incorrect. Instead, by learning the procedure, one can identify both which solutions are correct and which solutions are incorrect. In contrast, within the illusory truth paradigm, knowing that "Oslo is the capital of Finland" is false tells you nothing about the truth of "Marconi is the inventor of the wireless radio". If incorrect worked examples help students to gain an accurate representation of the correct solution, then it may not matter if students forget the specific label (correct or incorrect) that went with a given worked example.

One limitation of the current study was that the students did not fully learn the correct procedures. The participants correctly answered only a little more than half of the problems on both the immediate and delayed posttests. Our procedure was designed to match those used in the illusory truth and sleeper effect literatures and thus does not match how educators and educational researchers typically use incorrect worked examples. Future studies should examine students' memory after they experience a typical classroom lesson where the incorrect examples are directly compared to the correct examples. Such direct comparisons may increase learning of the correct procedures. It is also possible that students may have shown greater forgetting after a longer delay. Brown and Nix (1996) found that participants in an illusory truth experiment could remember which statements were labeled as true or false one week later, but after a month all repeated statements were rated as being more true, regardless of their initial labeling.

The current study represents a first step towards bringing together research on memory processes and research on learning to examine the long-term effects of different instructional techniques. We found small effects of forgetting on the rating scales typically used in the relevant memory research, but no negative effects on the students' actual problem solving skills. It remains to be determined whether the small amount of forgetting shown in this experiment would cause larger issues at a longer delay or if students can correctly learn the procedure even if they forget the initial correct and incorrect labels.

Acknowledgments

We would like to thank Meredith Sight and Octavia White for their help regarding this project.

References

Adams, D. M., McLaren, B. M., Durkin, K., Mayer, R. E., Rittle-Johnson, B., Isotani, S., & van Velsen, M. (2014). Using erroneous examples to improve mathematics learning with a web-based tutoring system. *Computers in Human Behavior*, *36*, 401–411.

Arkes, H. R., Hackett, C., & Boehm, L. (1989). The generality of the relation between familiarity and judged validity. *Journal of Behavioral Decision Making*, *2*(2), 81–94.

Bacon, F. T. (1979). Credibility of repeated statements: Memory for trivia. *Journal of Experimental Psychology: Human Learning and Memory*, 5, 241-252.

Barber, S. J., Rajaram, S., & Marsh, E. J. (2008). Fact learning: How information accuracy, delay, and repeated testing change retention and retrieval experience. *Memory*, *16*(8), 934–946.

Begg, I. M., Anas, A., & Farinacci, S. (1992). Dissociation of processes in belief: Source recollection, statement familiarity, and the illusion of truth. *Journal of Experimental Psychology: General*, *121*(4), 446–458.

Bink, M. L., Marsh, R. L., Hicks, J. L., & Howard, J. D. (1999). The credibility of a source influences the rate of unconscious plagiarism. *Memory*, *7*(3), 293–308.

Boehm, L. E. (1994). The validity effect: A search for mediating variables. *Personality and Social Psychology Bulletin*, *20*(3), 285–293.

Brown, A. S., & Nix, L. A. (1996). Turning lies into truths: Referential validation of falsehoods. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(5), 1088–1100.

Conway, M. A., Gardiner, J. M., Perfect, T. J., Anderson, S. J., & Cohen, G. M. (1997). Changes in memory awareness during learning: The acquisition of knowledge by psychology undergraduates. *Journal of Experimental Psychology: General*, *126*(4), 393–413.

Dechêne, A., Stahl, C., Hansen, J., & Wänke, M. (2009). Mix me a list: Context moderates the truth effect and the mere-exposure effect. *Journal of Experimental Social Psychology*, *45*(5), 1117–1122.

Dewhurst, S. A., Conway, M. A., & Brandt, K. R. (2009). Tracking the R-to-K shift: Changes in memory awareness across repeated tests. *Applied Cognitive Psychology*, *23*(6), 849–858.

Durkin, K., & Rittle-Johnson, B. (2012). The effectiveness of using incorrect examples to support learning about decimal magnitude. *Learning and Instruction*, *22*(3), 206–214.

Fazio, L. K., Brashier, N. M., Payne, B. K., & Marsh, E. J. (2015). Knowledge does not protect against illusory truth. *Journal of Experimental Psychology: General*, *144*(5), 993–1002.

Gigerenzer, G. (1984). External Validity of Laboratory Experiments: The Frequency-Validity Relationship. *The American Journal of Psychology*, *97*(2), 185–195.

Große, C. S., & Renkl, A. (2007). Finding and fixing errors in worked examples: Can this foster learning outcomes? *Learning and Instruction*, *17*(6), 612–634.

Hansen, J., & Wänke, M. (2009). Liking What's Familiar: The Importance of Unconscious Familiarity in the Mere-Exposure Effect. *Social Cognition*, *27*(2), 161–182.

Hasher, L., Goldstein, D., & Toppino, T. (1977). Frequency and the conference of referential validity. *Journal of Verbal Learning & Verbal Behavior*, *16*(1), 107–112.

Jarvis, B. G. (2014). MediaLab (Version 2014.1.108) [Computer Software]. New York, NY: Empirisoft Corporation.

Jarvis, B. G. (2014). DirectRT (Version 2014.1.104) [Computer Software]. New York, NY: Empirisoft Corporation.

Johar, G. V., & Roggeveen, A. L. (2007). Changing false beliefs from repeated advertising: The role of claim-refutation alignment. *Journal of Consumer Psychology*, *17*(2), 118–127.

Kirschner, P.A., Sweller, J., and Clark, R.E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41*, 75–86.

Kumkale, G. T., & Albarracín, D. (2004). The Sleeper Effect in Persuasion: A Meta-Analytic Review. *Psychological Bulletin*, *130*(1), 143–172.

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive Load Theory and Instructional Design: Recent Developments. *Educational Psychologist*, *38*(1), 1–4.

Renkl, A., & Atkinson, R. K. (2010). Learning from worked-out examples and problem solving. (pp. 91–108). Cambridge University Press (New York, NY, US).

Sweller, J., & Cooper, G. A. (1985). The Use of Worked Examples as a Substitute for Problem Solving in Learning Algebra. *Cognition and Instruction*, *2*(1), 59.

Sweller, J. (1999). *Instructional Design in Technical Areas. Australian Education Review, No. 43*. PCS Data Processing, Inc., 360 W. 31st, New York, NY 10001

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive Load Theory*. New York, NY: Springer New York

Topolinski, S., & Strack, F. (2010). False fame prevented: Avoiding fluency effects without judgmental correction. *Journal of Personality and Social Psychology*, *98*(5), 721–733.

Tsovaltzi, D., Melis, E., McLaren, B. M., Meyer, A.-K., Dietrich, M., & Goguadze, G. (2010). Learning from Erroneous Examples: When and How Do Students Benefit from Them? In M. Wolpers, P. A. Kirschner, M. Scheffel, S. Lindstaedt, & V. Dimitrova (Eds.), *Sustaining TEL: From Innovation to Learning and Practice* (pp. 357–373).

Unkelbach, C. (2007). Reversing the truth effect: Learning the interpretation of processing fluency in judgments of truth. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(1), 219–230.

Ward, M., & Sweller, J. (1990). Structuring effective worked examples. *Cognition and Instruction*, *7*(1), 1–39.

Zhu, X., & Simon, H. A. (1987). Learning mathematics from examples and by doing. *Cognition and Instruction*, 4, 137-166.