

Learning errors from fiction: Difficulties in reducing reliance on fictional stories

ELIZABETH J. MARSH and LISA K. FAZIO
Duke University, Durham, North Carolina

Readers rely on fiction as a source of information, even when fiction contradicts relatively well-known facts about the world (Marsh, Meade, & Roediger, 2003). Of interest was whether readers could monitor fiction for errors, in order to reduce suggestibility. In Experiment 1, warnings about errors in fiction did not reduce students' reliance on stories. In Experiment 2, all subjects were warned before reading stories written at 6th- or 12th-grade reading levels. Even though 6th-grade stories freed resources for monitoring, suggestibility was not reduced. In Experiment 3, suggestibility was reduced but not eliminated when subjects pressed a key each time they detected an error during story reading. Readers do not appear to spontaneously monitor fiction for its veracity, but can do so if reminded on a trial-by-trial basis.

Fictional narratives often refer to people, places, and concepts that exist (or have existed) in the real world. The movie *The Patriot* (Emmerich, 2000) is set during the American Revolution, the actor John Malkovich plays himself in the surreal film *Being John Malkovich* (Jonze, 1999), and *Tyrannosaurus rex* rules in *Jurassic Park* (Spielberg, 1993). In the classroom, fiction can be a powerful tool to engage student interest (see, e.g., Dubeck, Moshier, & Boss, 1988). However, by definition, fiction is not always true to life. Many British objected to the portrayal of the war in *The Patriot*, in real life there is no portal to John Malkovich's brain, and it is impossible to clone dinosaurs. Fiction can be a source of false beliefs about the world.

Research supports the intuition that readers associate fiction's errors with their related world knowledge. Studying false facts slows (interferes with) later retrieval of related knowledge (Lewis & Anderson, 1976; Peterson & Potts, 1985). In some circumstances, prior reading of false facts shifts beliefs as reported on Likert scales (Prentice, Gerrig, & Bailis, 1997; Strange & Leung, 1999; Wheeler, Green, & Brock, 1999). Story reading increases production of false facts on a test of general knowledge, even when guessing is discouraged (Marsh, Meade, & Roediger, 2003). That is, subjects who read stories containing factual errors such as *Wilmington is the capital of Delaware* (see the Appendix; Marsh, 2004) later answered the

factual question *What is the capital of Delaware?* with *Wilmington* at higher rates than did subjects who had not read the story. Perhaps even more surprising is that prior reading of misinformation reduced subjects' ability to correctly answer questions below baseline; having read the misinformation about *Wilmington* decreased correct responses (*Dover*) to a level below those of subjects who never read the story.

Given the importance of source memory problems in other memory errors, one might expect suggestibility to occur only after subjects have forgotten the fictional source of their knowledge (similarly to how low-credibility sources exert their influence on attitudes after a delay; Hovland, Lumsdaine, & Sheffield, 1949). Failure to monitor source is an important component of memory errors such as illusory truth (see, e.g., Begg, Anas, & Farinacci, 1992) and eyewitness susceptibility to postevent misinformation (Lindsay & Johnson, 1989). Similarly, in the domain of narrative persuasiveness, many readers could not recall whether passages had been labeled as fact (a newspaper article) or fiction (from a fiction magazine) (Green & Brock, 2000). However, source amnesia is not a necessary component of using false facts from fiction. That is, subjects rely on fictional stories even when they remember having read the facts in the stories (Marsh et al., 2003).

If not because of amnesia for the story source, why do readers rely on fictional sources? It is a puzzling result, since readers are at least somewhat aware of the problems with fiction. For example, a survey of Yale undergraduates revealed that many agreed that authors of fiction sometimes invent facts for storyline purposes (Prentice & Gerrig, 1999), and yet in other studies, story reading affected the beliefs of subjects from the same population. One way to gain traction on the mechanism underlying suggestibility is to manipulate well-understood variables, to see whether they affect the error level in this memory paradigm. We began with manipulations of warning, because

This work was supported by a collaborative activity award from the James S. McDonnell foundation. We thank Emily Lowe, Holli Sink, Elaina Pelky, Scott Weiskopf, Rebecca Wedel, Nancy Yang, and Angela Zhou for help with data collection, coding, and input. Keith Payne and Mike Hutson provided useful comments on an earlier draft of the manuscript. Correspondence concerning this article should be addressed to E. J. Marsh, Psychology and Neuroscience, Duke University, 9 Flowers Drive, Box 90086, Durham, NC 27708-0086 (e-mail: emarsh@psych.duke.edu).

effects of warning are quite consistent in other false memory paradigms. Across paradigms, a warning delivered before study is most effective at reducing later suggestibility. Preevent warnings have reduced suggestibility in the eyewitness misinformation paradigm (Greene, Flynn, & Loftus, 1982) and in the Deese/Roediger–McDermott (DRM) paradigm for creating false memories for words (see, e.g., Gallo, Roediger, & McDermott, 2001). In part, prestudy warnings work because they make subjects encode material more carefully. Warnings given after study are less likely to reduce false memories, but effects have been observed in the eyewitness misinformation paradigm (see, e.g., Chambers & Zaragoza, 2001) and in the social contagion paradigm (Meade & Roediger, 2002). In these cases, warnings may help by reminding subjects to think more carefully about the sources of their memories. Thus, if effects of warning (given before or after study) are observed in the fiction paradigm, it would suggest underlying mechanisms similar to those hypothesized in other paradigms.

However, the conditions facing the story reader may make it hard for him or her to heed the warnings and avoid fiction's errors. Consider an idea dating back to Coleridge (1817/1906), who argued that reading poetry necessitates a "willing suspension of disbelief" (p. 161). The idea was that a reader shifted processing modes upon encountering fiction, to allow acceptance of implausible events and ideas for the sake of enjoying a text. Although the idea of a "special mode" for fiction processing is less popular with modern researchers (e.g., Gerrig, 1993), a number of data points do support that one's general knowledge can become less accessible during the processing of fiction. That is, even if readers do not *willingly* suspend disbelief, involvement in a fictional narrative may interfere with retrieval from semantic memory. For example, readers were slower to verify the well-known fact that George Washington was the first president of the United States after reading a narrative that suggested that Washington did not want to be president (Gerrig, 1989).

When a reader encounters a factual error in the text, another problem arises. According to Gilbert (1991), disbelief requires cognitive effort, regardless of the type of material being read. Given that readers of fiction (1) may have willingly given up critical processing (as Coleridge proposed), and/or (2) may be less able to access related world knowledge needed to evaluate text (as demonstrated by Gerrig, 1989), it is a real question whether readers will be able to disbelieve a story's errors. The problem is compounded by the fact that readers need to devote many of their cognitive resources to following the storyline. That is, readers are busy applying schemas toward understanding story characters and building and updating mental models of the situation (see, e.g., Bower & Morrow, 1990; Johnson-Laird, 1983), with the consequence that fewer resources are available for evaluating and disbelieving background information. Accordingly, the more that readers are involved in a story's plot, the less able they are to rate parts of the narrative as "ringing false" to them (Green & Brock, 2000).

To examine these issues, in Experiment 1, we made it explicit that the fictional stories might contain errors. That is, it was not up to subjects to discover this upon encountering an erroneous fact. Given that subjects no longer had to discover the need to monitor, would they be able to do so? We compared performance in a no-warning control condition with that in two experimental conditions: a condition in which students received warnings before reading (which would allow them both to detect errors during reading and to later reduce reliance on the story) and a condition in which they received warnings after reading (which would allow them to reduce reliance on the story at test). The use of one or both of these monitoring strategies should reduce effects of story reading on the final general knowledge test.

EXPERIMENT 1

Method

Subjects. Forty-eight Washington University undergraduates participated in the experiment for partial fulfillment of a course requirement. The subjects were tested in small groups of 1 to 5 people.

Design. The experiment had a 2 (question ease: easy or hard) \times 3 (fact framing: not read baseline, correct, or misleading) \times 3 (warning condition: none, before, or after) design. Question ease and fact framing were manipulated within subjects; warning condition was a between-subjects factor. Dependent variables were the proportion of cued recall questions answered correctly versus with misinformation.

Materials. Eight fictional short stories were taken from Marsh (2004). Each story was two to three single-spaced pages long and contained characters, dialogue, and plot. These stories were divided into two sets of four stories; each subject read only one of the sets. Story reading was counterbalanced across subjects.

Each story contained eight references to facts from the Nelson and Narens (1980) norms. Half of the facts corresponded to easy questions; on average, 70% of Nelson and Narens's subjects had answered these questions correctly. The remaining fact corresponded to hard questions; on average, 15% of Nelson and Narens's subjects had answered these questions correctly.

Within a story, half of the facts were presented in a correct frame, and half in a misleading frame. Both frames referred to the later general knowledge question, but the correct frame provided the correct answer, whereas the misleading frame suggested an incorrect (but plausible) answer. For example, consider the question "What is the name of the navigation instrument used at sea to plot position by the stars?" When the item was in a correct frame, the dialogue read "This here, this is a sextant and it's the main tool used at sea to navigate via the stars." When the item was in the misleading condition, the dialogue read "This here, this is a compass and it's the main tool used at sea to navigate via the stars." Fact framing was counterbalanced across subjects.

Four different story books were created to counterbalance story set and fact framing. Four different reading comprehension books were constructed to match the four story books. Each comprehension book contained four pages (one corresponding to each of the stories) with four questions on each page. The first question was always "Did you finish reading the story?" The remaining three questions were about plot details.

The general knowledge test contained 96 questions. All questions were in cued recall format (e.g., "What is the name of the navigation instrument used at sea to plot position by the stars?"), and the subjects wrote each answer in a space provided next to the question. The 96 questions were tested in a single random order; 64 were critical questions, and the remaining 32 were filler questions.

Finally, debriefing sheets presented the corrected versions of facts that had been presented in the misleading condition.

Procedure. The experiment consisted of a series of reasoning tasks: reading comprehension, a visuospatial task (a filler), and the test of general world knowledge.

The subjects in the warning-before condition received a warning prior to the beginning of the reading phase. They were told that the stories would be fictional, and that "Authors of fiction often take liberties with certain facts or ideas in order to make the story flow better or be more entertaining. Therefore, some of the information you will read may be incorrect." Only the subjects in the warning-before condition received this warning printed on a piece of paper, which the experimenter read aloud.

In Phase 1, the story-reading phase, all of the subjects were told to read carefully and knew they would be asked questions about the stories. As established in previous experiments, the subjects were given up to 5 min to read each story; the experimenter warned the subjects at the 3.5-min mark. The subjects were instructed to read each story once, and to wait quietly if they finished reading early. The subjects completed four story-reading/comprehension-check cycles; as directed by the experimenter, they answered the comprehension questions between stories.

In Phase 2, the filler phase, the subjects solved visuospatial brainteasers for 7 min.

The subjects in the warning-after condition received a warning immediately prior to the general knowledge test. The warning was the same as in the warning-before condition, except that the last sentence read "Therefore, some of the information you read may have been incorrect." Only the subjects in the warning-after condition received this warning printed on a piece of paper, which the experimenter read aloud.

In Phase 3, the test phase, the subjects answered general knowledge questions. All of the subjects were warned not to guess and were instructed to draw a line in the answer space if they could not answer the question. They were given up to 11 min in which to complete this task; at three points in time, the experimenter instructed the subjects with regard to how many questions they should have answered thus far.

In Phase 4, the subjects were fully debriefed about the experiment's purpose. As part of this debriefing, they rated how surprising they found the corrected versions of facts that they had previously read in misleading frames.

Results

All results were significant at the .05 alpha level unless otherwise noted.

Correct answers. A 2 (question ease: easy or hard) \times 3 (fact framing: correct, misleading, or not read baseline) \times 3 (warning placement: none, before, or after story

reading) ANOVA was computed on proportion of critical questions answered correctly. The data are shown in the top panel of Table 1.

Replicating prior work, the subjects correctly answered more easy questions than hard [$F(1,45) = 624.02$, $MS_e = 0.02$]. Critically, there was a main effect of fact framing [$F(2,90) = 50.48$, $MS_e = 0.03$], and this interacted with question ease [$F(2,90) = 8.68$, $MS_e = 0.02$]. For easy questions, the subjects answered more questions correctly when they had read correct answers than when they had not read the relevant stories [$t(47) = 4.75$, $SEM = .03$]. A similar benefit of story reading was observed for hard questions [$t(47) = 6.33$, $SEM = .03$]. A different pattern emerged for having read misinformation. Reading misinformation dropped performance below baseline for easy questions [$t(47) = 5.09$, $SEM = .03$] but not for hard questions (likely due to a floor effect).

The novel variable was whether a warning was given prior to or after story reading. Interestingly, the effect of fact framing was not dependent on warning condition. That is, in all conditions, performance was highest after subjects had read the correct answers in the stories. In all conditions, there was a cost to having read misinformation in the stories for easy but not for hard questions. Receiving a warning, whether before or after story reading, did not change the pattern of correct responses ($F < 1$).

In summary, the effect of fact framing on correct responses was not dependent on warning condition. Reading misinformation reduced correct answers below baseline, and this result did not depend on warning condition.

Misinformation production. We defined *misinformation production* as producing the specific incorrect answer that was presented in the stories. Other incorrect answers were not counted as misinformation. As is shown in the bottom panel of Table 1, story reading led to misinformation production in all three conditions. Warnings had a general effect, with misinformation production reduced across conditions. But warnings did not lead the subjects to selectively edit out wrong answers that had appeared in the stories.

To support these observations, a 2 (question ease: easy or hard) \times 3 (fact framing: correct, misleading, or not

Table 1
Production of Correct and Misinformation Answers as a Function of Question Ease, Fact Frame, and Warning Condition (Data from Experiment 1)

Warning Condition	Easy Questions			Hard Questions		
	Correct	Baseline	Misleading	Correct	Baseline	Misleading
	Proportion Correct					
Control	.71	.64	.50	.32	.16	.16
Warning before	.79	.61	.49	.41	.14	.14
Warning after	.83	.71	.48	.39	.17	.18
<i>M</i>	.78	.65	.49	.37	.16	.16
	Proportion Misinformation					
Control	.10	.06	.23	.14	.10	.35
Warning before	.03	.04	.21	.03	.07	.29
Warning after	.01	.04	.22	.06	.08	.34
<i>M</i>	.05	.05	.22	.08	.09	.33

read baseline) \times 3 (warning placement: none, before, or after story reading) ANOVA was computed on proportion misinformation answers. There was a main effect of warning condition [$F(2,45) = 3.19$, $MS_e = 0.02$]. The subjects in the no-warning condition produced more misinformation ($M = .14$) than did the subjects who received a warning before story reading ($M = .10$) [$t(30) = 2.49$, $SEM = .02$]. There was also a trend for the control subjects to produce more misinformation overall than did the subjects who received a warning after story reading ($M = .11$) [$t(30) = 1.74$, $SEM = .02$, $p < .1$].

Warnings reduced misinformation production; however, that effect was not dependent on subjects' having read misinformation in the story. The only other significant effects were main effects of fact type and ease, to be discussed in the context of their significant interaction [$F(2,90) = 34.98$, $MS_e = 0.03$]. For easy questions, having read misinformation increased its production above baseline [$t(47) = 6.05$, $SEM = .03$]. A similar pattern occurred for hard questions [$t(47) = 8.30$, $SEM = .03$]. The significant interaction was driven by the fact that the misinformation effect was even larger for hard questions than for easy. As is shown in Table 1, however, these patterns emerged for all of the warning conditions.

Thus, although warnings decreased overall misinformation production, they did not help the subjects to selectively edit out the stories' errors.

Discussion

Warnings did have an effect: Overall, the subjects in the warning conditions were more conservative and produced less misinformation. However, this warning effect did not interact with fact framing or story reading. The subjects produced less misinformation in all conditions—even when they had not read the relevant stories. Warnings did *not* help the subjects to selectively edit out wrong facts from the stories.

The failure of warnings to reduce reliance on fiction is surprising, given findings obtained in other false memory paradigms. Of particular interest is that a prereading warning did not help, because warnings given prior to study are effective in reducing suggestibility in the eyewitness misinformation paradigm (Greene et al., 1982) and in the DRM paradigm for creating false memories for words (e.g., Gallo et al., 2001). Although it is possible that our warning was worded less strongly than those used in other studies, it is unlikely that this would explain the difference in results, since the warning did reduce errors overall. Rather, a major difference between the eyewitness and fiction paradigms is that in the fiction paradigm, misinformation is *intermixed* with the story, whereas in the eyewitness paradigm, the misinformation occurs *after* the to-be-remembered event. Unlike mock eyewitnesses, our subjects had to follow the story's plot while simultaneously retrieving general knowledge and monitoring for factual errors. The result was a dual-task situation; a warning could not be useful if the subjects did not have enough cognitive resources to both construct a mental

model of the story and monitor for factual errors. This account fits nicely with Prentice and Gerrig's (1999) dual-process model of how people learn from texts. Readers are proposed to use systematic processes to evaluate the text *if* they are both motivated *and* have the ability to do so. Consistent with Gilbert's (1991) idea that people automatically believe information, readers lacking in motivation or ability will not systematically evaluate information and, rather, are proposed to rely on more automatic processes.

To test these ideas, in Experiment 2 we manipulated the reading difficulty of the stories, on the assumption that easy-to-read stories would leave more cognitive resources available for systematic evaluation of the material (Inhoff & Fleming, 1989). Twelfth-grade stories had longer sentences and contained lower frequency words, whereas 6th-grade stories consisted of short sentences and higher frequency words. If the null effect of warning is due to limited resources to monitor the to-be-read information, suggestibility will be higher when the subjects read difficult stories than when they read easy ones.

EXPERIMENT 2

Method

Subjects. Forty-eight Duke University undergraduates participated in the experiment for partial fulfillment of a course requirement.

Design. A 2 (story reading level: 6th grade or 12th grade) \times 2 (question ease: easy or hard) \times 3 (fact framing: correct, neutral, or misleading) within-subjects design was used. The dependent measures were proportion correct and misinformation answers on the general knowledge test.

Materials. Two stories were modified and used for the second experiment. The stories each made 18 references to facts from the Nelson and Narens (1980) norms; again, half corresponded to easy questions and half corresponded to hard questions. Across subjects, items were rotated through correct, neutral, and misleading frames. For each of the stories, simpler and more complex versions were created. The 6th-grade version contained short sentences and high-frequency words, whereas the 12th-grade version contained longer, more convoluted sentences and lower frequency words. The average sentence length was 9.9 words for the 6th-grade stories and 19.9 words for the 12th-grade stories. In addition, the average number of syllables per word was 1.37 for the 6th-grade stories and 1.62 for the 12th-grade stories. The 6th- and 12th-grade versions of the stories were matched for total number of words (see the Appendix).

We also calculated the Flesch readability index for the two stories using the ReadMe tool (Norwalk, Becker, & Fassino, 2004). The Flesch readability index ranges from 0 to 100, with 0 being the most difficult to understand and 100 being the easiest. Stories created to be more difficult to read had an average Flesch readability index of 50.27, whereas the simpler stories had an average Flesch index of 80.85. These Flesch indexes would typically be interpreted as being at the high school and 6th-grade reading levels, respectively.

During pilot testing, the subjects rated the stories using Green and Brock's (2000) transportation scale. *Transportation* is a self-report measure of how involved a reader is in a story; the term comes from the idea that a reader who is deeply engaged in a story is mentally transported to that fictional world (Green & Brock, 2000). Transported readers endorse items such as "I wanted to learn how the story ended" and "The story affected me emotionally" but not items like "I found my mind wandering while reading the story." The 6th- and 12th-grade versions of the stories were equally transporting. The 6th-grade stories were rated at 3.86 on a 7-point scale, and the 12th-grade stories scored 3.89.

Across subjects, we counterbalanced whether each story was read in 6th- or 12th-grade form and whether the 6th- or 12th-grade story was read first.

Procedure. Phase 1, the reading phase, was conducted on the computer. The subjects were told that they would be reading two fictional short stories, one sentence at a time. All of the subjects were warned to “keep in mind that sometimes authors take liberties with facts or ideas, and that therefore some of the information you will read may be incorrect.” The subjects were told to read each sentence carefully, and to press the “next” key when ready for the next sentence. To ensure that the subjects read the sentences, they were told that they would later be asked questions about the stories. In addition, they were warned that sometimes, after a sentence disappeared, a box would appear and they would have to type in the sentence that they had just read. These catch trials were included to ensure that the subjects read carefully; 5% of the sentences were catch trials and always involved filler (noncritical) sentences. Immediately following each story, the subjects rated their involvement in the story using Green and Brock’s (2000) transportation scale.

All other portions of the experiment were similar to those in Experiment 1, except that all of the testing in Experiment 2 was done on the computer. The final test contained 72 items: 36 critical questions plus 36 fillers. As in Experiment 1, all of the subjects were warned against guessing on the final test and were told to skip any items that they could not answer.

Results

Reading catch-trial performance. The subjects recalled the just-read sentence on 5% of trials; catch trials were always filler sentences. Recall of each sentence was coded as a “0” (showed no memory for the prior sentence), “1” (showed memory for gist of sentence but suggested speed reading), or “2” (showed good memory for the just-read sentence). On average, the subjects showed good memory for what they had just read ($M = 1.69$), supporting the conclusion that they were carefully reading the stories. The subjects were better at recalling sentences from the 6th-grade stories ($M = 1.77$) than from the 12th-grade stories ($M = 1.50$) [$t(47) = 4.99$, $MS_e = 0.05$]; this effect is not surprising, however, given the coding system. Sentences from 12th-grade stories were much longer, and therefore it was expected that they would be recalled in a more summarized form. More important is that performance in both conditions indicated that the subjects were reading the sentences.

Reading times. Because the story sentences varied widely in length, the dependent measure was average read-

ing time per syllable. Syllable counts were obtained with the ReadMe tool (Norwalk et al., 2004).

The subjects read the first story more slowly ($M = 188.6$ msec per syllable) than the second story ($M = 157.4$) [$t(47) = 9.06$, $MS_e = 3.44$]. There was no difference in reading time per syllable between the 6th-grade ($M = 174.1$) and 12th-grade ($M = 171.9$) stories ($t < 1$).

Correct answers. A 3 (fact framing: correct, misleading, or neutral) \times 2 (question ease: easy or hard) \times 2 (story reading level: 6th grade or 12th grade) ANOVA was computed on proportion of correct answers. The data are shown in the top panel of Table 2. As in Experiment 1, the subjects correctly answered more easy questions than hard ones [$F(1,47) = 716.7$, $MS_e = 0.03$]. There was also a main effect of fact framing: The subjects answered more questions correctly after reading the correct answers in the stories ($M = .62$) than after reading neutral facts ($M = .42$). Performance dropped below the neutral baseline when the subjects had read misinformation in the stories ($M = .34$) [$F(2,94) = 62.0$, $MS_e = 0.07$]. Contrary to our predictions, there was no effect of story reading level ($F < 1$), nor did story reading level interact with any of the other factors ($F_s < 1$). In short, fact framing’s effect on correct responding did not differ as a function of the grade level of the stories.

Misinformation production. A 3 (fact framing: correct, misleading, or neutral) \times 2 (question ease: easy or hard) \times 2 (story reading level: 6th grade or 12th grade) ANOVA was computed on proportion of misinformation answers. As is shown in the bottom panel of Table 2, the subjects produced misinformation answers only after having read the misinformation in the stories. This resulted in the expected main effect of fact framing [$F(2,94) = 85.8$, $MS_e = 0.06$]. In addition, the subjects were more likely to produce misinformation as answers to hard questions than as answers to easy questions, as is shown in the interaction between fact framing and question ease [$F(1,47) = 5.7$, $MS_e = 0.04$]. There was no effect of story reading level on the amount of misinformation produced ($F < 1$). Reading the 6th-grade stories did not allow the subjects to better filter out the incorrect information, in comparison with reading the 12th-grade stories. Suggestibility was equal in the two conditions.

Table 2
Production of Correct and Misinformation Answers as a Function of Question Ease, Fact Frame, and Story Reading Level (Data from Experiment 2)

Reading Level	Easy Questions			Hard Questions		
	Correct	Baseline	Misleading	Correct	Baseline	Misleading
Proportion Correct						
6th-grade story	.83	.67	.51	.44	.17	.13
12th-grade story	.82	.63	.51	.40	.21	.18
<i>M</i>	.82	.65	.51	.42	.19	.16
Proportion Misinformation						
6th-grade story	.03	.06	.27	.04	.10	.43
12th-grade story	.03	.06	.26	.06	.08	.38
<i>M</i>	.03	.06	.27	.05	.09	.40

Discussion

Easier stories leave readers with more free cognitive resources (Inhoff & Fleming, 1989), potentially allowing them to learn more correct information and to avoid the stories' errors. However, suggestibility was the same regardless of whether the subjects read a story at a 6th-grade or a 12th-grade reading level. Increasing cognitive capacity (by reducing story difficulty to the 6th-grade level) did not reduce suggestibility.

Because our argument is based on a null effect, we did a power analysis and then collected additional data to support the conclusion that our manipulation had the intended effect. The power analysis was conducted to determine how many subjects would be required in order to achieve a significant effect of story reading level. Using an alpha level of .05 and a power level of .80, one would need 463 subjects in order to show a significant difference between the amount of misinformation produced in the 6th-grade and 12th-grade conditions. The need for such a large number of subjects to achieve significance was unreasonable; even doubling our sample size would be unlikely to yield a significant effect of story reading level. The null effect was not due to low numbers of subjects.

More critical was to ensure that the 6th- and 12th-grade stories actually differed in how much cognitive capacity was required for reading. Because the null effect was only meaningful to the extent that the stories required more or less attention to read, we decided to collect additional data (beyond the Flesch ratings) to support the manipulation. In a follow-up study, 18 new subjects completed only Phase 1 of the experiment. The key procedural difference was that story reading was accompanied by an additional task. The subjects were required to classify the pitch of a 50-msec tone that accompanied 30% of the sentences. The subjects pressed one key for a low tone and another key for a high tone. Accuracy overall was very high ($M = 97\%$ correct). However, the subjects took longer to classify the tones when they were reading a 12th-grade story ($M = 1,006$ msec) than when they were reading a 6th-grade story ($M = 915$ msec) [$t(17) = 2.85$, $SE = 32.03$]. When the subjects were reading a 12th-grade story, it was harder for them to switch to a secondary task, even though the story reading was self-paced. If readers normally monitor for errors while simultaneously following the plot, we would have predicted less monitoring (and more suggestibility) in the 12th-grade condition. Since we did not find reduced errors in the 6th-grade condition, these data suggest that these readers did not use the additional capacity available to them to monitor the text for errors, despite the warning.

The follow-up study yielded a second measure of how difficult the stories were to read; the subjects rated the difficulty of the stories on a 7-point scale. Ratings were made on four statements, which included "I frequently had to pause while reading the sentences in order to comprehend the story," and "The story would be appropriate in reading level for elementary school children." The 6th-grade stories were rated as easier to read ($M = 2.6$) than the 12th-grade stories ($M = 3.6$) [$t(17) = 3.24$, $SE = .31$].

Thus, three different data points converged on the conclusion that 6th-grade stories were easier to read: Sixth-grade stories had a lower Flesch level, they were rated as easier by the subjects themselves, and they interfered less with ability to do a secondary task. The availability of the extra resources, however, did not reduce suggestibility in the 6th-grade condition. When the subjects were reading the 6th-grade stories, they had been warned of the possibility of errors in the text and had the resources to look for errors, yet they showed no reduction in errors on the final general knowledge test.

In both Experiment 2 and the warning-before condition of Experiment 1, the warning was stated only once. No further warnings were given during or after story reading, and the subjects made no overt response upon discovering an error in the story. This procedure required the subjects to keep in mind that there could be errors in the story while they were reading. That is, the subject still needed to do several things in addition to following the storyline: He/she needed to remember to look for errors, and had to retrieve general knowledge to evaluate what he/she was reading. In Experiment 3, we removed the need to keep the monitoring goal in mind (a prospective memory load; Guynn, McDaniel, & Einstein, 1998). In one condition, the subjects were continually reminded of the text's possible problems. As in Experiment 2, all of the subjects were warned that they would read stories that might contain errors, and the stories were read sentence by sentence on the computer. In the control condition, as in Experiment 2, the subjects simply pressed a key when they were ready for the next sentence. In contrast, the subjects in the detection condition pressed one of two keys when they were ready for the next sentence: one key if the sentence was error-free, and another if they had detected an error in the sentence. Would the negative consequences of story reading persist in the detection condition, in which the subjects searched for errors on each sentence and actively marked any errors they discovered?

EXPERIMENT 3

Method

Subjects. Seventy-two Duke University undergraduates participated in the experiment for partial fulfillment of a course requirement.

Design. A 2 (question ease: easy or hard) \times 3 (fact framing: correct, neutral, or misleading) \times 2 (detection task: detect or control) mixed design was used. All factors except detection task were manipulated within subjects. The dependent measures were reading time, ability to detect errors during story reading, and proportion correct and misinformation answers on the general knowledge test.

Materials. The original versions of the two stories used in Experiment 2 were used in this experiment (see Marsh, 2004); the average reading level was 8th grade.

Procedure. The procedure was the same as in Experiment 2 except for the addition of the detection condition. The subjects in both conditions received the same general warning prior to story reading. Following the warning, the subjects in the control condition were told to read each sentence carefully, and to press the "next" key when ready for the next sentence. The subjects in the detection condition were also instructed to read each sentence carefully, and to press one of two keys when ready for the next sentence. They were instructed to press the "next" key if the sentence contained no errors, and to press an "error"

key if the sentence contained one or more errors. In addition, the number of catch trials was increased to 15% of the sentences.

Results

The Results section is arranged as follows. Three analyses will be presented from the reading phase: performance on the catch trials, time to read different kinds of sentences, and (for subjects in the detection condition) ability to detect errors in the sentences. Two analyses will be presented from the test phase: production of correct and target misinformation as answers on the final general knowledge test. Finally, one analysis will examine whether subjects answered general knowledge questions with misinformation that had been “detected” during the reading phase.

Reading catch-trial performance. Responses were scored in the same manner as in Experiment 2. On average, performance was excellent in both the control ($M = 1.95$) and detection ($M = 1.92$) conditions, supporting the conclusion that subjects were reading the stories in both conditions carefully.

Reading times. For the critical sentences, time to make a keypress did not depend on fact framing ($F < 1$). The subjects were no slower upon encountering misinformation ($M = 265.8$ msec per syllable) than when they had read a neutral ($M = 266$) or correct ($M = 260.6$) frame.

The subjects were slower in the detection condition ($M = 320.1$) than in the control condition ($M = 208.1$) [$F(1,70) = 41.35$, $MS_e = 32,731.1$]. The only other significant effects were a main effect of ease [$F(1,70) = 67.67$, $MS_e = 3,703.7$] and a significant interaction between reading condition and ease [$F(1,70) = 12.25$, $MS_e = 3,703.7$]. In the detection condition, the subjects were particularly slow to press a key for hard items ($M = 354.4$) as compared with easy items ($M = 285.8$); the difference in reaction times between hard items ($M = 222.1$) and easy items ($M = 194.3$) was smaller in the control condition.

In summary, the detection task slowed reading, and this was particularly true when the subjects were reading less well-known facts. The subjects were no slower when they read sentences containing errors in either condition. However, as the next set of analyses will support, this did *not* imply that they were unable to detect errors.

Error detection during story reading. For the subjects in the detection condition, a 2 (question ease: easy or hard) \times 3 (fact framing: correct, neutral, or misleading) ANOVA was computed on proportion of critical sentences labeled as containing an error (via keypress). As expected, the subjects were more likely to press the “error” key ($M = .33$) when the sentence contained misinformation than when it contained the correct answer ($M = .22$) or made only a neutral reference to the fact ($M = .21$). This main effect of fact framing was significant [$F(2,70) = 6.58$, $MS_e = 0.05$] and was further qualified by an interaction with question ease [$F(2,70) = 5.70$, $MS_e = 0.03$]. The subjects were better at discriminating truth from falsehood for better-known items (those corresponding to easy questions); for these items, they were least likely to press the “error” key for correct facts ($M = .17$), followed by the

neutral frames ($M = .26$), and were most likely to press it for misinformation items ($M = .35$). For lesser-known facts (those corresponding to hard questions), however, the subjects were least likely to press the “error” key for the neutral items ($M = .17$) and more likely to press it for correct ($M = .26$) or misinformation ($M = .31$) items.

In short, the subjects showed some ability to detect the misinformation items; error keypresses were most frequent for misinformation items. The subjects were better at correctly identifying contradictions of well-known facts (easy items) than of lesser-known items. However, performance was far from perfect; the subjects deemed 67% of misinformation sentences error free.

Correct answers on the general knowledge test. A 2 (question ease: easy or hard) \times 3 (fact framing: correct, neutral, or misleading) \times 2 (task: control or detection) mixed ANOVA was computed on proportion correct answers on the final test. As expected, the subjects answered more easy questions correctly [$F(1,71) = 445.33$, $MS_e = 0.04$]. Also significant was the main effect of fact framing [$F(2,142) = 150.23$, $MS_e = 0.03$]. Reading the correct answers in the story ($M = .64$) significantly boosted performance above baseline ($M = .37$) [$t(71) = 13.34$, $SEM = .02$]. In contrast, reading misinformation answers in the story ($M = .30$) led to performance significantly below baseline [$t(71) = 3.55$, $SEM = .02$]. No other effects were significant; detection did not significantly affect correct answers ($F < 1$), nor did it interact with question ease or fact type.

Misinformation production. Of particular interest was when the subjects used misinformation to answer questions on the final general knowledge test, even though they had been warned not to guess. A 2 (question ease: easy or hard) \times 3 (fact framing: correct, neutral, or misleading) \times 2 (task: detect or control) ANOVA was computed on proportion misinformation answers. The data are shown in Figure 1. Two major points are evident in the figure. First, replicating prior work, the subjects rarely produced misinformation answers except when they had read them in the stories; this was reflected in the main effect of fact framing [$F(2,140) = 138.35$, $MS_e = 0.26$]. Second, and most important, the misinformation effect was reduced in the detection condition, as reflected in the fact framing \times detection interaction [$F(2,140) = 3.52$, $MS_e = 0.26$]. In both conditions, more misinformation was produced after subjects read the errors in the stories than after they read correct or neutral frames; however, suggestibility to story errors was greater in the control condition than in the detection condition.

Were errors caught during study ever replicated on the final general knowledge test? To answer this question, a 2 (question ease: easy or hard) \times 3 (fact framing: correct, neutral, or misleading) ANOVA was computed on proportion of questions answered with misinformation that had been detected during reading. The effect of fact framing remained significant [$F(2,70) = 4.85$, $MS_e = 0.01$]. When subjects had read (and detected) misinformation in the stories, the same errors were produced as answers to 5% of corresponding questions at test. This 5% rate might seem low, but it was significantly different from zero [$t(35) =$

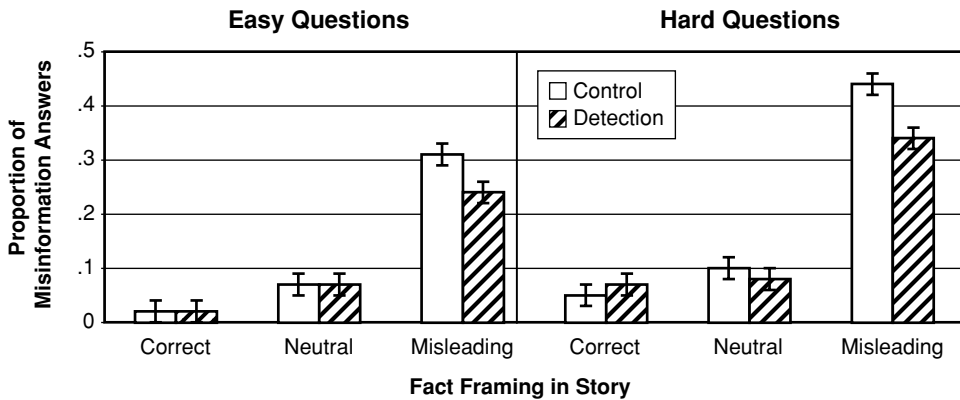


Figure 1. The effect of reading correct and misleading facts on proportion misinformation answers on the cued recall test, as a function of detection condition. Easy questions are on the left and hard questions are on the right. (Data from Experiment 3.)

3.95, $SE = .01$] and was higher than that observed after reading correct ($M = .01$) or neutral ($M = .025$) frames.

Discussion

The subjects in the detection condition were able to detect misinformation above chance, although they failed to detect the majority of the errors in the stories. Consider the easy items: On the basis of the norms, we would predict that the subjects would be able to correctly answer 65% of the cued recall questions (and that the subjects' recognition of the correct answer would likely be higher). Consistent with this, the subjects in the control condition answered 59% correctly, and again we would expect recognition rates to be higher. So the puzzle is why the subjects pressed the "error" key for only 32% of errors involving well-known facts, when both the norms and the data from the control subjects suggest that detection subjects had the knowledge to recognize the errors at a much higher rate.

Detection instructions significantly reduced, but did not eliminate, suggestibility. When analyses were restricted to items that the subjects had labeled as errors during reading, the suggestibility effect was small but significant; given "error" keypresses during study, the subjects were still more likely to produce target misinformation on the final cued recall test than if they had read neutral or correct information in the stories.

Detection instructions had a different effect than did warnings in Experiment 1. In Experiment 1, a warning served to reduce misinformation production in all conditions, rather than leading to the subjects' selectively editing out errors that they had read in the stories. In Experiment 3, detection instructions specifically reduced misinformation production after subjects had read misinformation, and not in other conditions. That is, detection did not make subjects more conservative overall, but rather helped them to avoid producing specific errors that they had read in the stories.

Detection did slow reading, probably both because the secondary task (monitoring) required careful reading, and

because this task required discriminating between two keys instead of one. Detection subjects were particularly slow when the monitoring task was difficult; they took more than an additional 65 msec per syllable to read hard items as compared with easy items (whereas control subjects were less than 30 msec slower for hard items). Detection did have consequences: The subjects were above chance in detecting errors, leading to a reduction in suggestibility. These results correspond nicely with investigations of reading times in the eyewitness domain; slower readers (by nature or by instruction) are more likely to detect errors in a postevent narrative and are less suggestible (Tousignant, Hall, & Loftus, 1986). However, the subjects were far from perfect in detecting errors, and although suggestibility was reduced, it was not eliminated in the detection condition.

GENERAL DISCUSSION

In three experiments, subjects persisted in relying on fictional sources despite experimental manipulations designed to reduce suggestibility. In Experiment 1, warnings provided before or after story reading did not lead to avoidance of story errors. Warnings reduced the overall error level; this effect was not specific to errors that had been read in the stories. In Experiment 2, subjects were just as suggestible when errors were embedded in a text that was easy to read; additional cognitive resources did not decrease learning of the stories' errors. In Experiment 3, suggestibility was reduced but not eliminated when subjects were required to press a key every time they discovered an error in a story sentence. Subjects were above chance but far from perfect in detecting errors; correspondingly, suggestibility was reduced but far from eliminated in the detection condition.

The main effect of detection found in Experiment 3 suggests that detecting errors during reading is the first step toward avoiding later reproduction of fiction's errors. The data from Experiments 1 and 2 suggest that it is quite difficult for readers to monitor for errors during story

reading, since neither warnings nor increased cognitive capacity decreased suggestibility. These results parallel those found in the domain of text comprehension research. For example, only 40% of readers of a passage about an airplane crash detected a problem with the sentence "The authorities were trying to decide where to bury the survivors," even though "bury" implies that the passengers died whereas "survivors" suggests that they did not die (Barton & Sanford, 1993). Similarly, the majority of subjects are willing to answer the question "How many animals of each kind did Moses take on the ark?" even though most subjects know that Noah, not Moses, was in charge of the ark (Erickson & Mattson, 1981). Although it could be argued that the preceding examples are rather subtle, detection rates are still below 50% when readers are warned and the errors are more blatant (e.g., *Ronald Reagan was the governor of Montana*; Baker, 1985). Semantic illusions increase when contradictions in expository text are not central to the focus of attention (Baker & Wagner, 1987) and when they are embedded in three-paragraph passages rather than in a single paragraph (Glenberg, Wilkinson, & Epstein, 1982). Consider the implications for the present research: Our stories were several pages long, and the misinformation was peripheral to the plot of the stories—conditions likely to hamper ability to detect an error. We suspect that better fiction would only increase the problem, given Green and Brock's (2000) finding that transportation into a narrative decreased subjects' likelihood to criticize the narrative; such story content manipulations await future research.

It is noteworthy that detection *did* take cognitive resources; subjects in that condition read more slowly than in the control condition. However, it appears more critical to remind subjects to monitor than to free up cognitive resources for them. Freeing cognitive resources in Experiment 2 (by reducing the difficulty of the to-be-read text) did not lead to reduced suggestibility. Rather, subjects need to be reminded continually to take advantage of increased cognitive resources to monitor information.

And detection is not the only prerequisite for avoiding suggestibility; the suggestibility effect remained in Experiment 3 when the analysis was restricted to items that had been labeled as "errors" during story reading. Even if an error is detected during story reading, it may still increase in familiarity, just as items explicitly labeled as "false" at encoding still gain in later rated truth value (Begg et al., 1992). In this way, the fiction paradigm is more closely linked to other paradigms involving familiarity-driven errors (e.g., Kelley & Lindsay, 1993), which may be immune to source-based warnings, than to paradigms such as the eyewitness misinformation paradigm (which may rely more on source amnesia).

Overall, our data can be interpreted using Prentice and Gerrig's (1999) dual-process model of how reading fiction affects people's attitudes and general knowledge store. They argued that both motivation and ability are prerequisites for rejecting an error and that, in their absence, subjects may rely on heuristics such as familiarity. The default mode of processing for fictional material is thought

to be nonsystematic; it is only in special situations that subjects will have both the motivation and the ability to reject false facts. In Experiment 1, we increased the subjects' motivation to read the stories critically by warning the subjects about the errors in the study. That increase in motivation, by itself, was not enough to decrease the effect of the false facts on the general knowledge test. Increasing the subjects' ability to reject the errors by increasing the available cognitive capacity in Experiment 2 also had no effect. It was only in Experiment 3, when the subjects had to actively detect the errors, that the influence of the misinformation was reduced. It is only when the monitoring goal is active throughout story reading that there is a reduction in the production of false facts.

REFERENCES

- BAKER, L. (1985). Differences in the standards used by college students for evaluating their comprehension of expository prose. *Reading Research Quarterly*, **20**, 297-313.
- BAKER, L., & WAGNER, J. L. (1987). Evaluating information for truthfulness: The effects of logical subordination. *Memory & Cognition*, **15**, 247-255.
- BARTON, S. B., & SANFORD, A. J. (1993). A case study of anomaly detection: Shallow semantic processing and cohesion establishment. *Memory & Cognition*, **21**, 477-487.
- BEGG, I. A., 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**, 446-458.
- BOWER, G. H., & MORROW, D. G. (1990). Mental models in narrative comprehension. *Science*, **247**, 44-48.
- CHAMBERS, K. L., & ZARAGOZA, M. S. (2001). Intended and unintended effects of explicit warnings on eyewitness suggestibility: Evidence from source identification tests. *Memory & Cognition*, **29**, 1120-1129.
- COLERIDGE, S. T. (1906). *Biographia literaria*. London: J. Dent. (Original work published 1817)
- DUBECK, L. W., MOSHIER, S. E., & BOSS, J. E. (1988). *Science in cinema: Teaching science fact through science fiction films*. New York: Columbia University, Teachers College Press.
- EMMERICH, R. (DIRECTOR) (2000). *The patriot* [Motion picture]. Columbia Pictures.
- ERICKSON, T. D., & MATTSON, M. E. (1981). From words to meaning: A semantic illusion. *Journal of Verbal Learning & Verbal Behavior*, **20**, 540-551.
- GALLO, D. A., ROEDIGER, H. L., III, & McDERMOTT, K. B. (2001). Associative false recognition occurs without strategic criterion shifts. *Psychonomic Bulletin & Review*, **8**, 579-586.
- GERRIG, R. J. (1989). Suspense in the absence of uncertainty. *Journal of Memory & Language*, **28**, 633-648.
- GERRIG, R. J. (1993). *Experiencing narrative worlds: On the psychological activities of reading*. New Haven: Yale University Press.
- GILBERT, D. T. (1991). How mental systems believe. *American Psychologist*, **46**, 107-119.
- GLENBERG, A. M., WILKINSON, A. C., & EPSTEIN, W. (1982). The illusion of knowing: Failure in the self-assessment of comprehension. *Memory & Cognition*, **10**, 597-602.
- GREEN, M. C., & BROCK, T. C. (2000). The role of transportation in the persuasiveness of public narratives. *Journal of Personality & Social Psychology*, **79**, 701-721.
- GREENE, E., FLYNN, M. S., & LOFTUS, E. F. (1982). Inducing resistance to misleading information. *Journal of Verbal Learning & Verbal Behavior*, **21**, 207-219.
- GUYN, M. J., MCDANIEL, M. A., & EINSTEIN, G. O. (1998). Prospective memory: When reminders fail. *Memory & Cognition*, **26**, 287-298.
- HOVLAND, C. I., LUMSDAINE, A. A., & SHEFFIELD, F. D. (1949). *Experiments on mass communication*. Princeton, NJ: Princeton University Press.
- INHOFF, A. W., & FLEMING, K. (1989). Probe-detection times during the

- reading of easy and difficult text. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **15**, 339-351.
- JOHNSON-LAIRD, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge, MA: Harvard University Press.
- JONZE, S. (DIRECTOR) (1999). *Being John Malkovich* [Motion picture]. USA Films.
- KELLEY, C. M., & LINDSAY, D. S. (1993). Remembering mistaken for knowing: Ease of retrieval as a basis for confidence in answers to general knowledge questions. *Journal of Memory & Language*, **32**, 1-24.
- LEWIS, C. H., & ANDERSON, J. R. (1976). Interference with real world knowledge. *Cognitive Psychology*, **8**, 311-335.
- LINDSAY, D. S., & JOHNSON, M. K. (1989). The eyewitness suggestibility effect and memory for source. *Memory & Cognition*, **17**, 349-358.
- MARSH, E. J. (2004). Story stimuli for creating false beliefs about the world. *Behavior Research Methods, Instruments, & Computers*, **36**, 650-655.
- MARSH, E. J., MEADE, M. L., & ROEDIGER, H. L., III (2003). Learning facts from fiction. *Journal of Memory & Language*, **49**, 519-536.
- MEADE, M. L., & ROEDIGER, H. L., III (2002). Explorations in the social contagion of memory. *Memory & Cognition*, **30**, 995-1009.
- 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.
- NORWALK, L., BECKER, A., & FASSINO, D. (2004). ReadMe (Version 2.0) [Computer software]. Retrieved September 1, 2004, from www.cba.nau.edu/becke-a/Accessibility/ReadMe.html.
- PETERSON, S. B., & POTTS, G. R. (1985). Incorporation versus compartmentalization in memory for discourse. *Journal of Memory & Language*, **24**, 107-118.
- PRENTICE, D. A., & GERRIG, R. J. (1999). Exploring the boundary between fiction and reality. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 529-546). New York: Guilford.
- PRENTICE, D. A., GERRIG, R. J., & BAILIS, D. S. (1997). What readers bring to the processing of fictional texts. *Psychonomic Bulletin & Review*, **4**, 416-420.
- SPIELBERG, S. (DIRECTOR) (1993). *Jurassic Park* [Motion picture]. Universal Pictures.
- STRANGE, J. J., & LEUNG, C. C. (1999). How anecdotal accounts in news and fiction can influence judgments of a social problem's urgency, causes, and cures. *Personality & Social Psychology Bulletin*, **25**, 436-449.
- TOUSIGNANT, J. P., HALL, D., & LOFTUS, E. F. (1986). Discrepancy detection and vulnerability to misleading postevent information. *Memory & Cognition*, **14**, 329-338.
- WHEELER, S. C., GREEN, M. C., & BROCK, T. C. (1999). Fictional narratives change beliefs: Replications of Prentice, Gerrig, and Bailis (1997) with mixed corroboration. *Psychonomic Bulletin & Review*, **6**, 136-141.

APPENDIX

Excerpt From the Inventor Story (as Used in Experiments 1 and 3)

It was a crisp fall day, Billy was ten years old and in Ms. Pringle's fifth grade class. The day began as a normal Friday—everyone was talking about their weekend plans, and planning their costumes for the upcoming Halloween parties. But everything changed for Billy when Ms. Pringle made her announcement:

"Boys and girls! Your attention please! I have exciting news." She paused, waiting for the students' full attention. "I'm happy to announce that this year's science fair will be held just three short weeks from today." She held up her hand to stop the groans from the children. "Now class, you know this can be fun. To help motivate you, we have a special prize this year—the winner of the science fair will win a trip to the national contest, which will be held in *Dover / xxxxx / Wilmington*, the capital of Delaware. And the winner of that contest will get to go to the international science fair in *Moscow / xxxxx / St. Petersburg*, the capital of Russia!"

Sample General Knowledge Questions

What is the capital of Delaware?

What is the capital of Russia?

Sample From the 6th-Grade Version of the Story (Compare to Second Paragraph Above)

"Boys and girls! Your attention please! I have good news." She stopped. The kids slowly quieted down. They were so excited for the weekend that it took a few minutes. They began to wonder what the news was. Perhaps snow was coming, and school would end early today! Or maybe an elephant was on the loose!

Sample From the 12th-Grade Version of the Story

"Boys and girls! Your attention please! I have an exhilarating announcement." She paused dramatically, allowing the children to speculate something thrilling enough to warrant their temporary silence. Perhaps there was a freak blizzard and school was being released early or perhaps Lula the circus elephant had escaped and was trampling towards the school as she spoke.