Applied Cognitive Psychology, *Appl. Cognit. Psychol.* **26**: 556–567 (2012)
Published online 12 January 2012 in Wiley Online Library (wileyonlinelibrary.com) **DOI**: 10.1002/acp.2827

Positive and Negative Effects of Monitoring Popular Films for Historical Inaccuracies

SHARDA UMANATH*, ANDREW C. BUTLER and ELIZABETH J. MARSH

Duke University, Durham, USA

Summary: History educators often use popular films in the classroom to teach critical thinking through an exercise that involves identifying historical inaccuracies in the films. We investigated how this exercise affects the acquisition of true and false historical knowledge. In two experiments, subjects studied texts about historical topics and watched clips from corresponding films. Each film contained one piece of information that contradicted the text (i.e. misinformation). Some subjects received instructions to monitor for inaccuracies in the films. After a delay, they were tested on the texts. Monitoring instructions did not reduce subjects' acquisition of misinformation, and even when subjects successfully detected the inaccuracies, they sometimes still reproduced the misinformation. However, when they received feedback about the inaccuracies, the production of misinformation was substantially reduced. Overall, these findings indicate that educators should provide feedback when using popular films for this critical thinking exercise so that students do not acquire false knowledge. Copyright © 2012 John Wiley & Sons, Ltd.

An important goal of education at all levels is to enable students to become critical consumers of information. Although this objective is relevant to every subject taught in school, it is particularly important for history education because students must learn to assess evidence from a variety of sources in order to form an accurate understanding of the past. One common method of teaching students this skill is to give them practice with assessing the accuracy of information in sources containing a mix of accurate and inaccurate information (Marcus & Stoddard, 2007; see also Berumen, 2008). History educators often choose to use popular films for this purpose as they are a prominent source for learning about history outside of the classroom and shape how the general public views the past (Burgoyne, 1997). Many popular films are based on real historical people and events (Niemi, 2006) and thus contain a lot of accurate information; however, the producers of these films often take liberties with the facts in order to tell a more entertaining story, so they also contain major historical inaccuracies (Jones, 1989; Carnes, 1995; Toplin, 1996; Stoddard & Marcus, 2010).

The primary aim of the present research was to investigate how the commonly used task of asking students to evaluate the accuracy of popular films affects their learning about history. Educators want students to develop a critical eye for the media's depiction of history through this learning activity; however, by exposing students to inaccurate information (i.e. misinformation) in the films, they risk the possibility that students will retain these historical inaccuracies (e.g. Butler, Zaromb, Lyle, & Roediger, 2009). Thus, we wanted to examine whether this pedagogical task is effective in terms of preventing students from acquiring false knowledge about history. A secondary goal was to explore whether the level of student engagement in the task influences its efficacy as a learning tool. Often, educators use popular films because they believe that the entertaining nature of these films will spark student interest in the material, thus increasing motivation and learning (e.g. Jones, 1989; Berumen, 2008; Michel, Roebers, & Schneider, 2007; Smilanich & Lafreniere, 2010). Of interest was the relationship between students' engagement with the materials and the learning of correct and incorrect information.

Before describing the present research, we provide some background to motivate these questions of interest. Drawing on the broader cognitive psychology literature, we first discuss findings that predict that asking students to assess the accuracy of popular films will be beneficial in terms of increasing the acquisition of correct information and decreasing the acquisition of misinformation. We then describe other findings that predict that this task might be detrimental to learning, thus undermining educators' objectives. Finally, we consider student engagement during learning and how it might affect learning from popular films.

POSITIVE EFFECTS OF MONITORING FOR ERRORS

In essence, the classroom exercise of identifying inaccuracies in a popular film is a monitoring task: students must evaluate the historical accuracy of information presented in the film with the goal of detecting major errors. Such a task has the potential both to reduce later reproduction of erroneous information and to enhance learning of correct information.

The task of detecting inaccuracies in films may decrease the acquisition and retention of the misinformation for several reasons. First, the detection task includes a warning about the presence of misinformation, and warnings have been shown to reduce the acquisition of incorrect information (e.g. Eakin, Schreiber, & Sergent-Marshall, 2003; Greene, Flynn, & Loftus, 1982). For example, Butler et al. (2009) found that informing students about what information in the film is inaccurate (a 'specific warning') led to a reduction in the production of misinformation on a later test. However, such passive tools are not always effective—general warnings were found to be ineffective in the same study (see also Green, Garst, Brock, & Chung, 2006). Second, explicitly asking students to detect inaccuracies encourages active monitoring of presented information, which is likely to produce long-lasting corrections of erroneous information. Previous studies have found that students do not appear to spontaneously engage in this sort of monitoring behavior (e.g. Lee, 1982). For example, when classroom activities involving

*Correspondence to: Sharda Umanath, Department of Psychology & Neuroscience, Duke University, Box 90086, Durham, NC 27708-0086, USA.
E-mail: sharda.umanath@duke.edu

history-based films do not compel students to analyze the films' portrayal of historical events (see Jones, 1989; Marcus, 2005; Hobbs, 2006; Marcus & Stoddard, 2007), students appear to use the films as a tool for memorizing the 'facts' of the historical topic (Stoddard & Marcus, 2010). Third, when students successfully detect inaccurate content, they are generating this knowledge for themselves, which should increase retention of the correction (Slamecka & Graf, 1978). Finally, successful detection may undermine the credibility of films as a source of correct historical information, making it less likely that students will acquire false knowledge from them (Lindsay & Johnson, 1989; Marcus, 2005).

Detection tasks have been shown to be effective in reducing later reproduction of misinformation (e.g. Loftus, 1979; Marsh & Fazio, 2006). For example, when subjects read fictional stories containing both correct and incorrect information, Marsh and Fazio (2006) found that simply asking the subjects to mark the errors as they read reduced the likelihood that they would later answer general knowledge questions with story errors. Successful detection further reduced the later production of misinformation, but it did not fully eliminate the effect. Thus, the active task of detecting historical inaccuracies in films may be very effective in preventing the acquisition and later reproduction of misinformation.

In addition to decreasing the acquisition and retention of misinformation, this monitoring task may simultaneously enhance students' learning of correct information from films and associated texts. Students' monitoring for misinformation in films is likely to increase the depth at which they process all of the materials, including accurate content in the films and the texts. The evaluative nature of the task may also promote relational processing of the texts and films as students attempt to compare their respective content to find the misinformation (Hunt & Einstein, 1981). Also, engaging in such a detection exercise may increase student engagement beyond the heightened level of engagement that comes from simply using popular films (e.g. Green & Brock, 2000; see section on "Engagement" below).

NEGATIVE EFFECTS OF MONITORING FOR ERRORS

The same detection exercise could also have no effect on the learning of misinformation or worse it might actually increase the acquisition and retention of the inaccuracies from viewed films. Previous research suggests several reasons why active monitoring for inaccuracies could produce this unintended effect. First, the possibility exists that students, despite their best efforts, fail to detect the inaccuracies in the films. Marsh and Fazio (2006) found that even when asked to detect errors while reading fictional stories, subjects did not notice many of them, including errors based on well-known facts. Critical evaluation of films is especially challenging for students (Lee, 1982) because films induce an entertainment-oriented mindset in the viewer rather than an evaluative one (Green et al., 2006). Second, research from Gilbert and colleagues suggests that people tend to believe that any information they encounter in the world is true and that disbelieving overall is an effortful process (Gilbert, 1991; Gilbert, Tafarodi, & Malone, 1993).

Even though films and other works of fiction are known to be less trustworthy, people tend not to discount them as inherently untrue (Prentice & Gerrig, 1999; Green et al., 2006). Thus, we might expect that if students are unable to correctly identify a film's inaccuracies, they may be likely to retain that misinformation. As mentioned above, this active task is likely to increase attention to the materials overall, but if students are unable to identify the inaccuracies or mistakenly label content as misinformation, students may discard accurate information and acquire misinformation instead. In addition, undermining the credibility of the films may reduce the acquisition of all information from the films, including accurate historical content (Greene et al., 1982).

Perhaps most frustrating to educators and researchers alike, students may correctly detect and identify the misinformation portrayed in the film and yet still go on to reproduce that erroneous content on a final test (e.g. Eslick, Fazio, & Marsh, 2011; Fazio & Marsh, 2008). When the films are the medium of narrative, it may be that the memory of the film, which is rich in sensory information, is more highly accessible than the memory of the accurate content from the text. Similarly, prior research demonstrates that people find distinguishing facts and fiction challenging, with facts and fiction being rated as equally true (Green & Brock, 2000; Green et al., 2006). Increased engagement may then enhance one's tendency to encode the inaccuracies as true, even if one initially detected them (Green, Garst, & Brock, 2004; Green & Brock, 2000).

ENGAGEMENT

A secondary aim of the current research was to examine the influence of student engagement in films and associated texts on the learning of history material. We define 'engagement' in the present research as a construct that incorporates interest (Silvia, 2008) and transportation into a narrative (Gerrig, 1993).

The construct of interest can be described as the intersection of novelty, complexity, and comprehensibility (Silvia, 2008; Silvia & Kashdan, 2009; Silvia & Berg, 2011). When something is new and complex, but not so complex as to be incomprehensible, it is likely to be interesting. Major motion pictures depicting historical figures and events can arouse a great deal of general interest in the historical topics they illustrate. For example, the release of the popular film *Amadeus* (1984) propelled Wolfgang Amadeus Mozart to the forefront of popular culture and precipitated a period known as 'Mozart Mania' (Brown, 1992). Interest is relevant to the present research because previous findings show that interest leads to increased learning (Silvia, 2006; Silvia, 2008). That is, interest motivates students to study for longer and read more carefully (among other learning behaviors), and thus, they acquire and retain more of the material (Silvia, 2006).

Films may also be considered engaging for their transporting quality. That is, individuals who are watching a film may be absorbed or 'transported' into the narrative world of the film. Transportation is a convergent process in which mental resources and attentional focus are directed at the events occurring in the world created by a narrative that is presented via a film, text, or any other medium (Gerrig, 1993; Green &

Brock, 2000). Transportation into the films and texts has the potential to increase learning by ensuring that students are fully concentrated on the narratives (Green, Brock, & Kaufman, 2004); however, this construct is also associated with a reduction of monitoring for accuracy because transported viewers suspend their disbelief and have fewer cognitive resources available (Green, Garst & Brock, 2004). Thus, although there are potential benefits of students being transported into the narratives of historical films, possible drawbacks also exist. On the basis of the prior research on interest and transportation, it appears that greater engagement (i.e. increased interest and transportation) is related to increased learning. However, this increased learning may be blind to the accuracy of the content—that is, it could occur for both true and false information. Of interest in the present study is whether engagement in the films and associated texts will predict the acquisition of correct and/or incorrect information.

EXPERIMENT 1

In the present set of experiments, we investigated how asking students to detect the inaccurate information in films would affect the acquisition of false knowledge. We also examined the influence of student engagement in films and associated texts on the process of learning from these materials. Experiment 1 consisted of a single session. First, subjects read historical texts and viewed associated films. The order in which the text and film were presented was manipulated to explore any order effects because sometimes educators show the film first and then have students read the text or vice versa. In addition, the film clip was not shown for some topics as a control condition. All of the subjects were instructed to learn the information in the texts for a future test and were told that the film clips were merely to illustrate the material. Subjects in the detection condition were further instructed to identify one major historical inaccuracy in each film clip (which would be contradicted by the text). After each text and film clip, they answered a series of four questions about their engagement while reading and viewing, respectively. Finally, subjects took a short answer test on the historical texts.

Method

Subjects and design

Thirty-six Duke University undergraduates participated in this study for payment. An additional three subjects were tested but excluded for not following instructions properly in the experiment. The experiment had a 3 (presentation order: read-only, read–view, view–read) × 2 (instructions: detection, control) mixed design. Presentation order was manipulated within-subjects, whereas instructions were manipulated between-subjects.

Materials and counterbalancing

Study materials consisted of nine film clips and nine corresponding historical texts that were adapted from Butler et al. (2009). Each film clip was taken from a major motion picture that depicted a historical event or figure and lasted approximately 5 minutes. The corresponding historical texts were each approximately 800 words and contained two pieces of information that overlapped with the film clip. One of these pieces of information was consistently presented in both the text and the film. The other piece of information was inconsistently presented: it was correctly described in the text, but incorrectly depicted in the film (i.e. the text directly contradicted the film). Critically, this piece of incorrect information in the film was a major historical inaccuracy. For example, the clip from the film *Amadeus* depicted a young Mozart behaving in a childish and eccentric manner; however, this portrayal is historically inaccurate because he was known for being quite respectful and polite in court despite his young age. The study materials were counterbalanced in several ways. First, each topic (i.e. a film and text pair) was rotated through the three presentation order conditions. Second, each topic was rotated through the nine presentation positions (i.e. first, second, third). Thus, overall each topic appeared in each condition, and each presentation position equally often across subjects.

In addition to the study materials, the experiment contained two types of questions that were given during the initial learning phase: engagement questions that assessed subjects' experience with studying the materials and prior knowledge questions about their previous experience with the materials. After either viewing a film or reading a text, subjects answered a series of four engagement questions that assessed their level of interest, involvement, desire to learn more, and mind wandering, respectively, while interacting with the materials (see Table 1). Subjects responded to each question by using a 5-point scale. In order to gauge prior knowledge, they were also asked how many times they had seen each film (if any) and how much they knew about the historical topic in each text.

A final test was constructed to assess subjects' memory for the texts and films. This test was adapted from materials used in Butler et al. (2009) and consisted of two types of critical questions: text/film consistent and text/film inconsistent.

Table 1. Mean rating given to the text and film clips on each of the four engagement questions and the aggregate engagement rating in Experiment 1

	Text	Film
How interesting was this text [film]?	3.5	3.8
How involved were you in the narrative while reading [viewing] the text [film]?	3.3	3.7
How much do you want to learn more about this topic?	3.0	3.4
How much did your mind wander while reading [watching]?	2.5	2.3
Aggregate engagement rating	3.4	3.7

Note: The mind-wandering question was reverse scored in creating the aggregate engagement rating.

Text/film consistent questions were drawn from information that was presented accurately in both the text and the film. Text/film inconsistent questions tapped the information that was inaccurately depicted in the film, but correctly described in the text. In addition, there were filler items that asked about information that was only presented in the text. For each topic, there was one text/film consistent question, one text/film inconsistent question, and two filler questions. Overall, the test contained 36 questions.

Procedure

The entire experiment was conducted on a computer using MEDIALAB and DIRECTRT software (Empirisoft Corporation, New York, NY, U.S.A) (Jarvis, 2008a, 2008b). Upon arrival, subjects were randomly assigned to one of the two instructions conditions. They were told that their goal was to learn the material in the texts for a later test and that the film clips were only to illustrate the text material. Subjects in the detection condition were also instructed to search for one major historical inaccuracy in each film clip (they were told that the inaccuracy would be directly contradicted in the corresponding text). After receiving the instructions, the subjects read the texts and viewed the film clips. A brief orienting paragraph preceded each film clip to provide some context about the characters and historical events. Reading of both orienting paragraphs and the historical texts was self-paced. Following each text and film, they answered the four engagement questions and the prior knowledge question. After all nine topics, subjects completed a visual search filler task for approximately 3 minutes. Finally, they took the short answer test that contained the text/film consistent, text/film inconsistent, and text-only filler questions. Subjects were instructed to answer these questions on the basis of their memory of the texts and not the films. The questions were presented in a random order, and responding was self-paced. Subjects were required to provide a response for every question. If they did not know the response, they were instructed to generate a plausible guess. After they completed the final test, they were debriefed.

Results

All results, unless otherwise stated, were significant at the .05 alpha level. Pair-wise comparisons were Bonferroni corrected to the .05 level. Eta-squared and Cohen's d are the measures of effect size reported for all significant effects in the ANOVA and t-test analyses, respectively. A Geisser–Greenhouse correction was used for violations of the sphericity assumption of ANOVA.

Scoring

Two coders independently scored all of the short answer responses. Both scorers were blind to condition and coded all the responses for a given question together to increase consistency in scoring. Cohen's kappa was calculated to assess inter-rater reliability. Reliability was high ($\kappa = .87$), and the second author resolved the disagreements in scoring.

Text study time

Overall, subjects studied each historical text for an average of 195 seconds (or 3.3 minutes). A 3 (presentation order) 2 (instructions) ANOVA showed no significant effects (all F's < 1), indicating that subjects studied the texts equally, regardless of the order of presentation of materials or the instructions they were given.

Prior knowledge

Subjects reported having little prior knowledge about the materials used in the experiment. With respect to the film clips, they indicated that 82% of the time they had not seen the film before; for films that they had seen, the average number of times that they reported having watched each film (grand $M = 0.3$) did not differ by experimental condition (all F's < 1). For the topics covered in the texts, subjects indicated that they had some knowledge about the historical figures and events, but not very much: the average rating of their prior knowledge was low (grand $M = 2.1$ out of 5) and did not vary as a function of experimental condition (all F's < 1).

Engagement

Table 1 contains the mean rating given to the text and film clips on each of the four engagement questions. As the goal was to compute a single variable that captured the construct of engagement from the multiple measures, the relationship between the ratings for each of the engagement questions was assessed. As expected, subjects' ratings on the interest, involvement, and learn more questions were all positively related with each other and negatively related to ratings on the mind-wandering question (all p's $< .001$) for the texts and film clips, respectively. Thus, an overall engagement rating was computed for each trial on which a subject read a text or viewed a film clip by averaging the ratings on the four engagement questions (the mind-wandering question was reverse scored).

Overall, subjects found the film clips significantly more engaging than the texts [3.65 vs. 3.35; $t(35) = 3.61$, $SEM = .08$, $d = .46$]. Additional analyses were conducted to assess whether engagement varied as a function of either of the independent variables. For engagement ratings about the film clips, a 2×2 ANOVA revealed no significant main effects (both F's < 1) or interaction [$F(1, 34) = 2.05$, $MSE = .13$, $p = .16$]. That is, engagement in the films did not differ based on presentation order or instructions. For text engagement ratings, a 3×2 ANOVA revealed a significant main effect of order [$F(2, 68) = 5.08$, $MSE = .10$, $p = .01$, $\eta^2 = .13$]. Follow-up pair-wise comparisons indicated that subjects rated the texts as more engaging in the view–read condition relative to read–view condition [3.43 vs. 3.21; $t(35) = 3.56$, $SEM = .08$, $d = .32$]. However, there was no difference between the view–read and the read-only condition [3.43 vs. 3.40; $t < 1$], and the difference between the read-only condition and the read–view condition was marginally significant [3.40 vs. 3.21; $t(35) = 2.33$, $MSE = .08$, $p = .08$]. Thus, seeing the film before reading the associated text resulted in the highest engagement, but this was not different from the control read-only condition. However, neither the main effect of instruction nor the interaction was significant (F's < 1).

Final test: text/film consistent questions

On questions for the content that the films portrayed accurately and was corroborated by the associated texts, subjects answered roughly the same proportion of questions correctly across the various conditions (grand $M = .68$). This observation was confirmed by a 3 (presentation order) $\times$ 2 (instructions) ANOVA that revealed that neither the main effects nor the interaction was significant (all F's < 1).

In order to understand if subjects' engagement with the materials influenced performance on these questions, we conducted two linear regressions in which the text engagement ratings and the film engagement ratings were used separately to predict the proportion of correct responses on text/film consistent questions. For each regression, the dummy-coded nominal scale independent variables (instructions and presentation order) were entered into the model first, and then the text or film engagement variable was added.[1] Engagement in the film predicted the proportion of correct responses on the text/film consistent questions [$\beta = -.27$, $t(67) = 2.27$; $\Delta R^2 = .07$, $F(1, 68) = 5.14$]. That is, controlling for the presentation order and the instructions given, when subjects were more engaged in the film clips, they were less likely to correctly answer questions for which the content was correctly represented in the clip (and corroborated by the associated text).

Final test: text/film inconsistent questions

On questions where the film depicted inaccurate content (and the associated text provided the correct information), subjects could produce three different types of responses for the text/film inconsistent items: the correct response, the misinformation (from the film), or some other incorrect response. It is important to note that subjects were explicitly told to answer the final test questions on the basis of what they remembered from the texts, disregarding what they saw in the associated film clips. Figure 1 shows the proportion of misinformation responses on text/film inconsistent questions as a function of instructions and presentation order. A 3 (presentation order) $\times$ 2 (instructions) ANOVA showed that there was a significant main effect of presentation order [$F(2, 55) = 9.49$, $MSE = .07$, $\eta_p^2 = .21$]. As expected, subjects produced a greater proportion of misinformation responses when they had seen the films (view–read and read–view conditions) than when they had not seen the films (read-only condition) [.31 vs. .06; $t(35) = 4.07$, $SEM = .06$, $d = 1.04$, and .19 vs. .06; $t(35) = 3.05$, $SEM = .04$, $d = .71$, respectively]. Although there appeared to be a numerical difference between the read–view and view–read conditions, this difference was not significant [.31 vs. .19; $t(35) = 1.84$, $SEM = .07$, $p = .23$] (a follow-up 2×2 ANOVA yielded the same outcome). In addition, neither the main effect of instructions [$F(1, 34) = 2.34$, $MSE = .05$, $p = .14$] nor the interaction ($F < 1$) reached significance. That is, attempting to detect the misinformation in the films did not affect subjects' production of misinformation on the test.

[1] The read-only condition was not included in the presentation order variable for the regressions with film engagement because subjects did not view the films in this condition and thus did not make any ratings. The results for the independent variables in the first step of the regressions are not reported because they are redundant with the ANOVA analyses conducted above.

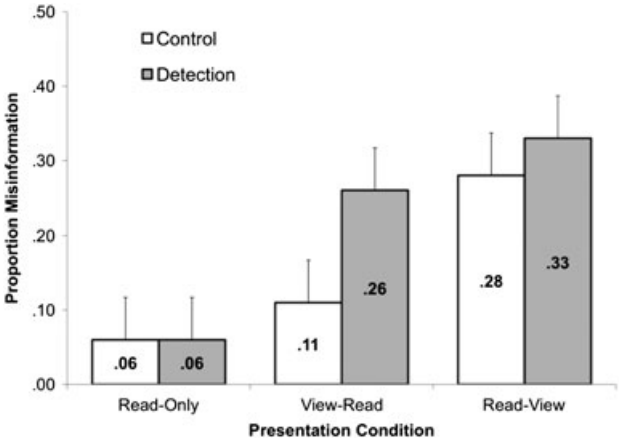

Figure 1. Proportion of misinformation responses on text/film inconsistent questions as a function of instructions and presentation order for Experiment 1. Error bars represent 95% confidence intervals

The pattern of results for the correct responses mirrored that of the misinformation responses with subjects producing a greater proportion of correct answers when they had not seen the films ($M = .64$) than when they had seen the films ($M = 0.53$) [$F(2, 68) = 4.28$, $MSE = .07$, $\eta^2 = .11$]. When we explored this further, pair-wise comparisons showed that a greater proportion of correct responses was produced in the read-only condition relative to the read–view [.64 vs. .46; $t(35) = 2.79$, $SEM = .06$, $d = .59$]. However, the view–read condition did not significantly differ from the read–view condition [.58 vs. .46; $t(35) = 2.13$, $SEM = .06$, $p = .11$] or the read-only condition [.64 vs. .58; $t < 1$]. The main effect of instructions was not significant [$F(1, 34) = 2.66$, $MSE = .13$, $p = .11$]. That is, engaging in the detection task did not lead subjects to answer more text/film inconsistent questions correctly. The interaction was not significant ($F < 1$).

Again, linear regressions were conducted in order to explore whether text engagement and film engagement ratings predicted the proportion of misinformation or correct responses on the text/film inconsistent questions. The regressions showed that engagement with the learning materials did not predict misinformation production or correct responding over and above instruction and presentation order (all F's > 1).

Discussion

The present results suggest that the detection task did not reduce the production of misinformation. Subjects who were explicitly asked to detect the inaccuracies in the films produced the misinformation that they saw in the films as much as subjects in the control condition; correspondingly, both groups retained similar amounts of correct information from the texts that contradicted misinformation in the films (i.e. text/film inconsistent items). However, exposure to film inaccuracies did reduce correct responses to the text/film inconsistent questions relative to when subjects did not view the films.

Contrary to our prediction, viewing the films did not boost the retention of the correct responses when the correct information was presented in the text and film. This result is surprising given that Butler et al. (2009) found that viewing the films and reading the text led to better retention of the

correct information relative to only reading the text. Two presentations of the correct information (text and film) would be expected to increase retention more than one presentation (text alone). One possible explanation is that the positive effect of films on the retention of correct information only emerges over time (see Experiment 2). The final test in the study by Butler et al. (2009) occurred after a 1-week delay, whereas our Experiment 1 had an immediate final test. On an immediate test, subjects may rely primarily on their memory for correct information in the texts (as instructed). However, they may forget the information from the texts over the course of a week, leading them to rely more on their memory for the films on a delayed test (even though they are instructed not to use this source). Thus, the benefit of film viewing may only emerge after a delay, which we will explore in Experiment 2.

With regard to engagement with the materials, as expected, the films were rated as more engaging than the texts. However, engagement with the films was not affected by whether subjects were asked to detect the inaccuracies or not. Although it might be expected that such an active task would increase engagement with the films, some studies have found that a focus on critical analysis does not change the degree of transportation into a narrative (e.g. Green, 2004), which is an indicator of engagement. Furthermore, viewing the films did not increase reported engagement with the associated texts relative to the read-only control condition. Perhaps it is only when looking back on the materials as a whole, which could encourage subjects to make a relative judgment, that the films are able to inspire greater interest in the texts (see Butler et al., 2009).

Finally, we conducted exploratory regression analyses by looking at the predictive power of subjects' engagement in the texts and the films. One significant relationship that emerged was that the more subjects were engaged in the films, the less likely they were to correctly answer the text/film consistent questions on the immediate final test. This finding is somewhat puzzling because greater engagement with the films might be expected to produce better retention of the correct information when the content was correctly represented in the film. One potential explanation is that subjects who were less engaged with the films relied more heavily on their memory of the text, which allowed them to perform well on an immediate final test. This possible explanation predicts that over longer delays, this relationship might reverse such that greater engagement would lead to increased retention (see Experiment 2). That is, regardless of their level of engagement in the film, subjects may forget correct information from the texts after a longer delay. Subjects who were more engaged in the films can fall back on their memory for the correct information in the films, whereas subjects who were less engaged in the films cannot because they do not remember the films as well.

EXPERIMENT 2

In a second experiment, we investigated some of the questions that were raised by the results of Experiment 1. First, detection success may play a role in the effectiveness of the detection task. However, it is unclear how often subjects were able to correctly identify the major inaccuracy in the film. If subjects in the detection condition are unable to catch the error while viewing the film, they may be more likely to acquire and reproduce it because of the additional cognitive processes devoted to evaluating the material. In order to examine this possibility in Experiment 2, we asked subjects to describe the major inaccuracy in each film. This detection check allowed us to gauge their effectiveness at correctly identifying the misinformation in the films. We also added a condition in which subjects were given feedback on the inaccuracy in the film after the detection check. Given that such tasks are often coupled with feedback from the teacher when used in the classroom, it is important for ecological validity to investigate the effects of feedback.

Second, given the finding in Experiment 1 that viewing the films did not increase the retention of information that was correctly presented in both the film and the text (cf. Butler et al., 2009), the retention interval was increased to 1 week. One possible reason for the lack of an effect may be the use of an immediate test in Experiment 1—the positive effect of viewing the films on the retention of correct information may only emerge over time. This change in the timing of the final test also provided the opportunity to examine whether the effects of the detection task and engagement in the material change over a longer retention interval.

Method

Subjects and design

Fifty-four Duke University undergraduates participated in this study for payment. The experiment had a 3 (presentation order: read-only, read–view, view–read) $\times$ 3 (instructions: detection, detection with feedback, control) mixed design. Presentation order was manipulated within-subjects, whereas instructions was manipulated between-subjects.

Materials and counterbalancing

The materials from Experiment 1 were used except for one change. As the five engagement questions were all highly inter-correlated, the number of questions was cut down to three (interest, involvement, and desire to learn more) in order to reduce the overall time taken to complete the experiment while still providing multiple measures of engagement.

Procedure

The procedure was the same as in Experiment 1 except for three major modifications. First, a detection check was added in order to gauge whether subjects were successful in detecting the historical inaccuracy in each film clip. After having both viewed the film clip and read the text, subjects were asked to describe the historical inaccuracy in the film. If they were unsure, they were told to make a plausible guess. Second, a 'detection with feedback' condition was added in which subjects were given the same instructions as those in the detection condition, but were also provided with feedback after completing the detection check. The feedback message identified the historical inaccuracy in the film and corrected it with the true information from the text. Subjects

were required to study the feedback for a minimum of 20 seconds, but they could continue studying for as long as they needed. Third, the delay between the initial learning phase and the final test was increased to 1 week. Finally, there was one minor change in which we instituted a minimum reading time (135 seconds) for each text to ensure that subjects spent enough time studying the material.

Results

Scoring

Two coders independently scored all of the short answer responses in the same manner as Experiment 1. Reliability was high for both the detection check responses ($\kappa = .90$) and the test responses ($\kappa = .81$). The second author resolved the disagreements.

Text study time

Overall, subjects studied the texts for an average of 202 seconds (or 3.4 minutes). Although study time did not differ much as a function of presentation order, there were differences among the instructions conditions. Subjects studied the texts for longer in the control condition relative to the detection condition and spent the least amount of time studying in the detection with feedback condition [$F(1, 52) = 3.73$, $MSE = 4407.55$, $\eta^2 = .13$]. One possible explanation for this effect is that when subjects knew that they could rely on the feedback to identify the inaccuracies, they spent less time studying the text whereas careful reading of the text was the only means of later identifying the misinformation content in the other conditions. Follow-up pair-wise comparisons showed that the control condition studied for longer than the detection with feedback condition [220 vs. 185 seconds; $t(34) = 3.07$, $SED = 11.35$, $d = .58$], but the differences between the control and detection conditions [220 vs. 200 seconds; $t(34) = 1.34$, $SED = 14.26$, $p = .42$] and the detection and detection with feedback conditions [200 vs. 185 seconds; $t(28) = 1.25$, $SED = 12.57$, $p = .68$] were not significant. Neither the main effect of order [$F(2, 104) = 2.24$, $MSE = 301.96$, $p = .11$] nor the interaction [$F(4,104) = 2.10$, $MSE = 301.96$, $p = .09$] was significant.

Engagement

Subjects' ratings on the three engagement questions were highly inter-correlated for the texts and film clips, respectively (all r's $> .68$, all p's $< .001$). Thus, these ratings were averaged to compute text and film engagement variables by using the same method as Experiment 1. Overall, subjects rated the film clips as more engaging than the texts [3.56 vs. 3.11; $t(53) = 7.97$, $SEM = .06$, $d = .70$]. For the film engagement ratings, a 2×2 ANOVA reveal no significant main effect of order [$F(1, 51) = 1.72$, $MSE = .22$, $p = .20$] or instructions [$F(2, 51) = 1.76$, $MSE = .89$, $p = .18$], and no interaction [$F(2, 51) = 1.93$, $MSE = .22$, $p = .16$]. That is, subjects' engagement in the films was not affected by the presentation order of the films versus texts or the instructions of the task. For the text engagement ratings, a 3×2 ANOVA showed a marginally significant main effect of instructions [$F(2, 51) = 2.64$, $MSE = 1.07$, $p = .08$], but no main effect of order or interaction (F's < 1).

Detection analyses

Overall, when explicitly asked to find the film inaccuracies (detection and detection with feedback conditions), subjects were relatively poor at detecting the misinformation in the films (grand $M = 0.35$), although they had been exposed to the accurate content in the texts associated with each film and thus had access to the correct knowledge. A 2 (instructions) $\times$ 2 (presentation order) ANOVA showed no main effect of instructions [$F(1, 30) = 1.41$, $MSE = .09$, $p = .24$], but the main effect of presentation order was marginally significant [$F(1, 34) = 3.62$, $MSE = .10$, $p = .07$], showing a trend that subjects were more successful at detecting the misinformation when they viewed the film then read the text ($M = 0.42$) than vice versa ($M = 0.28$). The interaction was not significant ($F < 1$).

Final test: text/film consistent questions

Figure 2 shows the proportion of correct responses on text/film consistent questions as a function of presentation order (the data are collapsed across instructions condition). Subjects answered more questions correctly when they had read the text and viewed the film than when they had only read the text as seen in a 3 (instructions) $\times$ 3 (presentation order) ANOVA, which confirmed these observations by revealing a main effect of presentation order [$F(2, 102) = 12.68$, $MSE = .06$, $\eta^2 = .19$]. Follow-up pair-wise comparisons showed that both the view–read and read–view conditions produced a significantly greater proportion of correct responses relative to the read-only condition [.53 vs. .32; $t(53) = 4.13$, $SEM = .05$, $d = .74$, and .53 vs. .32; $t(53) = 4.56$, $SEM = .05$, $d = .74$, respectively]. The view–read and read–view conditions were not significantly different [.53 vs. .53; $t < 1$]. However, the instructions did not seem to affect performance—all three groups produced about the same proportion of correct responses. Neither the main effect of instructions nor the interaction was significant (F's < 1).

As in Experiment 1, regression analyses were conducted in order to investigate whether engagement in the learning process predicted final test performance after accounting

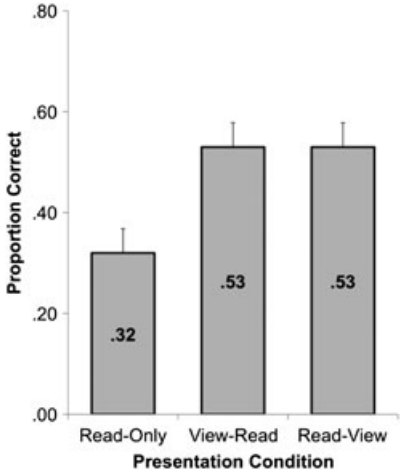

Figure 2. Proportion of correct responses on text/film consistent questions as a function of presentation order for Experiment 2. Data are collapsed across instructions condition. Error bars represent 95% confidence intervals

for the effects of the independent variables. The same two regressions from Experiment 1 were conducted with the addition of the detection with feedback condition to the instructions variable. Table 2 shows the total R^2 values for Model 1 (the main independent variables) along with the total R^2 values, the change in R^2 values, and the standardized beta weights from entering Model 2 (engagement ratings) into the regression. The text and film engagement variables were each significant predictors of correct responses on the text/film consistent questions. As engagement in the film or text increased, the proportion of correct responses on text/film consistent also increased.

Final test: text/film inconsistent questions

Figure 3 shows the proportion of misinformation responses on text/film inconsistent questions as a function of instructions and presentation order. Even when subjects had only read the texts and not seen the films, they still produced some misinformation (subjects were told to put a plausible guess if they were unsure and it was possible to guess some of the pieces of misinformation). Relative to this low baseline of misinformation production, when subjects viewed the films, the control and detection conditions produced a substantially greater proportion of misinformation responses although all subjects were told to answer the questions on the basis of the text material and not what they had seen in the films. This pattern held regardless of presentation order. However, when subjects received feedback after the detection check, they produced almost no misinformation at all.

As seen in a 3×3 ANOVA, a significant main effect of instructions [$F(2, 51) = 12.50$, $MSE = .05$, $\eta^2 = .33$] revealed that both the detection and control conditions produced significantly greater proportions of misinformation than the detection with feedback condition [.27 vs. .08; $t(25) = 4.44$, $SED = .04$, $d = 1.51$, and .27 vs. .08; $t(25) = 4.44$, $SED = .04$, $d = 1.76$, respectively] . However, the detection and control conditions did not differ [.27 vs. .27; $t < 1$]. For the significant main effect of presentation order [$F(2, 102) = 6.12$, $MSE = .05$, $p = .003$, $\eta^2 = .10$], the read–view condition produced a greater amount of misinformation than the read-only condition [.27 vs. .12; $t(53) = 3.32$, $SEM = .05$, $d = .62$], but the difference between the view–read and read-only conditions

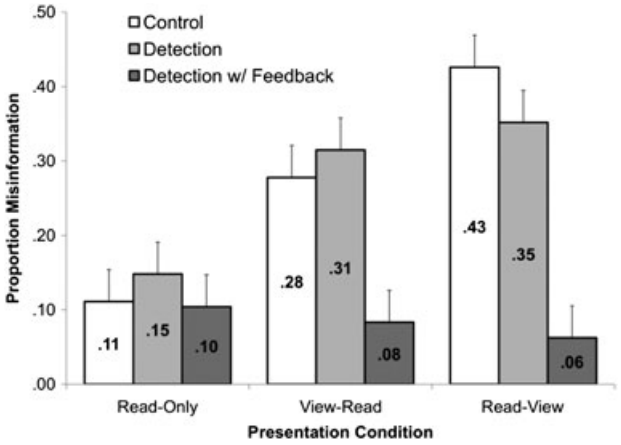

Figure 3. Proportion of misinformation responses on text/film inconsistent questions as a function of instructions and presentation order for Experiment 2. Error bars represent 95% confidence intervals

was only marginally significant [.22 vs. .12; $t(53) = 2.26$, $SEM = .04$, $p = .07$]. There was no significant difference between the read–view and view–read conditions [.27 vs. .22; $t < 1$]. In order to better understand the nature of the significant interaction [$F(4,102) = 3.15$, $MSE = .05$, $\eta^2 = .10$], we conducted a follow-up 3×2 ANOVA with just the view–read and read–view conditions (the read-only condition was excluded). This analysis revealed a significant main effect of instructions [$F(2, 51) = 14.38$, $MSE = .06$, $\eta^2 = .36$], but neither the main effect of presentation order [$F(1, 51) = 1.14$, $MSE = .06$, $p = .29$] nor the interaction was significant ($F < 1$). Thus, the numerical difference between the two order conditions was not significant.

The proportion of correct responses produced on text/film inconsistent questions mirrored the pattern of misinformation results (see Figure 4). A 3×2 ANOVA showed significant main effects of instructions [$F(2, 51) = 8.96$, $MSE = .11$, $\eta^2 = .26$] and presentation order [$F(2, 102) = 10.29$, $MSE = .06$, $\eta^2 = .13$], and a significant interaction [$F(4,102) = 8.08$, $MSE = .06$, $\eta^2 = .21$]. With respect to instructions, pair-wise comparisons showed that the detection with feedback condition produced significantly greater proportion of correct responses than both the detection and control conditions [.58 vs. .34; $t(25) = 3.35$, $SED = .07$, $d = 1.15$, and .58 vs. .35;

Table 2. Results of the six regression analyses conducted on the data from Experiment 2

	Model 1 (main IVs)	Model 2 (engagement ratings)		
Dependent variable	R^2	R^2	ΔR^2	β
Text engagement				
Text/film consistent—correct	.11***	.18	.08***	.29
Text/film inconsistent—correct	.19***	.28	.09***	.31
Text/film inconsistent—misinformation	.19***	.19	.00	-.02
Film engagement				
Text/film consistent—correct	.01	.07	.06**	.26
Text/film inconsistent—correct	.29***	.32	.03*	.18
Text/film inconsistent—misinformation	.23***	.23	.00	-.04

Note: Statistics are reported as a function of predictor variable (text engagement or film engagement) and dependent variable. The table shows total R^2 values for Model 1 (the main independent variables) along with the total R^2 values, the change in R^2 values, and the standardized beta weights from entering Model 2 (engagement ratings) into the regression.
*$p < .05$;
**$p < .01$;
***$p < .001$.

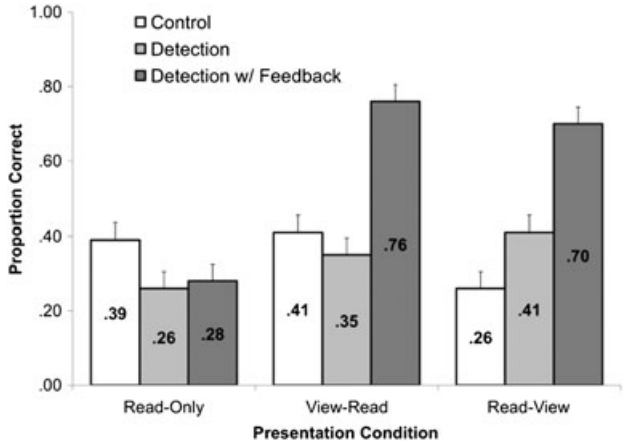

Figure 4. Proportion of correct responses on text/film inconsistent questions as a function of instructions and presentation order for Experiment 2. Error bars represent 95% confidence intervals

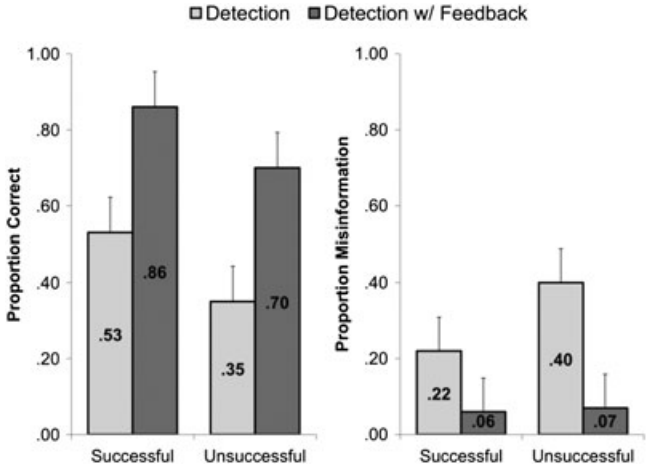

Figure 5. Proportion of correct responses (left panel) and misinformation responses (right panel) on text/film inconsistent questions as a function of instructions and detection success (successful vs. unsuccessful) for Experiment 2. Error bars represent 95% confidence intervals

$t(25) = 3.93$, $SED = .06$, $d = 1.36$, respectively]. The detection and control conditions were not significantly different [.35 vs. .34; $t < 1$]. For presentation order, pair-wise comparisons revealed that both the view–read and read–view condition produced significantly greater proportion of correct responses than the read-only condition [.51 vs. .31; $t(53) = 2.13$, $SEM = .06$, $d = .69$, and .46 vs. .31; $t(53) = 2.13$, $SEM = .05$, $d = .52$, respectively]. However, there was no significant difference between the two order conditions [.51 vs. .46; $t(53) = 1.16$, $SEM = .04$, $p = .73$].

Finally, we explored whether engagement predicted performance on the text/film inconsistent questions. The same four regressions from Experiment 1 were conducted with the addition of the detection with feedback condition to the instructions variable (see Table 2). The text and film engagement variables were each significant predictors of correct responses; however, neither film engagement nor text engagement predicted the production of misinformation.

Conditional analyses: effects of detection on final test performance

The conditional analysis were limited to subjects in the detection and detection with feedback conditions. The left panel of Figure 5 shows the proportion of correct responses on the text/film inconsistent questions as function of detection success (i.e. whether the misinformation was successfully detected) and the instructions condition. As expected, subjects produced a greater proportion of correct responses when they successfully detected the misinformation ($M = 0.70$) than when they failed to detect it ($M = 0.53$), confirmed in a 2×2 ANOVA revealing a significant main effect of instructions [$F(1, 30) = 12.79$, $MSE = .14$, $\eta^2 = .30$]. In addition, they answered more questions correctly when they received feedback after the detection check ($M = .78$) than when they did not receive feedback ($M = .44$), [$F(1, 30) = 6.65$, $MSE = .07$, $\eta^2 = .18$], but the interaction was not significant ($F < 1$).

The right panel of Figure 5 shows the proportion of misinformation responses on the text/film inconsistent questions as function of detection success and instructions condition. When subjects received feedback after the detection check, they were less likely to produce misinformation on the final test ($M = 0.07$) relative to when they did not receive feedback ($M = 0.31$) [$F(1, 30) = 15.23$, $MSE = .06$, $\eta^2 = .34$]. Somewhat surprisingly, they were only slightly less likely to produce the misinformation when they successfully detected the misinformation ($M = 0.14$) than when they failed to detect the misinformation ($M = 0.24$). However, it is important to note that this effect mainly applies to the detection condition because very little misinformation was produced in the detection with feedback condition. Neither the main effect of detection [$F(1, 30) = 2.20$, $MSE = .07$, $p = .15$] nor the interaction [$F(1, 30) = 1.94$, $MSE = .07$, $p = .17$] was significant.

DISCUSSION

Once again, the detection task did not reduce the production of misinformation, replicating the finding in Experiment 1. In fact, the amount of misinformation from the films produced on the final test after a week's delay was even more pronounced. One possible explanation for the ineffectiveness of this task in preventing the reproduction of misinformation in Experiment 1 was that subjects were relatively unsuccessful at identifying the inaccurate content. The results of Experiment 2 support this explanation—when asked to detect inaccurate content in films, subjects generally failed to identify the erroneous information, succeeding only 35% of the time. Moreover, even when subjects were able to detect the misinformation while viewing the films, they still reproduced some of that inaccurate content on the final test. However, when provided with feedback after the detection checks, subjects almost never produced the misinformation and they were more likely to produce the correct answers for the same questions.

The results of Experiment 2 also showed that a longer retention interval was critical to observing some of the effects of viewing the films. After a 1-week delay, having viewed the films increased the production of correct answers for questions about information that was depicted accurately in both the films and texts. This finding, which replicates

Butler et al. (2009), suggests that the lack of an effect in Experiment 1 was due to the immediate final test. As described above, one possibility is that subjects tend to rely on their memory for the texts on an immediate test, but over time they forget the information in the text and instead rely on their memory of the film (despite instructions against doing so). The changes in level of performance between Experiments 1 and 2 provide additional evidence to support this explanation. When subjects had only read the texts, the proportion of correct responses was substantially higher on the immediate test (.68) relative to the delayed test (.32), indicating that substantial forgetting of the text occurred over the course of a week. In contrast, when subjects had read the text and viewed the film, the decrease in correct responding between the immediate test (.68) and the delayed test (.53) was attenuated.

Another result that emerged with a longer retention interval was that the ratings of text and film engagement both predicted correct answers on text/film consistent and text/film inconsistent questions on the final test a week later. Although the results of Experiment 1 showed that engagement in the films was negatively related to memory for correct information on an immediate test, this relationship reversed over a longer retention interval. As described above (see discussion of Experiment 1), one way to explain the reversal from Experiment 1 to Experiment 2 is that subjects who are less engaged with the films may rely more heavily on their memory of the texts than subjects who are more engaged with the films. Greater reliance on the texts may be beneficial on an immediate test, thus resulting in a negative relationship between film engagement and the production of correct information. However, when the correct information in the text is forgotten over a longer delay, subjects who are more engaged in the films may have an advantage because they can rely on their memory for the correct information in the films. If so, the expected result would be a positive relationship between film engagement and the proportion of correct information produced on a delayed test, exactly as shown in Experiment 2.

GENERAL DISCUSSION

Together, the present experiments provide a fuller picture of how watching popular films affects learning about history, including the influence of student engagement with these materials. Exposing subjects to the historical inaccuracies in the films increased the acquisition of false knowledge (see also Butler et al., 2009). Critically, the negative effects of exposing students to the misinformation in popular films were not reduced by the pedagogical technique of asking students to detect the inaccuracies as they watched the film. In addition, this task did not promote greater engagement with the materials (see, Green, 2004). However, when feedback was given to subjects after the detection exercise, the acquisition and subsequent production of misinformation was almost completely eliminated. Finally, it is interesting to note that the effects of viewing the films, both positive and negative, seem to increase over time, suggesting that films may have a long-lasting influence on learning. Further research is clearly needed to investigate how the effects of viewing popular films on learning change as a function of retention interval. We turn now to discussing each of the key results in turn.

The current results show that the detection task was no better than passively watching the films for preventing the acquisition and retention of misinformation. Furthermore, this task did not promote the acquisition of the correct information from the texts even when the films corroborated their content. One could attribute the ineffectiveness of this task to subjects' difficulty with correctly identifying the inaccurate content in the films (Lee, 1982; Stoddard & Marcus, 2010). Yet, even when they were able to correctly identify the inaccuracies while viewing the films, they still later reproduced some of that misinformation (see also Eslick et al., 2011; Fazio & Marsh, 2008). One possible explanation for this finding is that subjects may have misattributed the inaccurate content as having appeared in both the text and the film (see Butler et al., 2009). The effect of films may be so powerful that the memory for the information in the films remains highly accessible after a week relative to the memory of the corresponding information in the text. If correct information learned through successful detection during initial learning was forgotten or rendered less accessible, then subjects may have fallen back on the inaccuracies in the film and erroneously attributed them to the texts.

While fundamentally interesting for the same reasons as other forms of narrative (Silvia & Berg, 2011), films may be special for several reasons: they are more complex but also more comprehensible, they contain visual and auditory information that are rich in detail and vividness, and they have narrative structure (Green, Brock, & Kaufman, 2004). The very reasons that make films more engaging than texts, as seen in both of our experiments, may be the ones that encourage the acquisition of content from them, regardless of whether this content is accurate or inaccurate. Suspension of disbelief is necessary for processing fictional narratives because it allows people to engage mentally with the story in a more immersive way, typically for enjoyment purposes (Green et al., 2006; Rubin, 1994). This absorption in the narrative makes one less motivated to constrain one's attention to critically analyze the accuracy of the propositions put forth within the story (Green & Brock, 2000; Green et al., 2006). At the same time, one may have fewer cognitive resources available for critical thinking when transported into a narrative, which may result in detecting fewer inaccuracies (Green & Brock, 2000).

Although the pedagogical task of detecting errors in popular films may have drawbacks, the powerful transporting quality of films does promote learning. After a week, in the present work, neither text nor film engagement was predictive of the production of misinformation answers on the final test. Instead, greater engagement predicted correct answers on the final test for both questions tapping information that was the consistently accurate in both the films and the texts as well as those for which the films portrayed inaccurate historical information. Thus, although exposing students to the inaccurate information in films may be problematic for learning, the engagement that they have with films can still encourage the learning of accurate historical content (cf. Green, Brock, & Kaufman, 2004: Silvia, 2006, 2008).

On another positive note, the results of Experiment 2 showed that feedback ameliorated the negative effects of exposing subjects to misinformation in the films. Although some studies have found that informing people about misinformation after it is presented is relatively ineffective (Eakin et al., 2003; Greene et al., 1982), subjects in the present study rarely produced the misinformation if they were given feedback after viewing the films. Broadly speaking, research on feedback has shown that it is a powerful tool for correcting errors in memory (e.g. Butterfield & Metcalfe, 2001; for review see Shute, 2008). Although feedback was more effective at promoting the retention of the correct information following the successful detection of an inaccuracy, it essentially eliminated the reproduction of misinformation regardless of whether the inaccuracy was detected. Our findings fit nicely in the previous research in which feedback has been shown to correct errors and help maintain correct responses (e.g. Butler, Karpicke, & Roediger, 2008).

EDUCATIONAL IMPLICATIONS

In pursuing the goal of fostering critical thinking and media literacy, educators must be aware that simply asking students to detect the historical inaccuracies while viewing popular films can actually lead to the acquisition of false knowledge, much like passively watching the films. The key to ensuring the efficacy of this pedagogical exercise is feedback. Coupled with this monitoring task, educators must explicitly tell students what was wrong about a film's portrayal of historical events in order for students to remember what is historically true in the long run. Films can be useful tools for enhancing learning in the classroom when used properly because they encourage student engagement with the material and increase the long-term retention of accurate historical information.

ACKNOWLEDGEMENTS

This research was supported by a Collaborative Activity Award from the James S. McDonnell Foundation's 21st Century Science Initiative in Bridging Brain, Mind and Behavior. The first author is supported by a National Science Foundation Graduate Research Fellowship. The authors would like to thank Suedabeh Walker for her help with creating the materials, programming, collecting data, and coding.

REFERENCES

Berumen, M. L. (2008). "Life" in movies. *The Science Teacher*, *75*, 26–31.

Brown, A. P. (1992). Amadeus and Mozart: Setting the record straight. *The American Scholar*, *61*, 49–66.

Burgoyne, R. (1997). *Film nation: Hollywood looks at U.S. history*. Minneapolis: University of Minnesota Press.

Butler, A. C., Karpicke, J. D., & Roediger, H. L., III. (2008). Correcting a metacognitive error: Feedback enhances retention of low confidence correct responses. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 918–928.

Butler, A. C., Zaromb, F. M., Lyle, K. B., & Roediger, H. L., III (2009). Using popular films to enhance classroom learning: The good, the bad, and the interesting. *Psychological Science*, *20*, 1161–1168.

Butterfield, B., & Metcalfe, J. (2001). Errors committed with high confidence are hypercorrected. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 1491–1494.

Carnes, M. C. (1995). *Past imperfect: History according to the movies*. New York: Henry Holt and Co.

Eakin, D., Schreiber, T., & Sergent-Marshall, S. (2003). Misinformation effects in eyewitness memory: The presence and absence of memory impairment as a function of warning and misinformation accessibility. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 813–825.

Eslick, A. N., Fazio, L. K., & Marsh, E. J. (2011). Ironic effects of drawing attention to story errors. *Memory*, *19*, 184–191.

Fazio, L. K., & Marsh, E. J. (2008). Slowing presentation speed increases illusion of knowledge. *Psychonomic Bulletin & Review*, *15*, 180–185.

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.

Gilbert, D. T., Tafarodi, R. W., & Malone, P. S. (1993). You can't not believe everything you read. *Journal of Personality and Social Psychology*, *65*, 221–233.

Green, M. C. (2004). Transportation into narrative worlds. The role of prior knowledge and perceived realism. *Discourse Processes*, *38*, 247–266.

Green, M. C., & Brock, T. C. (2000). The role of transportation in the persuasiveness of public narratives. *Journal of Personality and Social Psychology*, *79*, 701–721.

Green, M. C., Brock, T. C., & Kaufman, G. F. (2004). Understanding media enjoyment: The role of transportation into narrative worlds. *Communication Theory*, *14*, 311–327.

Green, M. C., Garst, J., & Brock, T. C. (2004). The power of fiction: Determinants and boundaries. In *The psychology of entertainment media: Blurring the lines between entertainment and persuasion*. Mahwah, NJ: Lawren Erlbaum Publishers.

Green, M. C., Garst, J., Brock, T. C., & Chung, S. (2006). Fact versus fiction labeling: Persuasion parity despite heightened scrutiny of fact. *Media Psychology*, *8*, 267–285.

Greene, E., Flynn, M., & Loftus, E. (1982). Inducing resistance to misleading information. *Journal of Verbal Learning and Verbal Behavior*, *21*, 207–219.

Hobbs, R. (2006). Non-optimal uses of video in the classroom. *Learning, Media and Technology*, *31*, 35–50.

Hunt, R. R., & Einstein, G. O. (1981). Relational and item-specific information in memory. *Journal of Verbal Learning and Verbal Behavior*, *20*, 497–514.

Jarvis, B. G. (2008a). *DirectRT (Version 2008.1.0.13) [Computer software]*. New York: Empirisoft Corporation.

Jarvis, B. G. (2008b). *MediaLab (Version 2008.1.33) [Computer software]*. New York: Empirisoft Corporation.

Jones, T. B. (1989). The good, the (not so) bad, and the ugly: Using 'historical' movies with adult learners in the American history classroom. *Film & History*, *19*, 74–86.

Lee, P. A. (1982). The frozen moment: History and some uses of film. *Film & History*, *12*, 35–43.

Lindsay, D. S., & Johnson, M. K. (1989). The eyewitness suggestibility effect and memory for source. *Memory & Cognition*, *17*, 349–358.

Loftus, E. F. (1979). Reactions to blatantly contradictory information. *Memory & Cognition*, *7*, 368–374.

Marcus, A. S.. (2005). "It is as it was": Feature film in the history classroom. *The Social Studies*, 61–67.

Marcus, A. S., & Stoddard, J. D. (2007). Tinsel town as teacher: Hollywood film in the high school classroom. *The History Teacher*, *40*, 303–330.

Marsh, E. J., & Fazio, L. K. (2006). Learning errors from fiction: Difficulties in reducing reliance on fictional stories. *Memory & Cognition*, *34*, 1140–1149.

Michel, E., Roebers, C. M., & Schneider, W. (2007). Educational films in the classroom: Increasing the benefit. *Learning and Instruction*, *17*, 172–183.

Niemi, R. (2006). *History in the media: Film and television*. Santa Barbara, CA: ABC-CLIO.

Prentice, D. A., & Gerrig, R. J. (1999) Exploring the boundary between fiction and reality. In *Dual process theories in social psychology* (pp. 529–546). New York, NY: Gifford Press.

Rubin, A. M. (1994). Media uses and effects: A uses and gratifications perspective. In *Media effects: Advances in theory and research* (pp. 417–436). Hillsdale, NJ: Erlhaum.

Shute, V. (2008). Focus on formative feedback. *Review of Educational Research*, *78*, 153–189.

Silvia, P. J. (2006). *Exploring the psychology of interest*. New York: Oxford University Press.

Silvia, P. J. (2008). Interest—the curious emotion. *Current Directions in Psychological Science*, *17*, 57–60.

Silvia, P. J., & Berg, C. (2011). Finding movies interesting: How appraisals and expertise influence the aesthetic experience of film. *Empirical Studies of the Arts*, *29*, 73–88.

Silvia, P. J., & Kashdan, T. B. (2009). Interesting things and curious people: Exploration and engagement as transient states and enduring strengths. *Social and Personality Psychology Compass*, *3*, 785–797.

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experiment Psychology: Human Learning and Memory*, *4*, 592–604.

Smilanich, B., & Lafreniere, N. (2010). Reel teaching = real learning: Motivating reluctant students through film studies. *Journal of Adolescent & Adult Literacy*, *53*, 604–606.

Stoddard, J. D., & Marcus, A. S. (2010). More than 'showing what happened': Exploring the potential of teaching history with film. *The High School Journal*, *93*, 83–90.

Toplin, R. B. (1996). *History by Hollywood: The use and abuse of the American past*. Chicago: University of Illinois Press.