Comprehension of Short Stories: Effects of Task Instructions on Literary Interpretation

Kathryn S. McCarthy & Susan R. Goldman

University of Illinois at Chicago

Accepted September 15th, 2014 to *Discourse Processes*

Abstract

An important purpose of reading literature is to move beyond the literal text to construct an interpretation of what the text conveys about the human condition and nature of the world (Langer, 2010; Lee, 2007; 2011). In two experiments, college students with no prior training in literary analysis read a short story and responded to one of four task instructions (plot, ambiguous, argument, theme) that were designed to bias either an interpretive or literal stance towards the text. Results indicated that the argument and theme instructions were more likely to lead to essays with more interpretive inferences than plot and ambiguous instructions. Results indicate that stance affected the kinds of inferences that were generated during reading. Implications for expanding current models of text comprehension are discussed.

Comprehension of Short Stories: Effects of Task Instructions on Literary Interpretation

An important purpose of reading literature is to construct a nonliteral interpretation of the text that speaks to a moral, message, or some greater meaning (Langer, 2010; Lee, 2011; Vipond & Hunt, 1984). Of course, the act of reading a text that has been deemed literary does not guarantee that the reader will engage in interpretive behavior. For example, George Orwell's *Animal Farm* can be read from a purely literal perspective. A literal reading of this book might result in a representation of how the animals of Old Manor Farm drove out the oppressive human owners, rebuilt the farm, and the struggle for control that ensued (Hamilton, 2011). In contrast, an interpretive reading of this book might result in a representation of a satire of the rise of communist Russia, with symbols, motifs, and themes about educating the poor, class stratification, and the abuse of power (e.g. Hamilton, 2011; Letemendia, 1992). Thus, in investigating literary reading, it is useful to distinguish between representations that are literal and limited to the world of the story and those that encompass interpretations that reach beyond the world of the story. The former result from readers adopting a *literal stance* toward the story, whereas the latter from readers adopting an *interpretive stance.* Adopting an interpretive stance does not replace a literal stance. Rather, an interpretive stance builds on an understanding of plot, character, and other elements of the "literal" story. However, evidence reviewed below indicates that many readers adopt only a literal stance and never move to an interpretive one (Graves & Frederiksen, 1991; Zeitz, 1994). Assuming an interpretive stance implies that readers invoke different criteria or standards of coherence (van den Broek, Lorch, Linderholm, & Gustafson, 2001), resulting in the generation of different kinds of inferences to satisfy these criteria (For further discussion, see Goldman, McCarthy, & Burkett, in press).

Many have noted the dearth of empirical research on the cognitive processes related to how, why, when, and even whether readers adopt an interpretive stance and construct these interpretive inferences (See Vol. 1 of *Scientific Study of Literature*). The current research examined the effect of reading purpose or goal, as induced by task instructions, on the propensity of college students (with no formal training in literary analysis) to adopt an interpretive stance when reading literature.

Interpretation and Current Models of Text Comprehension

Most text comprehension researchers acknowledge the utility of thinking about comprehension as involving the creation of at least three "levels" of representation: the *surface code, textbase,* and *situation model.* Originally proposed by van Dijk and Kintsch (1983) and subsequently further refined (see for discussion Goldman, 1997 and Goldman, Golden, & van den Broek, 2007), the surface code captures the specific words, syntax and rhetorical structure of the presented text and is typically relatively short-lived. The textbase expresses the meaning explicitly stated in the text and connections among the words, concepts, and propositions explicitly stated in the text. The situation model representation draws on prior knowledge relevant to the events and protagonists in the text to embellish the textbase representation and create a more complete and coherent model of the situation depicted in the text. Such a representation requires the generation of elaborative and causal inferences. The presence of a situation model is typically taken as evidence of understanding the text (Goldman, 1997).

Although certain types of elaborative information are likely to be represented in the situation model, it is less clear whether the author's intended message(s) or thematic interpretations of the stories are also represented within this level. There is the possibility that understanding plot-level information at the situation model level may be necessary for successful

comprehension of a literary work, but it may not be sufficient for representing interpretations of the point or message about the human condition. The representation of interpretive inferences may constitute an additional representational level. Indeed, some have proposed a *communicative* model that exists beyond the situation model to take into account readers' efforts to ascertain the communicative intent of the author (Graesser & McNamara, 2010; Graesser, Singer, & Trabasso, 1994; van Dijk & Kintsch, 1983). However, empirical research establishing the validity of "levels" or "models" beyond the situation model is scant. Efforts to extend beyond situation models have been most prevalent for work on comprehension and learning from multiple sources of information in history or science where information about the source (metadata) is represented as a node linked to, but outside of, the situation model. Sourcing information includes, for example, who authored the piece, when, for what purpose, and where it was published (See for elaboration, Goldman, 2004; Perfetti, Rouet, & Britt, 1999; Rouet, 2006; Rouet & Britt, 2011; Wiley, Goldman, Graesser, Sanchez, Ash, & Hemmerich, 2009).

Processing and Memory for Text

Much of the cognitively-oriented discourse processing research has focused on how readers establish and maintain coherence of the narrative at the textbase and situation model levels and the implications of such processing for memory (e.g. Graesser, Millis, & Zwaan, 1997; Graesser, Singer, & Trabasso, 1994; McNamara & Magliano, 2009). These studies are a far cry from literary reading. The research has tended to use relatively short, often researcher-generated texts that are highly controlled in order to isolate and manipulate one or two variables per experiment. Dubbed *textoids*, these narratives are often devoid of the rich and complex rhetorical devices found in *authentic* texts -- texts that are written for purposes other than laboratory experiments (Rapp, Komeda, & Hinze, 2011). Recently, researchers have begun

analyzing reading behaviors using longer, authentic works (e.g. McNerny, Goodwin, & Radvansky, 2011). However, the comprehension differences attributed to the manipulated variable(s) are still indexed by dependent variables such as reading time, free or prompted recall, and sentence and inference-verification tasks. Such assessments capture reproductive, and in some cases reconstructive, memory judged against what was presented in the text or spatial and causal inferences that do not go beyond the story world.

Modern literary theorists assume that authentic literary texts can support any number of interpretations. Indeed, some contend that part of literary reading is entertaining the possibility of multiple meanings that can be supported with evidence from the text and principles that govern the human condition such as morals and ethics (Langer, 2010; Lee, 2007, 2011; Rabinowitz, 1987; Schraw, 1997). Tapping into readers' interpretations of literature thus requires more open-ended response tasks that are done while having access to the literary works - quite a different situation from the typical laboratory experimental task using textoids and dependent variables that address memory for what was read.

Interpretive Processing of Literature

Evidence for complex interpretive reasoning during literary reading has been found using methodologies that tap reasoning during reading, specifically through think-aloud tasks. However, these interpretive behaviors tend to be limited to literary experts reading literary works (Graves & Frederiksen, 1991; Zeitz, 1994). For example, Graves and Frederiksen (1991) asked readers to read a passage from *The Color Purple* (Walker, 1982) and to provide a verbal description of the passage while reading it, commenting on its content and style. They found that their expert literary readers (University English department senior faculty) relied on prior knowledge to produce complex interpretive messages about the human condition. The literary

novices (sophomore students at the same university), on the other hand, produced restatements or summaries of the text that were factually correct with respect to the text, but contained little interpretive reasoning. In other work, Zeitz (1994) found that this type of interpretive processing by literary experts was not evident when they were processing non-literary works (e.g. an expository science text). These findings regarding literary experts are consistent with expert-novice contrasts in other disciplines that indicate that experts have more (and more organized) prior knowledge that results in different processing and representations of the content as compared to their novice counterparts (e.g. Chi, Feltovich, & Glaser, 1981).

This disparity between experts and novices is particularly concerning in light of achievement standards that call for classroom instruction that focuses on this type of interpretive work (Council of Chief State School Officers (CCSSO), 2010; National Assessment of Educational Progress (NAEP) 2008). In the current instructional climate, students rarely, if ever, spontaneously demonstrate this kind of reasoning behavior on their own (Earthman, 1992; Graves & Frederiksen, 1991; Peskin, 1998; Zeitz, 1994). In one investigation, Vipond and Hunt (1984) asked more than one hundred undergraduates questions about a short story (Updike's *A & P*) they had read several times. Even though some of these questions were intentionally leading, only 5% of the students acknowledged a possible interpretation or author purpose. These findings highlight the difficulty of getting novice readers to engage in literary interpretation.

In addition to the prior knowledge specific to the text (such as knowledge about the author or the historical context in which it was written), the knowledge bases of literary experts have been postulated to include more general knowledge that that texts have both a literal meaning and a subtext (Schraw, 1997) and that authors leave clues to the subtext in the surface text by relying on a set of conventions for use of language and rhetorical devices that the authors

expect their readers to be aware of (Rabinowitz, 1987). Additionally, one hallmark of literary expertise may be the awareness that literature often serves as a gateway into understanding the human condition (Langer, 2010; Lee, 2011). Thus, it may be that these characteristics of literary experts' knowledge of literature and literary conventions are responsible for their adopting an interpretive stance toward literature. Non-experts may have only some awareness of these goals and features of literature. Indeed, Claassen (2012) found that novice readers activated information about the author's possible interpretive meaning during reading, but in a think-aloud protocol, less than 5% of their statements were related to author intent, indicating that even though the information was activated, it was not being selected as relevant or important to the given task. These findings suggests that novices may possess the ability to engage in some level of interpretive reasoning, but they may struggle in distinguishing when adopting an interpretive stance is appropriate.

In the current work, we developed a task instruction manipulation designed to encourage novice readers to engage in more expert-like reasoning when reading literary works. We were interested in whether task instructions that suggested the appropriateness of literal versus interpretive stance would change non-expert readers' representation of short stories. The task instruction manipulation was motivated by prior work in text comprehension and problem solving indicating that task instructions moderate task interpretation, comprehension and reasoning processes, and the outcome of those processes as reflected in subsequent solutions and responses (e.g. Kotovsky, Hayes, & Simon, 1985; Wiley & Voss, 1999; Wiley, et al. 2009). Specifically, in a text comprehension situation, van den Broek, Lorch, Linderholm, & Gustafson (2001) presented students with several science texts and asked them either to "read for leisure" or to "read for study." Participants in the study condition produced more causal and explanatory

inferences, whereas participants provided with the leisure goal instruction offered more evaluations and associations. Similarly, using an authentic narrative text, Mason, Scirica, and Salvi (2006) found that a specific task instruction that asked readers to comment about the meaning, the author's intentions, and related personal experiences elicited more sophisticated interpretations of the text than did general instructions to comment on what readers had just read.

In the two experiments reported in this paper, we manipulated the instructions for an essay-writing task to test the hypothesis that tasks that were explicit about an interpretive level would make it more likely that readers' responses would reflect interpretive processing as compared to task instructions that were less likely to cue interpretation. The *plot* task instruction condition asked what happened in the story and was expected to yield essays that primarily related information at textbase and situation model levels. The *argument* instruction condition was intended to be a strong cue to adopt an interpretive stance. It explicitly asked readers to select one of two interpretations of the text and explain their choice. This condition was designed to make explicit the level of reasoning necessary for the task by offering two choices. It could be argued that choosing between two interpretations provides information that readers otherwise would not have constructed from the texts and that this that this prompt really evokes interpretive recognition processes. To evoke construction of interpretations, a third task instruction condition, *theme*, was included. A fourth condition left the level of interpretation relatively open by using a task instruction that asked what the story was about, which we labeled the *ambiguous* task instruction. In all conditions, the instructions indicated that evidence from the story should be used to support claims. The essays were analyzed for presence of literal and interpretive information.

We predicted that the plot instructions would yield essays that were largely literal retells of the story; that the argument and theme task instructions would yield essays that included more interpretive information; and that the ambiguous task instruction would yield essays that were a mix of retelling and interpretation. Experiment 1explored effects of these task instructions on interpretation for the short story *Harrison Bergeron* (Vonnegut, 1968). Experiment 2 was a replication using a second short story, *The Elephant* (Mrozek, 1972).

Experiment 1

Method

Participants

114 students (44 male) from an introductory psychology course at an urban university in the midwest received course credit for their participation. Mean age for participants was 19.98 (*SD* = 2.63). Two participants' essays were dropped from the analysis because their demographic information indicated that they were non-native English speakers who had been speaking English for less than 10 years. This resulted in 112 participants.

Design

The study was a 4-level (task instruction: plot, ambiguous, argument, theme) between-subjects design. Participants were randomly assigned to condition (Plot: N = 31; Ambiguous: N = 28; Argument: N = 28; Theme = 25). Dependent measures were derived from the essays subjects produced as described below.

Materials

Participants read *Harrison Bergeron* (Vonnegut, 1968). This science-fiction short story was selected because it is a fairly transparent allegory and is also similar in style and difficulty to many of the short stories students experience in late high school and college classes. The text

contained 2,201 words and readability was appropriate for high school students and beyond according to the Flesh-Kincaid Reading Ease score of 66.4, grade equivalent of 8.9, and lexile score of 750. However, it is not a staple in the typical literary canon, thus minimizing the likelihood that participants would have previously read or studied this particular story. Additionally, previous research indicates that the themes and conventions of science-fiction are pervasive enough in popular culture that they are relatively ubiquitous (Dixon & Bortolussi, 2009). Thus, by using a science-fiction text we hoped to minimize individual differences in prior knowledge about the genre.

Task-Instruction Conditions. Four different task instructions defined the between-subjects variable:

a) *Plot:* What happened in this story? Use evidence from the text to support your claims.

b) *Ambiguous*:What is this story about? Use evidence from the text to support your claims.

c) *Argument:* Critics often claim that this short-story is a political satire warning us of the dangers of letting "Big Brother" get out of control while others believe it is a story about human potential. Which do you think is the better interpretation? Use evidence from the text to support your claims.

d) *Theme:* Please discuss the theme of the text using evidence from the story to support your claims.

Questionnaire and Demographic Survey. A questionnaire was developed to obtain information about how the participants had interpreted the task instruction and to gather demographic information. The questionnaire asked participants to *Please paraphrase the essay*

question you were given and *Were the goals of the essay clear to you? If not, what was unclear? What would have helped you?* Participants were also asked for basic demographic information (e.g. age, year in school, native language, years of speaking English if a non-native speaker).

Procedure

Task instruction conditions were assigned randomly to each one-hour session of 5-10 participants. After giving informed consent, participants were provided with a hard copy of the *Harrison Bergeron* story and were asked to take 15 minutes to read the story. They were then provided with a response sheet that included the task instruction for the appropriate task instruction condition. Consistent with the circumstances of authentic literary interpretation, the text was available to the participants and participants were encouraged to use the text while composing their essays. Upon completion of the essay, participants filled in a hard-copy version of the questionnaire.

Scoring

The handwritten essays were transcribed and coding was done from the typed transcripts. These transcripts did not indicate the task instruction prompt nor the instructional condition. From these transcripts, essays were parsed into idea units, defined as a complete thought. A single idea unit sometimes corresponded to a single sentence, but some sentences contained more than one complete thought and were coded as multiple idea units. The two authors independently parsed a 20% sample of essays into idea units and achieved 90% agreement, with disagreements resolved in discussion. The first author parsed the remaining essays.

Each idea unit was coded into one of five categories: 1) verbatim or paraphrase, 2) text-based inference, 3) interpretive inference, 4) decision, or 5) other. The first three categories are ordered from verbatim copying of the text to interpretive thematic idea units as elaborated below

(additional information on the coding is provided in Appendix A.) The authors independently coded the idea units in a random subset of 40 essays. Reliability between the two independent coders on classifying the idea units into these five categories produced a Cohen's Kappa of .98. However, because the intact essay transcripts could not be completely blinded due to the differences among the task instructions, a further "reliability" assessment was conducted to check on possible bias in the coding decisions. A random sample of 60 of the idea units was extracted from different essays and randomly ordered by the first author. The second author then classified them into the five categories. Reliability with the original classifications was 95%. Disagreements were resolved through discussion. The first author then coded the remaining essays.

Verbatim and Paraphrase. Idea units were coded as verbatim if they exactly matched the text. Idea units coded as paraphrases preserved the meaning of the text, but used different words to express it. Paraphrases suggest a literal understanding of stated information and the presence of a textbase representation. For example, the text includes the following: *The Handicapper General came into the studio with a double-barreled ten-gauge shotgun. She fired twice, and the Emperor and the Empress were dead before they hit the floor.* The following was coded as a paraphrase: *Then he and the ballerina were killed by Diana Moon Glampers, the Handicapped [sic] General.* The verb phrase *were killed* captures the actions described in more detail in the text. The use of *he and the ballerina*, instead of *Emperor and Empress* reflects a local, referential connection.

As there were very few verbatim units in the essays, verbatim and paraphrase idea units were treated as a single category.

Text-based Inferences. Idea units were coded as text-based inferences if they connected information in the text via relatively simple reasoning processes (e.g., if, then; cause-effect) or via relevant prior knowledge. For example, *Diana Moon Glampers killed them because they tried to show their true selves* creates a causal connection between the murder and motive based on previous events in the story. Text-based inferences such as this one increase the coherence of the representation of a text by explicitly linking ideas in the text and by bringing additional information to the text. They contribute to the construction of a situation model representation of the story, but remained within the story world.

Interpretive Inferences. Idea units were coded as interpretive inferences if they went beyond the world of the story to construct a thematic interpretation or generalization. They either 1) discussed the symbolic meaning of something in the story, or 2) referred to the "moral" or "point" of the story. Examples of ideas coded as interpretive inferences are: *I feel that the theme of this text surrounds the idea that our differences should be celebrated* and *The author in these instances is trying to convey that a society totally on equality would fail because no accomplishments of value would be achieved*.

Decision. Because the argument condition offered two possible interpretations, idea units that indicated the choice of one interpretation over the other (i.e. *I think "Big Brother" is a better interpretation of this story*) were coded in this category.

Other. Idea units that did not meet the criteria for the preceding categories were coded as *other*. As reported below, less than 3% of total idea units were coded into this category.

Four idea unit scores were calculated for each participant: 1) Total, 2) Paraphrases, 3) Text-based inferences, and 4) Interpretive inferences. *Decision* and *other* categories represented only 2.4% of the idea units and were not analyzed.

Results

We predicted that there would be differences in the types of idea units comprising the essays produced in response to the four task instructions. In the plot instruction condition, we expected essays would contain paraphrases and text-based inferences that reflected what happened in the story, with few inferences that went beyond the story world. In the argument and theme conditions, we expected more evidence of interpretive inferences beyond the story world than in the plot condition, but we also expected that these essays would contain some paraphrases and text-based inferences. We expected essays in the ambiguous condition to fall between the plot and the other two conditions with respect to interpretive inferences. We made no specific predictions about the effect of task instruction on length of the essay.

Essay Length

The first analysis established that there was a task instruction effect on total number of idea units (Figure 1). A one-way between-subjects analysis of variance (ANOVA) indicated a main effect of task instruction condition, $F(3,108) = 16.19$, $p < .001$. Follow-up Tukey HSD tests indicated that essays from the plot condition ($M = 22.16$, $SD = 5.48$) contained more total idea units than essays from each of the other three conditions ($p < .02$). Additionally, essays from the ambiguous condition ($M = 17.64$, $SD = 7.56$) contained significantly more idea units than essays from the theme condition ($M = 12.08$, $SD = 3.49$, $p < .01$).

Types of Idea Units

A second set of analyses revealed significant differences related to task instruction condition for each of three categorical types of idea units: paraphrase, text-based inference, and interpretive inferences. Mean frequencies of idea units coded into these categories are also shown in Figure 1.[i] An ANOVA indicated a main effect of task instruction condition on the

number of paraphrase ideas in the essays, $F(3, 108) = 33.98$, $p < .001$. Planned contrasts indicated that the pattern of significant differences was consistent with predictions: The amount of paraphrases was not significantly different for essays from the argument and theme conditions, $t(108) = .54$, *ns*, $d = .28$. Essays in the plot condition ($M = 16.58$, $SD = 6.22$) contained more paraphrases than those in the argument ($M = 4.39$, $SD = 3.19$) and theme ($M = 5.20$, $SD = 2.61$) conditions, $t(108) = 10.04$, $p < .001$, $d = 2.43$. Essays in the plot condition also contained more paraphrases than those in the ambiguous condition (M = 9.82, $SD = 6.96$), $t(108) = 5.00$, $p < .001$, $d = 1.02$.

An ANOVA indicated a main effect of task instruction for text-based inferences, $F(3,108) = 5.76$, $p < .001$. Planned contrasts indicated that essays from the argument condition ($M = 4.64$, SD = 3.94) contained more text-based inferences than the theme condition ($M = 2.32$, $SD = 1.97$), $t(108) = 2.67$, $p < .01$, $d = .74$. Planned contrasts indicated that the plot condition produced more text-based inferences when compared against the combined mean ($M = 3.55$, $SD = 3.35$) for argument and theme conditions, the two expected to produce more interpretive processing, $t(108) = 1.99$, $p = .05$, $d = .45$. However, the argument condition ($M = 4.64$, SD = 3.94) contained more text-based inferences than the theme condition ($M = 2.32$, $SD = 1.97$), $t(108) = 2.67$, $p < .01$, $d = .74$.

A final ANOVA indicated a main effect of task instruction on the amount of interpretive inferences, $F(3,108) = 21.28$, $p < .001$. As predicted, planned comparisons showed no difference between the argument ($M = 5.29$, $SD = 3.41$) and theme ($M = 4.56$, $SD = 2.55$) conditions, $t(103) = 1.05$, *ns*, $d = .24$. Essays from the plot condition ($M = .68$, $SD = 1.56$) contained fewer interpretive inferences than those in the argument and theme conditions, $t(108) = 7.48$, $p < .001$, $d = 1.79$, and fewer than in the ambiguous condition ($M = 2.00$, $SD = 2.26$), $t(108) = 2.02$, $d =$

.68. These data indicate that, despite having written less overall, participants in the argument and theme conditions were constructing more interpretive inferences than participants in the other conditions.

We pursued two further analyses to examine whether the task instruction condition differences reported above in terms of means across participants were consistently reflected in individuals. The first analysis addressed the differential prevalence of paraphrasing across the four conditions. We categorized each essay as containing 50% or more of the idea units coded as paraphrases or as less than 50% coded as paraphrases. The resulting distribution by task instruction condition is shown in Table 1. A chi-square analysis indicated a significant relationship between essay composition and task instruction condition, $\chi^2(1, N = 112) = 35.92$, $p < .001$. To follow up this chi-square analysis, we calculated adjusted residuals that indicate a standardized measure of how much the observed frequency differs from expected. Absolute values greater than 1.96 indicate significant differences greater than $p < .05$ (Agresti, 2002). These residuals indicate that the significant chi-square is being driven by the plot and argument conditions. More plot condition participants than would be expected by chance produced essays with 50% or more of the idea units categorized as paraphrases. In contrast, more argument condition participants than would be expected by chance produced essays composed of less than 50% paraphrases. Though there was no significant difference in the theme condition, the trend is in the predicted direction in that 15 participants' essays included less than 50% paraphrase and only 10 participants' essays included 50% or more paraphrases. These findings provide support at the individual level for the overall trends in the means reported above.

The second analysis addressed the differential prevalence of interpretive inferences. We categorized each subject's essay in terms of whether it did or did not include any interpretive

inferences. That distribution is shown in Table 2. The trend is again clear, with a significant chi-square test indicating a relationship between task instruction and inclusion of any interpretive inference, $\chi^2(1, N = 112) = 49.39$, $p < .001$. Adjusted residuals indicated that plot condition participants were less likely than chance to include any interpretive inferences, but argument and theme condition participants were more likely than chance to include at least 1 interpretive inference.

To summarize, the data are consistent at both the aggregate and individual level with our predictions that the plot condition would produce essays reflecting the literal story, as evidenced by the prevalence of paraphrased information, but that the argument and theme instructions would encourage more interpretation, as evidenced by the occurrence of interpretive inferences. As the stacked bar graphs in Figure 1 show, essays in the argument and theme conditions were comprised of a greater proportion of interpretive inferences to total ideas than essays in the other two conditions. Essays in the plot condition predominantly consisted of paraphrases of information in the story. Finally, as expected the ambiguous condition was just that, with trends suggesting some tendency to interpret, but a strong reliance on paraphrasing. It is important to note that analyses of the questionnaires completed at the end of the experimental session did not suggest any systematic relationships between the patterns of findings and students' backgrounds or their interpretation of the instructions. Students' responses indicated that the task instructions were clear: the overwhelming majority of subjects in all conditions accurately restated what they had been asked to do by their instructions.

Discussion

Essay composition revealed a significant effect of task instruction on total number and types of idea units. Essays produced in the plot condition were the longest and contained little

evidence of interpretive inferences. Essays in the argument and theme conditions were the shortest, yet all but two subjects included at least one interpretive inference, providing evidence of interpretation beyond the world of the story. We speculate that the length differences reflect an effort on the part of those in the plot condition to include each major plot point, whereas those in the argument and theme conditions focused on inferences across larger spans of text to get at "the big picture" rather than the details. In all conditions, essays contained text-based inferences that helped build a coherent situation model. The essays produced in the ambiguous condition were similar to the plot condition essays with respect to paraphrasing and text-based inferences, but almost two-thirds of the essays also included at least one interpretive inference. We delay further discussion of the findings until after reporting Experiment 2.

Experiment 2

One limitation of Experiment 1 is that it only used one short story. Experiment 2 addressed the generalizability of the findings to a different short story. *The Elephant* (Mrozek, 1957) is a story about a career- and budget-minded zookeeper and his plan to display a fake, rubber elephant instead of a real one. Like *Harrison Bergeron,* it is a political satire, but is set in a more realistic context (realistic narrative genre) than a science fiction story (Stankiewicz, 1971).

Method

Participants

A new group of 116 students (56 males) from the same subject pool as in Experiment 1 received course credit for participation in Experiment 2. The mean age was 19.40 (*SD* = 1.98). Two participants were excluded from the analyses due to lack of English proficiency, leaving N = 114.

Materials

Text. The short story for this experiment was an English adaptation of *The Elephant* (Mrozek, 1957). Originally written in Polish, the English translation was 1,180 words with Flesh-Kincaid Reading Ease score of 67.1, grade equivalent of 7.7, and a lexile score of 1130. As with *Harrison Bergeron*, *The Elephant* is sometimes used in English Language Arts curricula, but infrequently enough that none of the participants reported being familiar with the text.

Task-Instruction Condition. The task instruction conditions were the same as for Experiment 1, with the exception of the specific wording of the argument task instruction to reflect interpretations specific to this story.

> Argument*: Critics often claim that this short story is a political satire about government corruption while others believe it is more simply about the dangers of putting the needs of one before the needs of many. Which do you think is the better interpretation? Use evidence from the text to support your claims*

Demographic Questionnaire. The same demographic questionnaire used in Experiment 1 was administered in Experiment 2. Since the responses regarding the paraphrasing and clarity of the task instruction in Experiment 1 indicated that participants' interpretations accurately reflected the instructions, we did not administer these questions in Experiment 2.

Procedure and Scoring

The procedure was the same as in Experiment 1. Essay scoring was identical to Experiment 1. The first author parsed the essays into idea units. Two new raters were trained to code for the five idea unit types (paraphrase, text-based inference, interpretive inference, decision, and other) on a subset of 30 essays and achieved inter-rater reliability of 94%. These raters coded the remaining essays, with disagreements resolved by the first author.

Results

Predictions for this experiment were similar to those for Experiment 1.

Essay Length

An ANOVA indicated a significant effect of task instruction on the total number of idea units in the essays $F(3,110) = 6.24$, $p < .001$, consistent with Experiment 1. However, as is evident in Figure 2, there were some discrepancies from Experiment 1 with respect to which conditions were significantly different from each other. Tukey HSD tests indicated that, as in Experiment 1, essays in the plot condition ($M = 15.84$, $SD = 5.81$) were longer than those in the argument condition ($M = 10.69$, $SD = 4.35$). Unlike Experiment 1, length in the argument condition was also significantly different from ambiguous ($M = 16.12$, $SD = 6.04$) and theme ($M = 14.57$, $SD = 5.16$; $p < .04$) and there were no differences among plot, ambiguous, and theme conditions.

Types of Idea Units

Predictions based on our a priori hypotheses and the findings from Experiment 1 were that essays from the plot condition would contain more paraphrases and text-based inferences, whereas those from the argument and theme conditions would contain more interpretive inferences; participants in the ambiguous condition were expected to fall somewhere in between. These predictions were confirmed and replicated the findings from Experiment 1.

An ANOVA revealed a main effect of task instruction condition on the number of paraphrases in the essays, $F(3,110) = 32.02$, $p < .001$.[ii] Planned contrasts indicated that the overall pattern of condition means replicated that of Experiment 1, with one exception. As in Experiment 1, essays from the plot condition ($M = 12.47$, $SD = 5.02$) contained more paraphrases than those in the argument ($M = 2.38$, $SD = 2.51$) and theme ($M = 5.21$, $SD = 4.51$) conditions, $t(110) = 8.26$, $p < .001$, $d = 1.99$, and there was no significant difference between the argument and theme conditions, $t(110) = 2.25$, $p < .03$, $d = .77$. However, unlike Experiment 1, there was

no significant difference on paraphrases between the plot and ambiguous (M = 12.00, SD = 6.38) conditions, $t(110) = .37$, *ns*, $d = .08$.

Unlike Experiment 1, an ANOVA indicated no significant effect of task instruction on number of text-based inferences, $F(3,110) = 2.63$, *ns*. As Figure 2 indicates, the amount of text-based inferences was relatively similar across conditions, although the relative proportion of paraphrases and text-based inferences to total composition of the essay preserves the pattern obtained in Experiment 1 wherein text-based inferences constituted a larger proportion of the total essay in the argument and theme conditions than in the plot and ambiguous conditions.

An ANOVA on number of interpretive inferences replicated Experiment 1 in finding a main effect of task instruction condition, $F(3,110) = 19.85$, $p < .001$. Planned contrasts replicated the patterns obtained in Experiment 1 with the exception of the ambiguous condition. As in Experiment 1, essays in the plot condition (M = .59, SD = 1.64) contained fewer interpretive inferences than those in the argument (M = 4.76, SD = 4.19) and theme (M = 4.61, SD = 2.88) conditions, $t(110) = 6.74$, $p < .001$, $d = 1.49$; and there was no significant difference between the argument and theme conditions $t(110) = .21$, *ns*, $d = .04$. Unlike Experiment 1, interpretive inferences in the plot and ambiguous (M = .88, SD = 1.13) conditions did not differ, $t(110) = .78$, *ns*, $d = .21$. Thus, in Experiment 1 the essay composition in the ambiguous condition appeared to fall between those in the plot and those in argument and theme conditions. In Experiment 2, ambiguous condition essays looked much like those produced in the plot condition.

The effects of task instruction condition at the aggregate level were reflected in the individual analyses: the plot condition differed from argument and theme conditions and the ambiguous condition behaved more similarly to the plot condition in Experiment 2 than it had in Experiment 1. Table 3 shows the distribution of essays with 50% or more versus less than 50%

of total idea units categorized as paraphrases. The chi-square analysis indicated a significant relationship between essay composition and task instruction condition, $\chi^2(1, N = 114) = 55.42$, $p < .001$. Adjusted residuals indicated that the plot and ambiguous conditions were more likely than chance to produce essays containing 50% or more paraphrases, whereas the argument and theme conditions were more likely than chance to produce essays containing less than 50% paraphrases.

Likewise, the likelihood of participants including at least one interpretive inference indicated that plot and ambiguous conditions were highly similar to each other and different than the argument and theme conditions. A chi-square analysis on the frequency distribution shown in Table 4 revealed a significant relationship between task instruction condition and inclusion of at least one interpretive inference in the essay, $\chi^2(1, N = 114) = 52.55$, $p < .001$. Adjusted residuals indicated that participants in the plot condition and ambiguous conditions were more likely than chance to produce essays that did not include any interpretive inferences, whereas those in the argument and theme conditions were more likely than chance to produce essays that included at least one interpretive inference.

Discussion

Experiment 2 replicated the key findings of Experiment 1: Essays produced in the argument and theme conditions produced more interpretive inferences than essays in the plot condition. Participants in the plot condition produced more paraphrases than those in the argument and theme conditions. The composition of essays produced in the ambiguous instruction condition resembled the plot condition more in Experiment 2 than in Experiment 1. In addition, the essays in Experiment 2, regardless of condition, appeared to be, on average, shorter than those in Experiment 1. These differences are likely due to the specific content of the two

stories. Nevertheless, the effect of the task instructions designed to induce an interpretive stance (argument and theme) was robust in both experiments.

General Discussion

These studies explored the effects of task instructions designed to induce literal versus interpretive stances toward text on the propensity to include interpretations in essays. Plot-oriented instructions led to essays that were dominated by paraphrases of story events and text-based inferences that connected them. Conversely, argument and theme-oriented instructions produced essays that included more emphasis on interpretations that extended beyond the story world. Given the existing research suggesting that interpretive reasoning behavior is largely absence in classrooms and is not evident on national assessments (NAEP, 2009), it was surprising to find robust differences in the composition of the essays with such a modest manipulation.

The major difference across the two experiments was in the behavior of those in the ambiguous task instruction condition. In the ambiguous condition of Experiment 1, about half of the essay content reflected inferences and about half paraphrases. In Experiment 2, the ambiguous condition essays were composed mostly of paraphrases and text-based inferences -- almost indistinguishable from the essays produced in the plot condition. The ambiguous condition was designed to serve as a neutral instruction with respect to stance. We expected that essays produced in this condition would reflect the "default" stance that readers take to literary texts. Readers default stances would, in turn, depend on knowledge of conventions literary author use to convey their messages, knowledge of the specific author, their understandings of the purposes of reading literature, as well as interest and motivational dimensions. Thus, we expected that some participants would spontaneously engage in interpretive reasoning, while

others would maintain a literal stance. The data, especially for Experiment 1, indicate that this is apparently what happened.

The differences in essay composition in the ambiguous condition across the two experiments also suggests that characteristics of different texts may make it more or less likely that readers will adopt an interpretive stance. That is, the text in Experiment 1 (*Harrison Bergeron*) is science-fiction, whereas *The Elephant* is realistic fiction. As mentioned, previous work indicates that non-expert readers are quite familiar with the themes and conventions of science-fiction stories (Dixon & Bortolussi, 2009), so it may be that science-fiction is a genre that is more likely to induce an interpretive stance even in readers who are relatively naïve with respect to conventions of literary interpretation. In future work, it would be beneficial to replicate task instruction manipulations across a wider variety of texts and text genre to explore whether and what particular characteristics of texts are more likely to induce interpretive stances on the part of individuals not experienced with literary analysis.

In addition to broadening the types of texts, it would be useful to include measures of online processing that would provide insight into the mechanisms that underlie the construction of interpretive inferences. We speculate that interpretive inferences are likely to occur at points in stories were readers experience breaks in coherence for which typical repair strategies are not effective. If task instructions highlight the importance of going beyond the literal, as the argument and theme instructions appear to have done in the present experiments, it may be that readers shift their default standards of coherence (van den Broek, Young, Tzeng, & Linderholm, 1999). For example, when readers take an interpretive stance toward text they may be more tolerant of odd juxtapositions of events, or entertain multiple or ambiguous explanatory frameworks for characters actions. Ultimately, it is from these more fragmentary elements that

readers may be generating inferences that speak to the human condition beyond the world of the story.

The present research also cannot speak to the issue of the longer-term impact of a single exposure to task instructions that explicitly direct them to take an interpretive stance toward text. We are doubtful that a single exposure would be sufficient to effect long-lasting change in readers propensities to adopt an interpretive stance. We expect that repeated exposure to such instructions and opportunities to apply interpretive reasoning to a range of texts is necessary before students are likely to exhibit interpretive reasoning without prompting. It might also be a useful instructional strategy to have students contrast the reasoning required to respond to different types of task instructions such as the ones used in the present experiments. Such an exercise could make more explicit and transparent different purposes for reading and appropriate forms of reasoning and response.

The lack of processing data in the research reported in this paper, prevents us from differentiating among various ways in which the task instructions might have impacted processing, including what readers attended to in the stories, or what aspects of prior knowledge were activated and when, as well as shifting standards of coherence. Nevertheless, the present experiments do provide support for the need to expand models of text representation to accommodate interpretive inferences that go beyond the world of the story. Such expansion might take the form of attaching nodes to a situation model representation or by adding specification to other levels that have been proposed, specifically communicative levels (Graesser & McNamara, 2010; Graesser, Singer, & Trabasso, 1994; van Dijk & Kintsch, 1983). At the communicative level, information that might be represented could include particular authors' favorite themes, their political and social orientations and causes, and messages from

other literary works. Such sourcing information would allow readers to place a particular work by an author in the context of other works by the same author. This information could then inform subsequent interpretations of additional literary works by the same author, reinterpretations of previously read works, or serve as a basis for comparisons across literary works by the same or by different authors who were writing from different perspectives and contexts. Just as sourcing and contextualizing are important processes in historical inquiry (Wineburg, 1991) and models of mental representations of texts have been expanded to account for this information and its role in historical interpretation, (e.g., Perfetti, Rouet, & Britt, 1999; Rouet, 2006; Rouet & Britt, 2011), representations for literary works will need to do so as well (Goldman, et al., in press). Although use of source and context information is evident in the interpretation processes employed by literary experts, whether and how such information is connected to the representation of a particular literary work and drawn upon subsequently has not received serious attention.

The present experiments indicate that when readers who have had no training in literary analysis are given an external goal of making an interpretation, their responses appear more expert-like (Graves & Frederiksen, 1991; Zeitz, 1994): Those in the argument and theme conditions produced messages about the human condition and generalizations about human nature. Nevertheless, the interpretations these students created were not as sophisticated as those that might be produced by literary experts. Indeed, the "evidence" cited by readers in the present studies tended to be specific events from the stories. These interpretations and the reasoning supporting them made little contact with disciplinary-based practices of close textual analysis, tracing authors' craft and style in using language to communicate nonliteral messages. Expert evaluations use disciplinary criteria for such judgments. In addition to what happened to whom

in the text, expert literary analysis draws on how authors select and use language to convey their messages. Use of language includes specific rhetorical devices such as symbolism and irony, repeated use of words or phrases, and other forms of intentional "manipulation" of language and language structures (Hillocks & Ludlow, 1984; Rabinowitz, 1987). As well, experts connect particular plot structures and character types to literary movements and philosophical perspectives that extend well beyond the world of the story. Further work on comprehension as literary interpretation is needed to examine how the knowledge that underlies these aspects of literary reasoning is acquired and activated in processing literary works.

References

Agresti, A. (2002). Categorical data analysis. NJ: Wiley.

Chi, M. T. H., Feltovich, P., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, 5, 121-152.

Claassen, E. (2012). *Author representation in literary reading*. John Benjamins: Utrecht, Netherlands.

Council of Chief State School Officers (CCSSO). (2010). *Common Core State Standards.* Washington, DC: National Governors Association Center for Best Practices, Council of Chief State School Officers.

Dixon, P. & Bortolussi, M. (2009). Readers' knowledge of popular genre. *Discourse Processes, 46*, 541-571.

Earthman, E. (1992). Creating the virtual work: Readers' processes in understanding literary texts. *Research in the Teaching of English. 26,* 351-384.

Goldman, S. R. (1997). Learning from text: Reflections on the past and suggestions for the future. *Discourse Processes, 23,* 357-398.

Goldman, S. R. (2004). Cognitive aspects of constructing meaning through and across multiple texts. In N. Shuart-Ferris & D.M. Bloome (Eds.), *Uses of intertextuality in classroom and educational research.* (pp. 313-347). Greenwich, CT: Information Age Publishing.

Goldman, S. R., Golden, R. M., & van den Broek, P. (2007). Why are computational models of text comprehension useful? In F. Schmalhoefer & C. A. Perfetti (Eds.), *Higher Level Language Processes in the Brain: Inference and Comprehension Processes.* Mahwah, NJ: Erlbaum.

Goldman, S. R., McCarthy, K. S., & Burkett, C. (in press). Interpretive inferences in literature. In E. O'Brien, A. Cook, & R. Lorch (Eds.), *Inferences during reading*. MA: Cambridge University Press.

Graesser, A. C. & McNamara, D. S. (2011). Computational analyses of multilevel discourse comprehension. *Topics in Cognitive Science, 3,* 371-398.

Graesser, A. C., Millis, K. K., & Zwaan, R. A. (1997). Discourse comprehension. *Annual Review Psychology, 48,* 168-189.

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review, 101,* 371-395.

Graves, B. & Fredriksen, C. (1991). Literary expertise in the description of fictional narrative. *Poetics, 20,* 1-26.

Hamilton, C. (2011). Allegory, blending, and censorship in modern literature. *Journal of Literary Semantics, 40,* 23-42.

Hillocks, G. & Ludlow, L. H. (1984). A taxonomy of skills in reading and interpreting fiction. *American Educational Research Journal, 21,* 7-24.

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness.* Cambridge, MA: Harvard University Press.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model, *Psychological Review, 95,* 163-182.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge, England: Cambridge University Press.

Kotovsky, K., Hayes, J. R., & Simon, H. A. (1985). Why are some problems hard? Evidence from tower of Hanoi. *Cognitive Psychology, 17,* 248-294.

Langer, J. A. (2010). *Envisioning Literature: Literary understanding and literature instruction, 2nd edition.* NY: Teachers College Press.

Lee, C. D. (2007). *The role of culture in academic literacies: Conducting our blooming in the midst of the whirlwind*. NY: Teachers College Press.

Lee, C. D. (2011). Education and the study of literature. *Scientific Study of Literature, 1,* 49-58.

Letmendia, V.C. (1992). Revolution on animal farm: Orwell's neglected commentary. *Journal of Modern Literature, 28,* 127-137.

Mason, L., Scirica, F., & Salvi, L. (2006). Effects of beliefs about meaning construction and task instructions on interpretation of narrative texts. *Contemporary Educational Psychology, 31,* 411-437.

McNerny, M. W., Goodwin, K. A., & Radvansky, G. A. (2011). A novel study: A situation model analysis of reading times. *Discourse Processes, 48,* 453-474.

Mrozek, S. (1972). *The Elephant* (Słoń). Westport, CT: Greenwood Press.

National Assessment of Educational Progress (NAEP). (2009). *NAEP 2008 Trends in Academic Progress (NCES 2009–479).* Prepared by Rampey, B. D., Dion, G. S., & Donahue, P. L. for the National Center for Education Statistics, Institute of Education Sciences. Washington, DC: US Department of Education.

National Assessment Governing Board (2008). *Reading Framework for the 2009 National Assessment of Educational Progress.* Washington D. C.: American Institutes for Research.

Perfetti, C.A., Rouet, J-F, & Britt, M.A. (1999). Towards a theory of documents representation. In H. van Oostendorp & S.R. Goldman (Eds.), *The Construction of Mental Representations During Reading* (pp. 99-122). Mahwah, NJ: Erlbaum Associates.

Peskin, J. (1998). Constructing meaning when reading poetry: An expert-novice study. *Cognition and Instruction, 16,* 135-263.

Rabinowitz, P. (1987). *Before reading: Narrative conventions and the politics of interpretation.* Ithaca, NY: Cornell University Press.

Rapp, D. N., Komeda, H., & Hinze, S. R. (2011). Vivifications of literary investigation. *Scientific Study of Literature, 1,* 123-135.

Rouet, J-F. (2006). *The Skills of Document Use: From Text Comprehension to Web-Based Learning*. Mahwah, NJ: Erlbaum.

Rouet, J-F. & Britt, M. A. (2011). *Relevance processes in multiple document comprehension.* In M.T. McCrudden, J. P. Magliano, & G. Schraw (Eds.) Text Relevance and Learning from Text (pp. 19-52). Charlotte, NC: Information Age.

Stankiewicz, M. G. (1971). Slawomir Mrozek: Two forms of the absurd. *Contemporary, Literature, 12,* 188-203.

Van den Broek, Lorch, Linderholm, & Gustafson (2001). The effects of readers' goals on inference generation and memory for text. *Memory & Cognition, 29,* 1081-1087.

Van Dijk, T. A. & Kintsch, W. (1983). *Strategies of discourse comprehension.* New York: Academic Press.

Vipond, D., & Hunt, R. A. (1984). Point-driven understanding: Pragmatic and cognitive dimensions of literary reading. *Poetics, 13,* 261-277.

Vonnegut, K. (1968). Harrison Bergeron. In *Welcome to the Monkey House*, 7-13. New York: Delacorte.

Walker, A. (1982). *The Color Purple.* New York: Pocket Books.

Wiley, J. & Voss, J. F. (1999). Constructing arguments from multiple sources: Tasks that promote understanding and not just memory for text. *Journal of Educational Psychology, 91*, 301-311.

Wiley, J., Goldman, S. R., Graesser, A. C., Sanchez, C. A., Ash, I. K., Hemmerich, J. A. (2009). Source evaluation, comprehension, and learning in internet science inquiry tasks. *American Educational Research Journal, 46,* 1060-1106.

Wineburg, S. (1991). Historical problem solving: A study of the cognitive processes used in the evaluation of documentary and pictorial evidence. *Journal of Educational Psychology, 83*, 73-87.

Zeitz, C. M. (1994). Expert-novice differences in memory, abstraction, and reasoning in the domain of literature. *Cognition and Instruction, 12*, 277-312.

Table 1

Paraphrasing in Essays: Number of subjects with less than 50% or 50% or more for each task instruction

	Less than 50% of idea units coded as paraphrases	50% or more idea units coded as paraphrases	Adj. Residual
Plot	3	28	5
Ambiguous	12	16	0.7
Argument	24	4	-4.6
Theme	15	10	-1.3

Table 2

Interpretive Inference Inclusion in Essays: Number of subjects not including or including at least one interpretive inference for each task instruction condition

	Did not include any interpretive inferences	Included at least 1 interpretive inference	Adj. Residual
Plot	24	7	6.5
Ambiguous	8	19	0.1
Argument	1	27	-3.6
Theme	1	24	-3.3

Table 3

Paraphrasing in Essays: Number of subjects with less than 50% or 50% or more for each task instruction

	Composition of Essay		
	Less than 50% of idea units coded as paraphrases	50% or more idea units coded as paraphrases	Adj. Residual
Plot	2	30	5.4
Ambiguous	5	20	3
Argument	25	4	-5
Theme	21	7	-3.5

Table 4

Interpretive Inference Inclusion: Number of subjects not including or including at least one for each task instruction condition

Composition of Essay

	Did not include any interpretive inference	Included at least one interpretive inference	Adj. Residual
Plot	27	5	-6.1
Ambiguous	14	11	-1.5
Argument	4	25	3.3
Theme	1	27	4.5

Appendix A

Description and Examples of the Types of Idea Units

	Description	**Van Dijk & Kintsch**	**Example from Experiment 1**	**Example from Experiment 2**
Verbatim	Copied directly from the text	Surface Code	*The Handicapper General, came into the studio with a double-barreled ten-gauge shotgun. She fired twice, and the Emperor and the Empress were dead before they hit the floor.*	*The schoolchildren who had witnessed the scene in the zoo soon started neglecting their studies and turned into hooligans. It is reported they drink liquor and break windows. And they no longer believe in elephants.*
Paraphrase	Rewording of the sentences from the text; Summary or combining of multiple sentences from the text	Textbase	*Then [Harrison] and the ballerina were killed by Diana Moon Glampers, the [sic] Handicapped General.*	*After seeing this the students gave up on education became drunks and stopped believing in elephants.*
Text-Based Inference	Reasoning-based on information presented in the story, with some use of prior knowledge; connecting information from two parts of the text	Situation Model	*Diana Moon Glampers killed them because they tried to show their true selves.*	*After being deceited [sic] by the fake elephant, the children became poor students, and grew up behaving badly because they were lied to.*
Interpretive Inference	Inferences that reflect nonliteral, thematic interpretations of the text	Communicative	*It shows what kind of a place the world can turn out to be if we let [the government] get out of control.*	*The theme is that being lied to ends the innocence of the young boys and girls.*

[i] A parallel set of analyses was conducted using proportions of total idea units rather than frequencies to adjust for differences in length among conditions. These analyses showed similar patterns of results as the analyses of frequencies. There was a significant effect of task instruction condition on the proportion of paraphrase to total idea units, $F(3,108) = 22.97$, $p < .001$. Planned contrasts indicated the plot condition yielded a higher proportion of paraphrases than the argument and theme conditions, $t(108) = 7.88$, $p < .001$. It also had a higher proportion of paraphrases than the ambiguous condition, $F(3, 108) = 4.18$, $p < .001$. Theme had a higher proportion of paraphrases than the argument condition, $F(3, 108) = 2.32$, $p < .05$. For the proportion of text-based inferences, there was a significant effect of task instruction, $F(3, 108) = 6.75$, $p < .001$. The argument condition yielded a higher proportion of text-based inferences than the theme condition, $t(108) = 2.53$, $p < .02$. There was no significant difference between the plot condition and the argument and theme conditions, $t(108) = .43$, *ns.* The ambiguous condition yielded a higher proportion of text-based inferences than the plot condition, $t(108) = 3.37$, $p < .001$. For proportion of interpretive inferences, there was an overall effect of task instruction condition, $F(3, 108) = 29.32$, $p < .001$. There was no difference between the argument and theme conditions, $t(108) = .66$, *ns.* Argument and theme conditions yielded a significantly higher proportion of interpretive inferences than participants in the plot condition, $t(108) = 8.67$, $p < .001$. The proportion of interpretive inferences in the ambiguous condition versus the plot condition was not significant, $t(108) = 1.83$, $p = .069$.

[ii] As in Experiment 1, a parallel set of analyses was conducted using proportions of total idea units rather than frequencies to adjust for differences in length among conditions. These analyses generally showed the same overall effects as the analyses of the frequencies.
There was an effect of task instruction condition on the proportion of paraphrases, $F(3, 110) = 32.59$, $p < .001$. Planned contrasts indicated no difference between the argument and theme conditions, $t(110) = 1.43$, *ns.* The plot condition yielded a higher proportion of paraphrases than the argument and theme conditions, $t(110) = 8.91$, $p < .001$. There was no significant difference between the plot and ambiguous conditions, $t(110) = 1.32$, *ns.* For proportion of text-based inferences, there was a significant effect of task instruction, $F(3, 110) = 4.86$, $p < .01$. Planned contrasts revealed no significant difference between the argument and theme conditions, $t(110) = .20$, *ns.* The plot condition yielded a significantly lower proportion of text-based inferences than the argument and theme conditions, $t(110) = 3.74$, $p < .001$. There was no significant difference between the plot and ambiguous conditions, $t(110) = 1.58$, *ns.* Finally, there was a significant effect of task instruction on the proportion of interpretive inferences, $F(3, 110) = 21.84$, $p < .001$, and an identical There was no difference in the proportion of interpretive inferences between the argument and theme conditions, $t(108) = 1.47$, *ns.* Participants in the argument and theme conditions yielded higher proportions of interpretive inferences than those in the plot condition, $t(110) = 6.89$, $p < .001$. There was no significant difference between the plot and ambiguous conditions, $t(110) = .31$, *ns.*