

---

## Basic Studies in Literature

---

### Project READI Technical Report #8

Susan R. Goldman, Joseph Magliano,  
Carol D. Lee, Stephen Briner, Sarah  
Levine, Katherine S. McCarthy, and  
Candice Burkett

# PROJECT **READi**

inquirium



Citation for this Report: Goldman, S. R., Magliano, J., Lee, C. D., Briner, S., McCathy, K., & Burkett, C. (2016). *Basic studies in literature*. READI Technical Report #8. Retrieved from URL: [projectreadi.org](http://projectreadi.org)

The authors would like to acknowledge the contributions of the following members of the Project READI literature team: Teresa Sosa, MariAnne George, Rick Coppola, Angela, Fortune, Allison Hall, Jessica Chambers, and Courtney Milligan.

Please send us comments, questions, etc.: [info.projectreadi@gmail.com](mailto:info.projectreadi@gmail.com)

Project READI was supported by the *Reading for Understanding (RFU)* initiative of the Institute for Education Sciences, U. S. Department of Education through Grant R305F100007 to the University of Illinois at Chicago from July 1, 2010 – June 30, 2016. The opinions expressed are those of the authors and do not represent views of the Institute or the U. S. Department of Education.

Project READI operated as a multi-institution collaboration among the Learning Sciences Research Institute, University of Illinois at Chicago; Northern Illinois University; Northwestern University; WestEd's Strategic Literacy Initiative; and Inquirium, LLC. Project READI developed and researched interventions in collaboration with classroom teachers that were designed to improve reading comprehension through argumentation from multiple sources in literature, history, and the sciences appropriate for adolescent learners. Curriculum materials in the READI modules were developed based on enacted instruction and are intended as case examples of the READI approach to deep and meaningful disciplinary literacy and learning.

©2016 Project READI

## Overview

Basic studies in literature pursued three goals throughout Project READI. The first was to contribute to the knowledge base on adolescent’s understanding of literary works. This work was informed by several areas of prior empirical work, including models of text comprehension, expert – novice studies of literary reading, and the Hillocks and Ludlow (1984) hierarchical taxonomy of literary comprehension processes. The second goal was to explore various forms of instructional tools and strategies for supporting students in making sense of literary works, including the wording of task instructions. Each of these goals contributed to the third goal, assessment development. One assessment tapped epistemic cognition in literature (Yukhymenko, et al., 2016; Yukhymenko, et al., READI Technical Report #5). The second instrument was designed to elicit students’ comprehension and evidence-based argumentation (EBA) performance for pairs of literary works. The development of the EBA assessment is reported in READI Technical Report #10. All three of the literature basic studies strands contributed to and were influenced by the iterative design-based development and research of the middle and high school interventions to promote interpretation of literary works. Indeed, a large subset of the basic studies literature team were also members of the intervention design team and some of the basic process studies were carried out in the classrooms of design-team teachers.

All of the work in literature was heavily shaped by the articulation of the overarching Project READI Disciplinary Literacy framework, the development of which began with the inception of the project. Thus, the core constructs related to why and how we make sense of literary works for purposes of understanding the human condition, the inquiry and problem solving strategies and features of literary works important to interpretation lie at the very heart of the work in literature. The core constructs and articulation of the epistemology of literature and implications of this for literary problem solving and interpretation are detailed in Goldman, et al. (2016), Lee (2007), and Lee, Goldman, Levine, & Magliano (2016).

This technical report describes the work conducted during Project READI related to the first two goals: (1) processes of literary comprehension and interpretation and (2) supports for literary comprehension and interpretation. Within each section, the work is organized topically and then within topic, chronologically. We selected the rhetorical device of symbolism in stories that dealt with the theme “coming of age” because symbolism is ubiquitous in young adult literature from middle school onward and the transition from childhood to adulthood is the essence of adolescence. This rhetorical device and theme guided the selection of literary works for the basic studies, assessment development and figured heavily in the instructional goals for the intervention. This alignment across the multiple strands of work increased the coherence of the efforts.

## I. Literary Comprehension and Interpretation Processes

### 1. Baseline Study 1:

Baseline studies to assess students’ literary interpretation skills provided a starting point and basis for further basic comprehension and interpretation process studies as well as for the other strands of work (instructional supports and assessment development). We designed a task to assess students’ abilities to interpret symbolism in two short stories and to argue how language

functioned across two texts to convey symbolic meanings. Thus, we intended to elicit students' reasoning about how authors were using symbolic language to convey messages about the human condition. Message and use of language correspond to Hillocks and Ludlow's (1984) most advanced interpretive skills. However, we were operating from the premise that the more complex interpretive skills depended on students understanding the basic plot and character aspects of the text. Thus, the baseline task "walked" students through detailing the events in the plot and basic character dimensions for each of the two stories prior to engaging in the interpretive comparison of the two texts in the open-ended written response. We also assumed that the processes of comprehension and interpretation were similar at the younger adolescent grade bands and the older grade bands but that what differed was the complexity of the texts for which students were expected to utilize these skills. Thus, in these initial baseline studies we selected three texts—one common text across the grade levels along with one for grades 6–8 and a second for grades 9–12.

## **Methods and Procedure**

The sample consisted of 150 secondary students from the Chicago area in grades 9–11, including both regular and honors classes. We selected three stories that posed similar thematic categories (i.e. coming of age) and embodied a core symbolic reference that was the target of interpretation" "Marigolds" by Eugenia Collier, "The Jacket" by Gary Soto, and "Eleven" by Sandra Cisneros. We kept one story ("The Jacket) constant across grade levels in order to begin to consider potential developmental dimensions involved in literary reasoning. As we began this work, we found that one of the stories in the original pairing – "Marigolds" – was of such a length as to constrain the time students were able to devote to the entire task. Thus, 51 of the 9<sup>th</sup> graders read "Marigolds" and "The Jacket," and 22 of the 9<sup>th</sup> graders read "Eleven" and "The Jacket." All 10<sup>th</sup> and 11<sup>th</sup> graders read "Marigolds" and "The Jacket."

The students' essays were analyzed by a team of four coders. The essays were typed into a database according to idea units. Inter-rater reliability on the parsing of idea units was 90% for a 10% sample of essays. Idea units were analyzed by a coding scheme developed by the Literature team to assess the four targeted areas we were interested in: symbolism as an interpretive device (labeled addressing the Task), comparison across stories, quality of literary reasoning, and structure of argument.

*Addressing the Task:* Does the student's response address the task (i.e., how the story's images function to convey information about characters or their worlds).

*Synthesis across stories:* Does the student address the stories separately, or is there an attempt to discuss how the characters or their worlds are similar or different?

*Quality of Literary Reasoning:* Does the essay contain inaccurate claims, and are the accurate claims supported by evidence from one or both stories?

*Structure of argumentation:* Does the essay contain claims and evidence? The analysis of the structure of arguments was based on the work of Stephen Toulmin, identifying claims, evidence, warrants and backing, nested claims, and counter arguments.

Detailed rubrics for each category in the coding scheme were developed through discussion among a team of 3 scorers working on a 10% sample of essays. With each round of discussion, the group scored a new 10% set of essays. Once the final set of rubrics were determined the three scorers scored the entire sample of essays. Inter-rater reliability on classification of essays for each of the categories averaged 85% across scorers and categories

## **Key Findings**

*Addressing the Task:* While not statistically significant, we did find a trend for 10<sup>th</sup> and 11<sup>th</sup> graders to score higher on addressing the task than 9<sup>th</sup> graders who read the same set of stories. Even so, less than half of the students across grades addressed the task, with the exception of 9<sup>th</sup> graders who read “The Jacket” and “Eleven,” 50% of whom scored a one, indicating that they mentioned and attempted to relate the symbols (jacket, being 11 years of age) to the meaning of the stories.

*Synthesis:* Synthesis across stories was scored on a scale from 0–4. Comparing students who read “The Jacket” and “Marigolds,” 54% of 9<sup>th</sup> graders, 41% of 10<sup>th</sup> graders, and 60% of 11<sup>th</sup> graders scored between a 1 and 2 out of the 0–4 scale. This means that only 7% of 9<sup>th</sup> graders, 26% of 10<sup>th</sup> graders, and 11% of 11<sup>th</sup> graders scored in the highest range. For the 9<sup>th</sup> graders who read “The Jacket” and “Eleven” (the two shorter stories), only 15% scored in the top range of 3–4. While we do see some interesting trends with regard to grade level, the vast majority of students were not able to synthesize across texts with sophistication.

*Quality:* Quality of literary reasoning was evaluated with a rubric ranging from 0–4. Of the students who read “Marigolds” and “Jacket,” 28% of 9<sup>th</sup> graders, 16% of 10<sup>th</sup> graders, and 12% of 11<sup>th</sup> graders showed no evidence of literary reasoning. 10<sup>th</sup> and 11<sup>th</sup> graders who read “Marigolds” tended to score higher than 9<sup>th</sup> graders. 42% of 10<sup>th</sup> graders and 48% of 11<sup>th</sup> graders scored 2 or 3, and 6% of 11<sup>th</sup> graders scored a 4, the highest score for this category. These differences were statistically significant. This suggests that skill in literary reasoning does appear to advance across grades.

*Structure:* Structure of argumentation was scored on a scale from 0–5. Again, there was a slight tendency for 9<sup>th</sup> graders who read “Eleven” to score higher on Structure than 9<sup>th</sup> graders who read “Marigolds,” and a tendency for 11<sup>th</sup> graders to score higher than 10<sup>th</sup> graders and 9<sup>th</sup> graders who read “Marigolds.” However, 70% of the 9<sup>th</sup>, 10<sup>th</sup>, and 11<sup>th</sup> grade students who read Marigolds scored below a 3; 50% of the students who read the shorter story set scored below 3. No more than 1 student at any grade level scored a 4 or 5.

We also examined correlations among synthesizing, quality of literary reasoning and structure scores. Synthesizing was moderately correlated with Quality of Literary Reasoning and Structure, while Quality of Literary Reasoning and Structure were strongly correlated with each other.

## **Conclusions from Baseline Study 1**

Our overall take away from the first baseline study was that high school students showed virtually no interpretive reasoning in their essays, given these texts and the framing of the task. Furthermore, providing students with intermediary tasks that asked them to analyze plot and character before focusing on the symbolic and thematic took more time than was allocated for the assignment. Hence, based on the task order counterbalancing we had done we were able to ascertain that approximately at least half of those who had the plot and character tasks prior to writing their essays wrote very little in their essays. Conversely, those who wrote their essays first provided very little for the plot and character analysis tasks.

We also examined the nature of the understanding students seemed to be taking from “Marigolds” and decided, given its length, to conduct a second “baseline” study with different stories. It was also clear from this baseline study that the interventions needed to address three critical aspects of the interpretive process: functions of language in literary texts, identifying and interpreting symbolism, and writing to synthesize arguments across texts

## **2. Baseline Study 2**

Baseline 2 differed from the first baseline study in terms of the grade bands of the participants and the story sets that were used. However, our interests remained understanding how students across grade levels were able to identify what was symbolic and to examine how the author used language in order to convey propositions about the world beyond the text, specifically in terms of a coming of age experience. In order to consider possible developmental trends, we decided to expand the pool in this second iteration to include 6<sup>th</sup> graders and more 12<sup>th</sup> graders. Furthermore, we decided to focus more directly on developmental differences and so decided to use the same text sets across grade levels.

### **Method and Procedure**

The sample consisted of 77 students from the Chicago area, 46 of whom were 12<sup>th</sup> grade students and 31 of whom were 6<sup>th</sup> graders. Two separate sets of texts were selected. The first set consisted of “Eleven” by Sandra Cisneros and “The Butterfly” by James Hanley. The second set consisted of “Flowers” by Alice Walker and an excerpt from “We Were the Mulvaney’s” by Joyce Carol Oates. Twenty three seniors and thirteen 6<sup>th</sup> graders read the first set, whereas twenty three seniors and eighteen 6<sup>th</sup> graders read the second set. We also removed the intermediary tasks for plot and character and went directly to the comparative essay task.

Several shortcomings of the rubric used in Baseline Study 1 led to revamping the rubric to capture a number of dimensions of literary argumentation. This revamping became the basis for the dimensions used in the EBA literature assessment task. Students received rubric scores on categories of argument (claims, how claims were used in the essay, evidence, reasoning); symbolism (attention to, interpretation); rhetorical structure of the essay: organization, language, and structure; and synthesis. (For additional information on the dimensions, see READI Tech Report #10.)

### **Key Findings**

The developmental patterns for the two texts sets and stories were similar in that 12<sup>th</sup> graders generally performed better overall. However, which dimensions of the essay were significantly different between the two grade levels was not identical. In the Eleven/Butterfly text set, for “Eleven,” there was a general developmental trend, with 12<sup>th</sup> graders scoring higher than 6<sup>th</sup> graders on all rubrics except for Evidence and Reasoning. The chi-square for Function of Claims was significant,  $\chi^2(2, N = 36) = 8.96, p = .01$ . The difference between 6<sup>th</sup> and 12<sup>th</sup> graders approached significance for Organization,  $\chi^2(3, N = 36) = 3.89, p = .14$ , and Language and Syntax,  $\chi^2(1, N = 36) = 2.48, p = .11$ . For “The Butterfly,” 12<sup>th</sup> graders scored higher on all rubrics than 6<sup>th</sup> graders. The chi-square analyses demonstrated that this difference was significant for Function of the Claims,  $\chi^2(2, N = 36) = 8.44, p = .01$ . The differences approached significant for the Organization rubric,  $\chi^2(2, N = 36) = 3.89, p = .14$ , the Language & Syntax rubric,  $\chi^2(2, N = 36) = 4.99, p = .08$ , and the Symbolism rubric,  $\chi^2(4, N = 36) = 7.11, p = .13$ .

In the Mulvaney/Flowers text set, for “Mulvaney,” 12<sup>th</sup> graders scored higher on all rubrics than 6<sup>th</sup> graders. These differences approached significance for Coming of Age,  $\chi^2(2, N = 41) = 4.23, p = .12$ , and Symbolism,  $\chi^2(2, N = 41) = 3.44, p = .17$ . For “Flowers,” 12<sup>th</sup> graders scored higher than 6<sup>th</sup> graders on all rubrics. These trends were significant for Function of the Claims,  $\chi^2(3, N = 41) = 7.79, p = .05$ , and approached significance for Organization,  $\chi^2(2, N = 41) = 3.58, p = .16$ , and Evidence,  $\chi^2(3, N = 41) = 6.07, p = .11$ .

Twelfth graders scored higher than 6<sup>th</sup> graders on synthesis for both text sets, however these differences did not approach significance, with the majority of students in each grade level showing little evidence of synthesis.

## **Conclusions from Baseline Study 2**

Overall, we continued to find that students, regardless of grade level, struggled with reasoning through warrants and nested arguments in constructing arguments about the meaning of symbolic targets in the texts. They also struggled with analyzing the rhetorical functions of language, particularly in how the author was using that which was symbolic to forward an interpretive meaning. We decided to pursue the development of the dimensional rubric and the EBA assessment in the context of the interventions. The details on the rubric development are included in READI Technical Report #10. The use of the EBA literature task was used as a pre/post assessment in the iterative design of the literature interventions and results on it are reported in the Technical Report describing those (#10). We dedicated the basic studies to developing more in-depth understandings of how students were processing literary works and the instructional scaffolds, including task instructions, that might support them.

## **3. Think Aloud Study: Spontaneous and Probed Interpretations of Literature**

Given the paucity of research on literary interpretation by adolescents, we conducted studies to inform us regarding (1) how individuals process short stories spontaneously and when specifically probed, and (2) the impact of different forms of essay-writing instructions. Prior research on literary comprehension is limited with respect to adolescents. Researchers do report differences between adult experts and novices reading literary texts: experts produce more interpretations, more sophisticated reasoning about texts, and elaborations on rhetorical devices such as symbolism (Graves & Frederiksen, 1991; Zeitz, 1994). Furthermore, authors often use

rhetorical devices to cue readers that certain aspects of a text may be important, thus guiding attention during reading (Rabinowitz, 1987). However, communication can fail if there is a mismatch between the author’s use and the reader’s awareness of such devices.

We decided to examine how adolescent readers identify and interpret rhetorical devices. We also included a sample of undergraduates as a comparison sample on the specific short stories that we presented to the adolescent readers. We used a prompted think-aloud investigation into the processing of literary texts, focusing on the identification and processing of the rhetorical device symbolism. These foci allowed us to align this work with the intervention design work. As well, we were able to include a sample of students who had participated in the implementation of the symbolism module so we could explore how their approach to the short stories compared to similarly-aged students who had not had the intervention.

### Methods and Procedures

The participants were 20 undergraduates (67% female), 9 seventh graders (67% male), 4 eighth graders (50% male) who had not participate in the READI symbolism intervention and 6 eighth graders (67% male) who had participated in the READI symbolism intervention for two-thirds of a semester. All adolescents were from urban public schools in the Midwest. Undergraduates participated to fulfill a course requirement in introductory psychology.

Participants were asked to read the short story “The Butterfly” by James Hanley (991 words). They received instructions to read out loud and say what they were thinking at the end of each sentence. After reading, participants summarized the story and answered a series of questions. Think-aloud statements were coded for (1) type of processing including elaborations, paraphrases, evaluations, comprehension problems and successes (Wolfe & Goldman, 2005); (2) notice of rhetorical device (i.e., explicit mention of literary device, connection to title, noticing discrepancy and repetition); and (3) presence of interpretation, symbolic or otherwise; and (4) metacognitive activity (monitoring of understanding (positive or negative), strategy specification, revision of understanding. Appendix A contains details on the instructions and coding of the data.

### Key Findings

The results indicated few developmental differences and little evidence of spontaneous interpretive behavior. Specifically, the average length of the think-alouds was between 1000 and 1800 words on average. Undergraduates tended to provide the shortest protocols and the 8<sup>th</sup> grade intervention group the longest. Average time to read and think –aloud was 40 minutes.

The following table reports descriptive statistics for the think-aloud participants.

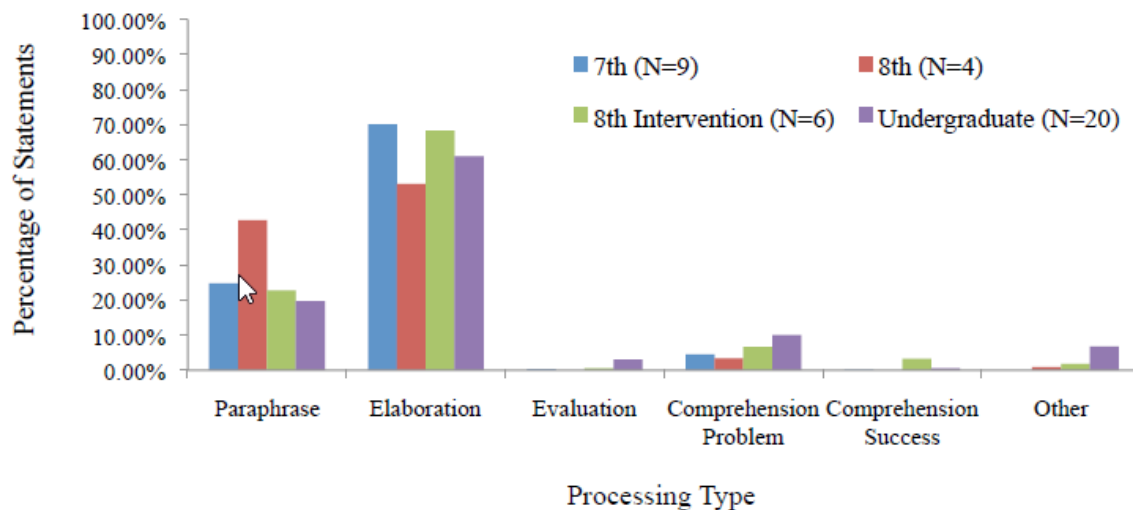
Measure	7 <sup>th</sup> Graders	8 <sup>th</sup> Graders – No intervention	8 <sup>th</sup> Graders - Intervention	Undergraduates
Statements (Mean #)	63.6	83.5	69.2	73.7
Total Words (Mean #)	1214.7	1514.0	1820.0	1009.3
Statement Length (Mean words per sentence)	18.66	17.1	27.9	13.7
Length of think-aloud >	44%	50%	100%	35%



1000 words (% of students)				
----------------------------	--	--	--	--

*Types of processing.* There was little difference among 7<sup>th</sup> graders, Intervention 8<sup>th</sup> graders and Undergraduates in percentage of paraphrases. However, when compared to the other groups, Non-intervention 8<sup>th</sup> graders had a considerably higher percentage of paraphrases than the other three groups. Paraphrasing is a type of processing that reflects relatively superficial meaning making as contrasted with elaborative processing. All groups had a higher percentage of elaborative statements than paraphrases, although the difference was smallest for the Nonintervention 8<sup>th</sup> grade students. The percentage of think-aloud statements that were elaborative was similar across the other three groups. In general, there were very few statements of evaluation, comprehension problems and comprehension successes. Elaborative processing was further examined for evidence of self-explanation, a type of processing generally associated with better understanding of text. Across all four groups of participants, self-explanations constituted more than 70% of the elaborative comments, with surface text connections comprising the bulk of the remainder.

The following figure provides a graphic representation of the types of processing.



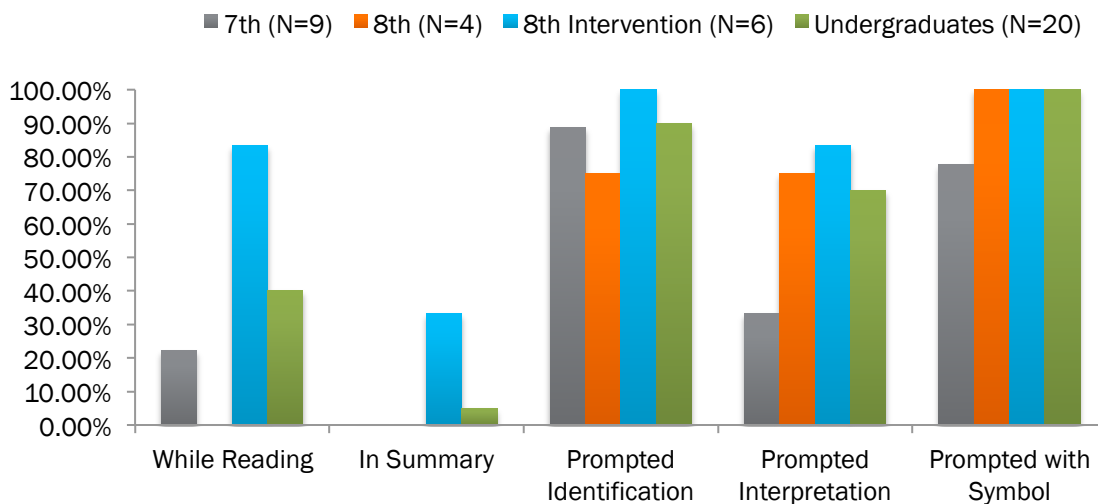
*Rhetorical Devices and Interpretations.* In general there was not a lot of thinking aloud about either rhetorical devices or interpretation. Regarding rhetorical devices only 2 participants explicitly mentioned a literary device during the think aloud (one 7<sup>th</sup> grader mentioned personification and one 8<sup>th</sup> grade Intervention student mentioned both alliteration and personification). In addition, the intervention 8<sup>th</sup> graders (4 out of 6) are the only participants in any of the groups who made connections to the story title during the think-aloud sessions. Furthermore, there were only 4 non-undergraduate participants who exhibited notice of repetition during think aloud sessions (1 7<sup>th</sup> grader and 3 Undergraduates). However, 70% of the undergraduates (14 out of 20) reported at least one notice of repetition during think aloud sessions. Noticing discrepancy was the most frequently coded literature-related category. Overall, the majority of 7<sup>th</sup> and non-intervention 8<sup>th</sup> graders did not make comments indicating

that they noticed discrepancies. In contrast, approximately 67% (4 out of 6) of the intervention 8<sup>th</sup> graders and 90% (18 out of 20) of the undergraduates made comments regarding notice of discrepancies during their think-alouds.

Overall, there was little evidence of interpretation during the think-aloud sessions. For example, only fourteen participants (out of 39 total) provided interpretations of symbols (7<sup>th</sup>: 2 of 9; 8<sup>th</sup> Intervention: 4 of 6, Undergraduates: 8 of 20). Only 4 participants (1 Undergraduate and 3 Intervention 8<sup>th</sup> graders) made other interpretive comments (i.e., those not related to symbolism) during the think-aloud session.

*Spontaneous and Probed Interpretations of Symbolism:* Although during the reading/think-aloud, there were few un-prompted statements of symbolic interpretation (approximately 20 out of 2,659 statements) during reading, indicating that interpretations were generally not stated. However, these data do not necessarily indicate that participants are incapable of making interpretations; they indicate that they do not report them while reading. To pursue what kind of interpretations that might be able to make, we investigated what interpretations were provided after reading, either during summarization, or when they were specifically prompted through a direct question to identify a symbol in the story, and to interpret an experimenter-provided symbol from the story.

The following figure provides rates of symbolic interpretation produced spontaneously and when probed.



The comments in the think-alouds indicated that students rarely offered interpretations during reading, replicating previously reported differences between novice readers of literature and experts. Experts' comments during reading mention noticed features of the text (see Rules of Notice codes in Appendix A). As well, they frequently offer interpretations during reading and make comments that tie their interpretations to the specific language the author has put into the text. However, the data from the adolescents and undergraduates indicate that when specifically asked to make interpretations they were more than capable of doing so. This finding suggested that major issues to tackle in the instructional interventions were (1) raising awareness of the

definitional role of interpretation in literary reading and (2) providing opportunities for students to develop the literary reading practices that attuned them to noticing discrepancies, repetitions, potential symbols and other rhetorical devices with importance for interpretations. In addition the findings from the 8<sup>th</sup> graders who had participated in an initial READI intervention designed to foster literary reading suggested that the early designs were headed in the right direction. We thus pursued both foci in the subsequent iterations of the designs. In the basic studies we explored various instructional techniques for fostering and guiding interpretative practices.

#### **4. Do story titles cue interpretation?**

The purpose of this study was to investigate the impact of the title of a story on readers' literal and interpretive understanding. According to literary theorists, titles are an example of text that is in a *privileged position*. Privileged position is one of the rules of notice that Rabinowitz (1987) postulates guide readers attention to potential text elements that may be important to go beyond the literal meaning to construct interpretative meaning. Indeed, research in memory and discourse processing research has shown that titles affect what readers pay attention to in texts, how they comprehend them, and the way they are remembered (Anderson & Pichert, 1978; Bransford & Johnson, 1972; Daniel & Raney, 2007). This study was an initial effort to explore the impact of titles on literal and interpretive meanings that adolescents constructed for short stories. We manipulated the title of two literary short stories to ascertain the impact of the title on the ongoing development of meaning during reading and well as once the story had been read. We chose two stories that afforded the possibility of multiple titles due to the existence of multiple prominent characters, objects, or events in the stories.

#### **Methods and Procedure**

Eighth-grade students (N = 92) were assigned to one of four reading conditions: 1) *The Butterfly* by James Hanley presented with its original title, 2) *The Butterfly*, but presented with the manipulated title *The Caterpillar*, 3) *The Flowers* by Alice Walker presented with its original title, or 4) *The Flowers*, but presented with the manipulated title *The Dead Body*. Both stories are thought to convey a theme of coming of age and feature prominent symbolism. In *The Butterfly*, there are only two mentions of a butterfly; however a caterpillar figures prominently in the concluding episode of the story. In *The Flowers*, flowers play a significant role in the beginning of the story and set a playful mood. In the concluding episode of the story, the mood changes when the main character trips over a dead body and the remains of a noose. In both stories the original and the manipulated titles thus could take on symbolic significance. Our main interest in this study was whether the title biased the literal and/or interpretive meaning that readers constructed.

Students read printed versions of the stories from a booklet in a self-paced reading task that was completed in a single class period of approximately 45 minutes. At three points during reading (after the first paragraph, approximately halfway through the story, and at the end of the story), the story contained a prompt that asked students a question they were told to respond to at that point in the story reading: *What do you think the story is about?* Each story was segmented on different pages in the booklet to provide space for writing the response to each question prior to reading the rest of the story.

After reading, students were asked to respond to five additional questions that tapped into their overall interpretations as well as their understanding of the importance of titles in literary reading,

- 1) *What does this story tell you about the world we live in?*
- 2) *Why do you think the author chose [title] as the title?*
- 3) *What makes [title] a good title for this story?*
- 4) *What makes [title] a bad title for this story?*
- 5) *If you were the author, what would you have titled the story? Why?*

## Key Findings

Responses to the question *What is this story about?* were regarded as reflecting the literal meaning that readers constructed. Two coders coded the responses to this question for mention of “dead body” (or semantically similar terms, such as “dead man”) and of “flower(s)”. The raters reached 98.9% reliability and the single discrepancy was reconciled through discussion.

Responses to the question *What do you think this story tells us about the world we live in?* were regarded as reflecting the interpretive meaning that readers constructed. Responses to this question were coded for thematically-related terms or phrases. For “dead body”, we noted reference to “death”, “kill”, and “murder”. For “flowers” we noted reference to “nature” and specific flowers like “rose”. Raters first-pass coding achieved reliability of 100%.

The first set of analyses explored whether or not responses to the literal meaning question *What is this story about?* at the end of reading included title-related information. Chi-square analyses revealed a significant association between the title and mention of dead body,  $\chi^2(1, N = 44) = 6.15, p < .02$ , where participants given the title *The Dead Body* had more mentions of dead or death than would have been predicted by chance. There was no significant association between title and mention of flowers,  $\chi^2(1, N = 44) = 1.63, ns$ .

The second set of analyses looked at the interpretive meaning responses. The title manipulation affected the interpretation of the text, but only for mentioning dead body/man. Chi-square analyses indicated a significant association between title and mention of death,  $\chi^2(1, N = 44) = 6.20, p < .02$ , but no significant relationship between title and mention of flowers in the interpretation,  $\chi^2(1, N = 44) = .41, ns$ .

Interestingly, responses from participants in *The Dead Body* condition seemed to have more emotion words in them. To explore this, Linguistic Inquiry and Word Count (LIWC; Pennebaker, Booth, & Francis, 2007) was employed to identify emotion words in the students’ interpretations. The analysis revealed no difference in the amount of positive emotion words,  $t(42) = -.97, ns$ , but that the title *The Dead Body* led to more negative emotion words ( $M = 4.32, SD = 5.15$ ) than the original title, *The Flowers* ( $M = 1.40, SD = 2.63$ ),  $t(42) = -2.40, p < .05$ . This finding is consistent with other findings that demonstrate a relationship between affective evaluation and constructed interpretations (Levine & Horton, 2013).

## 5. Affect Priming and Literary Interpretation

The impact of title on interpretation suggested that title did indeed affect the literal representation of the text. Of greater interest is that the title manipulation also appeared to have affected the emotional valence of the interpretative meaning. That is, the title may have been directly priming the reader's affect, and not simply activating prior knowledge. To test this possibility, we ran an additional study in which we conducted an "emotion priming" study using one of the stories from the title studies, *The Flowers*. We removed the title from the story, and used positive, negative, and neutral images to prime an emotion prior to reading. All other aspects of the design were the same. We predicted that the reader's affect would determine which parts of the text the reader would attend to most (the reader's literal representation) which would, in turn, affect the nonliteral interpretation of the story's meaning.

### Methods and Procedure

Participants were a "convenience" sample of 61 undergraduates. In a between subjects design, they were primed with one of three affectively-valenced images (positive, negative, neutral) and then read the short story *The Flowers*. To assess their literal representation, participants were asked to summarize the story. To assess their interpretive representation, participants provided their own interpretation of the story and then ranked five possible (and plausible) story themes from most (1) to least (5) appropriate. Finally, participants were asked if both *The Flowers* and *The Dead Body* would be good titles for the text and to justify their answer.

We predicted that readers would be more attentive to sentences in the text that are congruent with their induced affect. We used Linguistic Inquiry and Word Count (LIWC), to identify emotion words in participants' interpretations. We predicted an affective-consistency bias, most evident in the negative condition, where those who saw a negative priming image would yield more negative words in their summary and theme responses.

### Key Findings

Our manipulation check indicated that the neutral image invoked negative emotions, so this condition was omitted from analyses. Preliminary analyses explored the responses to the summary and interpretive question prompt. We ran a 2(image: positive prime, negative prime) x 2 (LIWC: positive words, negative words) mixed design ANOVA with number of emotion words in the summary response as the dependent variable. This indicated no main effect of image,  $F < 1.00$ , but a main effect of emotion word, such that all participants, regardless of prime condition, included more positive words in their summary,  $F(1, 37) = 7.76, p < .01$ . There was no interaction.

Another 2x2 ANOVA was conducted on the number of emotion words in the interpretation response. This analysis showed a similar pattern in which there was a main effect for emotion word, with positive words being more prevalent than negative words,  $F(1, 37) = 4.45, p < .05$ . There was no main effect of image nor an interaction,  $F < 1.00$ . Additionally, a

chi-square analysis indicated no effect of image on the selection of one of the five provided themes,  $\chi^2(1, n = 38) = 5.45, ns$ .

Despite the text being identical for all participants, we believed that reader's affect would change how the text is understood both literally and interpreted. However, our results indicate that the emotion induced by aspects of the text itself washed out any effect of the affective prime. The results suggest that inherent features of the text play an important role in literary meaning making above and beyond specific aspects of the reader. We believe these findings suggest that, in the previous study, it was not simply that the title *The Dead Body* made the readers' feel more negatively. Instead, it seems that this title helped readers to make a more accurate literal representation of the text, which resulted in a more negative representation of the text as a whole. We are currently analyzing the sentence verification responses and the participants' responses to the appropriateness of each title to better understand how the affective prime may have altered their representations of the story. Results from this study were presented at the 2015 UIC Student Research Forum and the Midwestern Psychological Association Conference..

## 5. Beyond Symbolism: Situational Irony

The Common Core standards identify reasoning about irony as an important skill for English Language Arts, and yet there is a lack of research about how secondary students reason about situational irony. By situational irony, we refer to events in which actions have the opposite outcome from what was intended. In this study, we investigated how adolescent students in 8<sup>th</sup> grade reasoned about situational irony compared to an adult sample. We hypothesized that we would observe age-related differences such that adults would find ironic comics more humorous than adolescents, and also make more appeals to irony than adolescents when justifying their reasons for their humor ratings.

### Methods and Procedure

We recruited 118 eighth grade students for our adolescent sample and 103 college students for our adult sample. We obtained informed consent from our adult sample, and our adolescent sample participated with their assent and with the consent of their parents, teachers, and schools.

We selected eight comic strips for this study. We chose comic strips as opposed to short stories because comic strips afford the opportunity to examine a greater number of scenarios than would be possible when exploring prose fiction. A group of experts in English Language Arts rated these comics for anticipated humor, level of situational irony, how likely adolescents were to get the humor, and how appropriate the subject matter was for adolescents. Our eight comics were equivalent in terms of anticipated humor, likelihood of getting the joke, and appropriateness of subject matter, and differed only in their level of irony such that we obtained 4 situationally ironic comic strips and 4 situationally unironic comic strips.

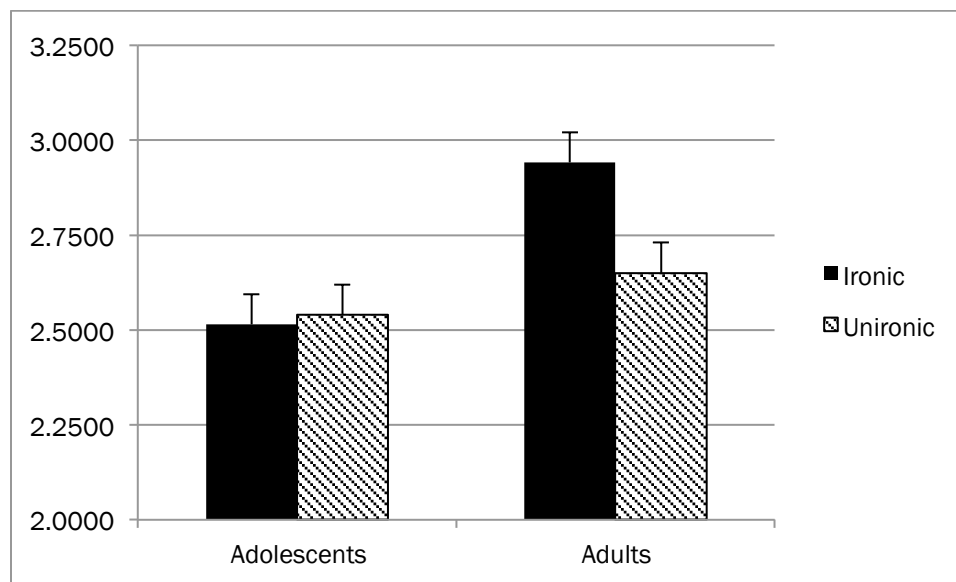
Adolescents and adults read the comic strips and rated each comic strip for how funny it was on a scale of 1 (not funny at all) to 5 (extremely funny). After rating the humor of each

comic strip, participants provided reasons as to why they found the strip to be funny or not funny.

### Key Findings.

**Humor Ratings.** In examining the humor ratings for ironic and unironic comics, we had two major questions: 1) What are the differences between age groups in how people perceive humor in irony? and 2) Within each age group, do participants rate ironic comics as funnier than unironic comics?

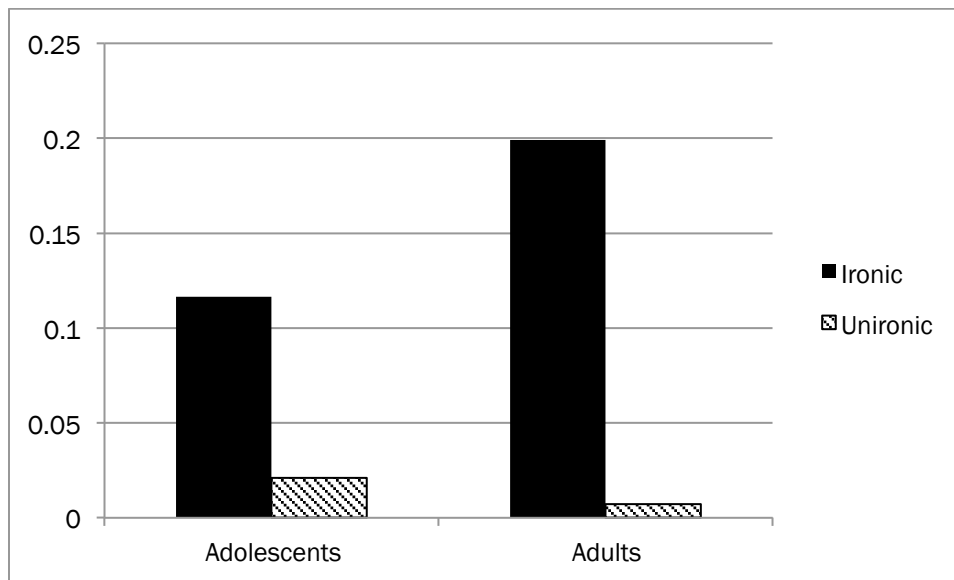
To investigate these questions, we conducted a 2x2 ANOVA with a within subjects factor of irony (ironic, non-ironic), and a between-subjects factor of age (adolescent; adult), with humor ratings as the dependent variable. See Figure X for means and standard errors. This revealed a significant interaction between irony and age,  $F(1, 219) = 7.12, p = .008$ . To further understand this interaction, we conducted a series of planned follow-up comparisons.



*Differences between age groups.* We investigated how adolescents and adults rated ironic and unironic comics using independent samples  $t$  tests. The critical value for these tests was lowered from .05 to .025 to adjust for multiple comparisons. This revealed that adults rated the ironic comics as significantly funnier than did the adolescents,  $t(219) = 3.64, p < .001$ . However, there were no differences between how adolescents and adults rated unironic comics,  $t(219) = 1.03, p = .31$ .

*Differences within age groups.* Within each age group, we investigated how individuals rated ironic and unironic comics using paired samples  $t$ -tests. The critical value was lowered from .05 to .025 to adjust for multiple comparisons. There was no difference in how adolescents rated the ironic and unironic comics,  $t(117) = .302, p = .763$ . However, adults rated the ironic comics as significantly funnier than the unironic comics,  $t(102) = 3.52, p = .001$ .

**Reasoning About Irony.** We explored participants' reasoning about irony by examining the proportion of participants in each age group and irony condition made explicit appeals to irony when explaining why they found the comics funny or unfunny. To investigate this, we conducted a 2x2 ANOVA with a within subjects factor of irony (ironic, non-ironic), and a between-subjects factor of age (adolescent; adult), and the average rate of appeals to irony as the dependent variable. See Figure X for more information on the proportion of participants appealing to irony. There was a significant interaction between age group and irony,  $F(219) = 10.77$ ,  $p = .001$ . To further understand this interaction, we conducted a series of planned follow-up comparisons.



**Differences Between Age Groups.** We conducted independent-samples t-tests for the rates of appeals to irony in both the ironic and the unironic conditions. To correct for multiple comparisons, we lowered the critical  $p$  from .05 to .025. Adults were more likely than adolescents to make appeals to irony when reasoning about the ironic comics,  $t(219) = 2.85$ ,  $p = .005$ . Interestingly, there was a noticeable trend for adolescents to make appeals to irony more often when reasoning about the unironic comics,  $t(219) = 1.757$ ,  $p = .08$ .

**Differences within age groups.** Both adults and adolescents made appeals to irony when reasoning about the ironic comics than when reasoning about the unironic comics ( $p < .001$  in both cases).

## 7. Beyond Symbolism: Comprehension of Story Parodies

Parodies are works that imitate the character or style of a serious work of literature or art, often for humorous effect. For example, the parody *Bored of the Rings* mocks J.R.R. Tolkien's *Lord of the Rings* by replacing Tolkien's complicated writing style and lofty topics by using short, colloquial prose and frequently absurd situations. Parodies are ubiquitous in our society, and yet there is a dearth of empirical research on how readers comprehend or reason about parody.



Parodies provide an especially fertile ground for studying how readers reason about multiple documents in literature. This is because readers likely must maintain a representation of both the parody and the work being parodied (i.e., the target) for successful comprehension. The overlap between a parody and its target may serve as a retrieval cue when readers comprehend parody, but it is unclear what sorts of features (i.e., subject matter or writing style) provide sufficient cues for retrieval.

## Methods and Procedures

In this study, we recruited 123 undergraduate students from an urban Midwestern University. All participants received course credit in exchange for their participation, and all were proficient speakers of English.

To test participant's reasoning about parody, we selected a short story, "Hills Like White Elephants" (aka "White Elephants") by Ernest Hemingway, and a parody of that story, "Hills Like White Hills" (aka "White Hills") by Kevin Catalano. "White Hills" has been featured in the International Imitation Hemingway Competition, an annual writing contest for Hemingway parodies. Similar to "White Elephants," the parody text features a man and a woman drinking in a Spanish bar, and discussing the woman's pregnancy, and thus the two stories share similar topic matter. Both stories also feature similar writing styles: short paragraphs and sentences, which are indicative of Hemingway's writing style (Whissel, 1994), and both stories feature a high proportion of dialogue.

We also selected three other short stories which shared a similar writing style, topic, or neither style nor topic with "White Elephants." To match the topic of "White Elephants" we chose "Moseley," an excerpt from William Faulkner's *As I Lay Dying*, in which a young woman expresses hesitation about her pregnancy. Thus, this story shares a central topic with "White Elephants." Importantly however, the writing styles between these two stories are very different, with Faulkner's story contains long sentences and paragraphs, and a lower proportion of dialogue than either "White Elephants" or "White Hills. To match the writing style of "Hills Like White Elephants," we selected another Hemingway short story, "An Old Man at the Bridge." This story matches the writing style of "White Elephants," featuring short sentences and paragraphs and a high proportion of dialogue. However, the story is about a war refugee, and so it does not share the same topic as "White Elephants." Finally, we selected a story with neither thematic nor stylistic overlap to "Hills Like White Elephants" as a control text. This control text, "War" by Sherwood Anderson, is about two war refugees, and thus does not share the topic of "White Elephants." "War" also features long sentences and paragraphs, but very little dialogue, as opposed to the shorter, dialogue-heavy sentences of "White Elephants."

Using these short stories, we created four conditions: *Style and Topic Overlap*, *Topic Overlap Only*, *Style Overlap Only*, and *No Overlap*. Each condition featured the parody text "White Hills." In the Topic and Style Overlap condition, this story was paired with "White Elephants." In the Topic Overlap Only condition, "White Hills" was paired with "Moseley." In the Style Overlap Only condition, "White Hills" was paired with "An Old Man at the Bridge." And in the No Overlap condition, "White Hills" was paired with "War."

To evaluate students reading abilities, we employed an updated version of the Author Recognition Test (ART; Acheson et al., 2008). In this test, participants view a list of names, some of which are authors. Participants must select the names they know for certain are authors, and are penalized for incorrect responses. The author names in this version of the ART have been updated from the original version (Stanovich & West, 1989) to account for authors that current participants are likely to have encountered. Scores on this test are positively correlated with scores on both the English and Reading portions of the ACT and with self reports of reading frequency and enjoyment of reading.

Participants began the experiment by reading both stories. In each condition, the parody story “Like White Hills” was always presented last in order to test the prediction that echoes from previously read stories should influence perception of parody. Participants read the stories at their own pace, and were allowed to take notes or otherwise annotate the texts. Upon completion of the stories, participants rated how easy, interesting, and funny each story was, on a scale of 1 (not funny/interesting/easy at all) to 7 (extremely funny/interesting/easy). Participants then responded to a short essay question about parody. For this essay, students were given a brief description of parody and were then asked if they felt that one of the stories was a parody of the other, and why. Finally, participants completed the Author Recognition Test.

### Key Findings

A Chi square showed that detection of the parody was significantly related to condition,  $\chi^2(3, N = 123) = 22.85, p < .001$ . Overall, we see that the largest proportion of participants who claimed “White Hills” was a parody occurred in the Style & Topic Overlap condition, followed by the Style Only and Topic Overlap conditions, with the smallest proportion being found in the Control condition.

#### *Percentages of participants identifying “White Hills” as a parody in each condition*

Condition	Percent Detecting Parody
Topic & Style Overlap	87.9%
Topic Overlap Only	55.2%
Style Overlap Only	56.7%
No Overlap (Control)	29.0%

We then conducted follow-up Chi-squares on each pair of conditions to further explore the relation between condition and parody detection. For these followups, we adjusted the alpha level to .044, following Keppel’s (1991) suggested correction for Chi-square follow-ups. These followups revealed that the relation between parody detection and condition was significant between the Story & Topic Overlap condition and the Topic Overlap Only condition,  $\chi^2(1, N = 62) = 8.29, p = .004$ ; between the Style & Topic Overlap Condition and the Style Overlap Only condition,  $\chi^2(1, N = 63) = 7.78, p = .005$ ; and between the Topic & Style Overlap condition and the Control condition,  $\chi^2(1, N = 64) = 4.21, p = .04$ ; and between the Topic Overlap Only condition and the Control condition  $\chi^2(1, N = 60) = 4.21, p = .04$ ; and between the

Style Overlap Only condition and the Control condition  $\chi^2(1, N = 61) = 4.76, p = .03$ . However, there was no relation found when we compared the Topic Overlap Only and the Style Overlap Only conditions,  $\chi^2 < 1$ .

Results suggest that participants understood the parody's intent most often when it was paired with its target and least often when the parody was paired with an unrelated text. Interestingly, participants identified the parody's intent more often when paired with a story with a similar writing style or topic compared to the unrelated condition. However, when the parody was related to the second text only by writing style or topic, most participants indicated that the parody was a parody simply because it was "funny" instead of appealing to the overlap in writing style or topic. This suggests that interventions aimed at improving students' reasoning about parody may benefit from teaching general rules of the structure of parodies, to help students better identify them. This study is being written up for publication (Briner, Goldman, & Magliano in preparation).

## **8. Technology-based tools for strategic reading assessment: RSAT and Read & Answer.**

Our original plans called for exploratory work with two tools, for assessing moment to moment processing during task-oriented reading, RSAT (Magliano, et al., 2011) and Read&Answer (Vidal-Abarca, & Cerdán, 2013), RSAT is a tool developed under a previously funded IES project (R305G040055). It measures moment-to-moment processing by having participants read texts one sentence at a time and periodically respond to indirect questions (what are you thinking now?), similar to verbal think – aloud protocols. Responses to these indirect prompts expose comprehension processes such as paraphrasing, bridging, and elaborating, and summarizing. Read&Answer is a tool that tracks readers' movement through the use of a "masking" technique. That is, to reveal actual words or sentences, readers "click" on the word or sentence they want to read. When they click to see another segment of the text, the one they have read is masked again. In this way, a reader's inspection of the text is tracked, similar to eye tracking methodologies. Readers can click on any segment in any order. The researcher sets the "unit" that is masked and unmasked per click. It can be letter, word, phrase or sentence. Data on what and in what order text was unmasked and for how long is recorded for analysis. The reading can be done apart from specific questions or questions can be posed and reading in response to those questions can be tracked to reveal search and reading processes in the context of specific kinds of questions. Computer algorithms operate on the output to generate estimates of inferencing and comprehension processes.

We conducted a series of basic studies using these tools on convenience samples of college students at NIU or in regional community colleges. The entering skill levels of these students are typically similar to average and struggling high school students.

*Understanding the differences in text processes between developmental and non-developmental college students.* Unfortunately, many students enrolled in universities and community colleges leave high school underprepared to meet expectations for reading literacy in their post high-school courses. As such, many students are referred to programs, called Developmental Education (DE) programs, intended to bolster their reading literacy skills.

Referral criteria are usually based on performance on high stakes, summative tests of comprehension skills. These assessment tools do not provide information about what students do as they read, nor are they well suited as formative assessments. In several of these exploratory studies, we used RSAT to assess whether there are differences in processing of text between students enrolled in DE programs and comparable freshmen admitted to college through traditional criteria and therefore not enrolled in compensatory courses (Boonthum, & Magliano 2011; Chan, Ray, Armstrong, & Magliano, 2013; Magliano, 2011). Specifically, the computational linguistic algorithms that RSAT uses to analyze students open-ended responses were used to provide assessments of readers engaging in bridging and elaborative inferences during reading. Importantly, while DE and non-DE students did not differ in bridging scores, DE students had lower elaboration scores than non-DE students. These experiments suggest that this population of students may not be using relevant background knowledge when reading informational texts. We are currently conducting a replication study involving community college students.

In support of this endeavor, we have also conducted several studies exploring new approaches for classifying students based on the range of strategies that they demonstrate. We have learned that combinations of strategies that demonstrate that readers are establishing coherence (e.g., paraphrasing and bridging) are more indicative of positive learning outcomes than when readers tend to rely on one strategy when taking RSAT (Boonthum, & Magliano 2011; Boonthum, McCarthy, Jackson, Magliano, McNamara, 2010; Magliano, 2010). This research has been reported at the Society for Computers in Psychology.

*Exploring changes in bridging and elaborative processes during reading.* This study investigated how comprehension processes dynamically occur over the course of reading a text (Magliano, Durik, Holt, 2012). We examined how elaborative and bridging inference processes changed over the course of moment-to-moment processing of a text. Participants produced concurrent verbal protocols while reading expository texts in the context of the RSAT. A novel application of growth curve modeling was used to assess how these processes changed over the course of the texts. The results showed that elaborations are high during the initial phases of reading a text, but then decline over the course of the text, whereas bridging inference increase over the course of a text. The results are interpreted in terms of the stages of structure building during reading. Reports of this work were presented at the annual conferences for the Psychonomics Society, Society for Computers, the American Educational Research Association, and the Society for Text and Discourse.

*Exploring the role of bridging skill on task-oriented reading.* It is widely acknowledged that reading in academic settings is typically grounded in a task or purpose. One common form of grounding reading in a task is to have questions accompany the texts. This is certainly the most common approach for assessing comprehension skill on standardized tests. In these contexts, the text is available to the reader when answering the questions and readers can adopt a variety of test-taking strategies that have a dramatic affect on how and when they read the text. For example, readers can choose to skim the text, to read deeply, or to not read at all before turning to the questions. In contrast, students preparing for a closed-book quiz on an assigned textbook chapter have no option but to engage in a close reading of the text to prepare for the exam if they want to perform at an optimal level.

When a text is available to the test taker, do they really need to comprehend the text to answer questions at an optimal level? The goal of the present study was to assess the extent that a reader's drive for coherence was related to performance on comprehension questions as a function of whether or not the text was available when answering the questions (Higgs & Magliano, 2012; Higgs, Magliano, Vidal-Abarca, & McNamara, 2012). Participants were administered RSAT to obtain a measure of a readers drive for coherence, as operationalized by the RSAT bridging score. Read & Answer was used to assess how students engaged in task-oriented reading. Specifically, participants read science and history texts and then answered short answer questions that tapped their understanding of relationships between discourse constituents. We manipulated whether participants were allowed to search the texts when answering the questions.

We found that RSAT bridging scores were equally and robustly correlated with performance on the comprehension questions when the text was available and when it was not available when answer the questions. We interpret these results as indicating that comprehension is important regardless of whether students can search the text for answers. These finding have important implications for high stakes tests because students are often advised to not bother reading the text for comprehension, but rather to read the questions and search for text segments that provide the answers (Rupp, Ferne & Choi, 2006). This study indicates that this strategy is not optimal and will likely lead to inferior performance. Reports of this work were presented at the annual conferences for the Psychonomics Society, Society for Computers in Psychology, and Society for Text and Discourse.

## **II. Supports for Literary Comprehension and Interpretation**

We conducted a number of studies to explore the use of particular instructional techniques to make the why and how of literary reading more explicit to students. Basic studies of supports focused on prior knowledge of interpretive devices, task instructions, instructional scaffolds, and affective evaluation.

### **1. Prior Knowledge and Literary Argumentation**

Building on prior knowledge is a form of scaffolding in that students' prior knowledge and experiences provide a framework for processing canonical texts (Lee, 2007). Lee's (2007) Cultural Modeling is an approach to literary analysis that involves using culturally relevant media to make students aware of meaning-making processes and ways of thinking that they use to make sense of their everyday worlds. Having been made visible and explicit, these processes and ways of thinking can then be referenced in other contexts, In the present contexts, students are encouraged to use them in the context of literary analysis of canonical texts (Lee, 2007; Morrell, 2009).

Implementing the Cultural Modeling approach requires identifying a cultural data set that involves the same kinds of processing as the literary work students will be interpreting subsequently. One critical assumption is that cultural data sets are familiar to the target students, enabling them to readily identify processes they use that can then be transferred to processing

rhetorical features of literary works. In this way, students have a means of using their prior experiences to engage in deep, critical literary analysis. The study discussed here tested the familiarity assumption by manipulating the familiarity of the materials that preceded a poetic interpretation task.

## Methods and Procedure

Participants were college students with demographics similar to those of graduating high school seniors from urban areas. Participants in the treatment conditions were exposed to a cultural data set prior to a task that required them to identify and interpret the poem, *Don't Go Gently into the Night*, by Dylan Thomas. The cultural data sets were songs that varied in familiarity: familiar song (current song on the pop charts), less familiar song (Number 1 song from approximately 20 years ago), least familiar (a poem that was described as lyrics to an old R & B song). In a fourth condition (control), participants read instructions on the nature of symbols in literature. A total of 110 participants were randomly assigned to the four conditions as follows: Familiar = 32, Less Familiar = 27; Least Familiar = 31; Control = 20.

## Key Findings

We assessed the impact of the cultural texts on the literary analysis of *Don't Go Gently into the Night* in a variety of ways, but most critically we assessed the extent to which participants identified the canonical interpretation. The results were most consistent with a *recency hypothesis* that assumes that the effectiveness of a cultural data set is a function of familiarity: Participants were more likely to identify the canonical interpretation if they were provided with the familiar or less familiar song than if they were provided with the least familiar song or the information on symbols in literature. These means and standard deviations are presented in the table below.

	Dependent Measure		
	Total Symbols	Symbols repeated in text	Symbols not based in text
Familiar (n = 32)	22.45(4.85)	11.62 (5.70)	.53 (1.24)
Less Familiar (n = 27)	18.03 (6.03)	8.37 (3.31)	.55 (1.55)
Least Familiar (n = 31)	19.77 (5.83)	8.81 (5.38)	1.03 (2.91)
Control (n = 20)	19.25 (5.83)	7.85 (4.54)	2.25 (4.70)

An ANOVA on the total number of symbols revealed a main effect of familiarity ( $F(3, 103) = 3.87, p = .011$ ), with posthoc tests indicating that more symbols were produced in the recent song condition than in the other three conditions. ANOVA on the total number of repeated symbols mimicked this pattern: a main effect of familiarity, ( $F(3, 103) = 4.94, p = .003$ ), with posthoc tests showing more identification of symbols repeated in the poem for participants in the recent song condition than in the other three. Finally, the data suggest an interesting and unexpected trend for non-text based symbols: More were produced in the control condition than in the familiarity conditions

The following table presents the frequencies of canonical interpretations by condition:

	Interpretation			
	None	Partial	Full	% Full
Familiar (n = 32)	4	14	14	45%
Less Recent (n = 27)	3	9	15	55%
Least Recent (n = 31)	8	13	10	32%
Control (n = 20)	9	7	4	20%

To further test these effects, 4 x 2 Chi-square was conducted. There was a significant relationship between condition and Full versus the combined Partial/None categories,  $X^2 = 13.95$ ,  $p = .03$ . Participants in the least recent and control conditions were less likely to produce the canonical interpretation than the other two conditions.

These results are consistent with claims that Cultural Modeling is most effective when cultural data sets are familiar to students. In terms of instructional implications, these results support the design principle of including cultural data sets in the READI interventions. However, the familiarity effect also implies that the exact content of cultural data sets is not only a function of the focal rhetorical device but need to be adapted to reflect the sociocultural context of the students. Cultural data sets appropriate for older as compared to younger students will not necessarily be the same. Likewise, cultural data sets will vary with respect to their appropriateness for different ethnic groups.

## 2. Instructional Orientation to Literary Interpretation: What are symbols telling us?

This study manipulated how students were instructed in literary interpretation, specifically regarding symbolic and thematic interpretation. Previous research and Project READI classroom observations suggest that typical instruction in literary interpretation relies heavily on declarative knowledge, providing students with static definitions of rhetorical devices such as symbols and surface level identification procedures. This labeling and procedural identification of literary techniques tends to be isolated from any relationships such techniques might have for thematic level understanding and student interpretations. Project READI's work sought to make that relationship explicit for students by scaffolding their understanding of authors' craft – the practices authors of literature use to communicate their worldviews and messages. In this quasi-experimental study we compared typical instruction in symbolism to instruction that made explicit the connection between symbols and interpretations in authors' craft.

### Methods and Procedure

Two 7<sup>th</sup> and two 8<sup>th</sup> grade intact classrooms of students participated in the study. At each grade level, one of the classrooms was randomly assigned to the typical instructional condition (n = 12 consented students in 7<sup>th</sup> grade and 21 in 8<sup>th</sup> grade) and one to the explicit craft condition

(n = 16 consented students in 7<sup>th</sup> and 17 in 8<sup>th</sup> grade). In the typical instruction students were taught the definition of the literary term “symbol” and taught to look for possible symbols in a literary text. They were then asked to interpret possible themes of the text. In the explicit craft instructional condition, students were not asked to define or look for symbols, but instead discussed ways in which authors use language to communicate worldviews and big ideas about human nature. Students then practiced constructing thematic interpretations of literary texts by looking at textual cues that might contribute to overall meaning, many of which were symbols.

The study took place over three days, however the instructional portion was limited to only a single day due to constraints on the amount of classroom time the teachers were willing to allocate to the external research. On day 1, both the typical instruction and the explicit-craft classrooms completed a pre-test. Instruction occurred on Day 2 and on Day 3 a post-test was administered. On Day 2, both classes received about an hour of instruction (typical or explicit craft) with two READI researchers (both formerly high school English Language Arts teachers) doing the instruction, and using discussion and small group work. On Day 3, both groups were given a post-test. The pre- and post-test texts were counterbalanced within classroom. Each researcher-teacher taught one typical class and one explicit-craft class, crossed with one 7<sup>th</sup> and one 8<sup>th</sup> grade class each. Although grade taught and instructional condition were confounded, this “counterbalancing” method at least somewhat controlled for “quality and style” of teaching. Both pre- and post-tests provided students with a story for which they were prompted to write about their interpretation of the message of the story and how the author communicated it. Two different short texts were counterbalanced for students so that each student wrote about a different story at pre and at post; for instructional condition within grade so that about half of the students in each instructional condition at each grade level worked on each story at pre and the other at posttest.

## **Key Findings**

Analyses of the essays indicated that students were responsive to the instruction in that the emphasis during the instruction was clearly reflected in the organization and content of the posttest essays. Note that pretest essays reflected reiterations of plot and story summaries across both conditions and grade levels. In contrast, at posttest, the students who participated in the typical instruction tended to name a number of symbols and provide evidence similar to that discussed during instruction (e.g., it’s repeated a lot; it’s the title). On the other hand, the majority of essays in the explicit craft instruction began with a statement of the message/theme of the story and then referred to specific content in the story as evidence that connected the evidence to the claim, albeit with limited reasoning. This was an encouraging finding considering the very short duration of the “instruction” and limited opportunity to practice.

### **3. Effects of Sentence Stems on Writing About Literature: Study 1**

A significant challenge to students in ELA classes is the move from discussion to written articulation of literary interpretations. In general, this challenge may arise from a general lack of attention to writing instruction (Hillocks, 2005) and lack of integration between different events in an ELA classroom, such as discussion and writing (Langer, 1999). Another general source of this challenge may be students and teachers’ uncertainty about the nature of literary



interpretation. For example, research suggests that some teachers and students still view thematic interpretation as a process of identifying a single message suggested by a text (Applebee, 1993; Grossman, 1990; Hamel, 2003).

READI teachers and researchers faced those same challenges when designing modules to help students write about literature. In particular, the READI team found that students struggled to frame and articulate their ideas about overall themes and “deeper meanings” of literary texts. Teacher-led discussions were often rich, sometimes with high levels of student participation, but students’ written literary interpretations generally did not reflect that richness.

READI sought to address this issue through the design and use of sentence stems that acted as a framework in which students could articulate their ideas. For example, one sentence stem read, “This text suggests that the world is a place where \_\_\_\_\_.” The sentence stems are a form of scaffolding (Vygotsky, 1978; Wiggins & McTighe, 2005), and an attempt to help students become acculturated to discipline-specific discourse – the language that experienced readers might use to talk about literature (Ball & Freedman, 2004; Gee, 2008; Graff, 2003). Just as important, the sentence stems were designed to help students and teachers reconceptualize theme not as a single “message” or “deeper meaning,” but instead as a range of worldviews constructed through a reader’s interpretation of a text (Bruner, 1986). This kind of conceptualization may support students in understanding and interpreting texts.

One of the design team teachers in the READI project gave her class repeated practice using these sentence stems. On an essay test assessing growth in interpretive writing, her class showed significant gains in thematic inference (see Annual Report, Year 3). The READI project designed a companion study to isolate the potential influence of the sentence stems. If the sentence stems by themselves support interpretive writing, they could become a low-cost tool for teachers and students engaged in moving from discussion to interpretive writing.

## **Methods and Procedure**

The study took place in one of the READI project’s participating high schools. Students in two sophomore classes (N= 67) participated. In each classroom, students were assigned to one of two conditions:

1. A “universal language” condition (n = 32), in which the sentence stem was: “Reading this story helps us see that the world can be a place where \_\_\_\_\_.”
2. A control condition (n = 35), in which the sentence stem read, “Some of my interpretations of themes in this story are \_\_\_\_\_.”

The conditions were counterbalanced so that in each class, about 1/2 of the students were assigned to each condition.

All students were given a one-page excerpt from a novel called “Prisoner’s Dilemma” (Powers, 1996). The excerpt had been previously used as a stand-alone short story for the internationally normed International Baccalaureate English literature exam. Along with the novel excerpt, the students were given a worksheet that asked them to complete three written tasks.

First, they were asked to write a summary of the story. Second, students were given a definition of theme based on the definition they had been using in class during the course of the year:

**“Themes** are your interpretation of a story’s messages, big ideas, or underlying meanings. *Below, please discuss your interpretation of some of the themes in the story you just read. Please use the sentence starter below to begin your writing.*”

Students then made a claim about the story’s themes using the sentence stem. Finally, after completing their claim, students were asked to write a paragraph or more to “support or explain the interpretation you wrote above.”

## Key Findings

Two independent raters coded both the claims and supporting paragraphs as *interpretive* or *literal* based on codes adapted from empirical research on readers’ interpretive responses (Lehr, 1988; Svensson, 1987). Interrater agreement was 89% (Cohen’s  $\kappa = 0.73$ ). Claims coded as “literal” took the form of summary or paraphrase, even if couched in universal language. For example, a student in the “universal” condition wrote, “Reading this story helps us see that the world can be a place where kids are with their dads learning about the stars.” In claims coded as “interpretive,” a reader derived or constructed at least one abstract or conceptual inference that could be supported in some way by the features of the text. Such interpretations might include inferences about characters’ motivations or feelings, symbolic interpretations, or construction of connotations for concrete imagery. For example, another student in the “universal” condition wrote, “Reading this story helps us see that the world can be a place where we are trapped and alone.”

A similar coding system was used for students’ supporting body paragraphs. In this case, students’ paragraphs were coded as either literal, interpretive with supporting evidence from the text, or interpretive without supporting evidence.

Students in the “universal” condition made significantly more interpretive claims than those in the control group (chi-square with one degree of freedom = 5.945,  $p = .015$ ). This result is not too surprising, since the wording of the sentence stem should have prompted more universal language. However, beyond the gains in interpretive claims, the results show that students in the “universal” condition also made significant gains in the level of interpretation and text-based support in their supporting body paragraphs (chi-square with one degree of freedom = 4.992,  $p = .025$ ).

These results have immediate practical implications. Practically speaking, the results suggest that simple sentence stems may not only help students adopt academic discourse, but may help move them from summary to interpretive inference. Further, using universal language in a claim may influence other aspects of students’ writing; specifically, their focus on textual evidence that supports interpretive claims. These results, along with further research in this area, may also contribute to building a model of interpretive response to literature that can inform further interventions and teacher strategies.

## 4. Effects of sentence stems on writing about literature: Study 2

### Purpose

This study expands on previous READI work on instruction in literary interpretation to examine ways in which teachers can offer instructional scaffolds to help readers build interpretive skills and strategies. Specifically, this study examines the role of sentence stems – also called sentence frames or templates – in supporting students’ interpretive sense-making.

### Methods and Procedure

The study took place at one urban and two suburban public schools in the Midwest. The participating students were representative of their schools in terms of race, income, and reading skills. The urban school students – all 9<sup>th</sup> graders (n = 17) – were African American and Latino, with 44% scoring at “meets standards” on a state-wide standardized reading test, and 87% of the student body eligible for free or reduced lunch.

Eleventh graders (n = 110) and 8<sup>th</sup> graders (n = 58) from the two suburban schools were also representative of race, income, and reading scores. At both schools, about 5% of the students were African American, 40% Caucasian, 30% East and South Asian, and about 20% Latino. About 45% of students at both schools were eligible for free or reduced lunch. About 40% scored as “meets standards” on the same state-wide reading test referenced earlier, and 6% (School A) and 18% (School B) scored at “exceeds standards.”

The study was administered to all participating classes at the beginning of their spring quarter. Only the work of consented students was used in this study.

Students read a one-page story excerpted from the novel “Prisoner’s Dilemma” (Powers, 1996). In the excerpt, a father and his children look at the constellations with their father on a cold winter night. In the end, the father seems either to have died or have become so involved in his own thoughts that he does not respond to the children. He is “no longer warm.” The text complex and did not necessarily invite obvious clichés or “happiness bound” morals. All participating teachers judged it to be accessible to their students, and the text yielded a Flesch Kincaid reading ease score of 70 (indicating a 7<sup>th</sup> grade reading level). After reading the story, students were asked to summarize the story. Next, students read the following instructions: “Please discuss your interpretation of some of the themes in the story you just read.” Depending on the condition to which they were assigned, students received one of three possible sentence stems:

1. A **“Control” condition**: *Some of my interpretations of themes in this story are \_\_\_\_\_.*
2. An **“Emotion” condition**: Students were asked to choose only one of the sentence stems below to begin their writing.

*Choice A: By the end, reading this story leads us to feel that the author has a positive outlook on life, because the author seems to believe that \_\_\_\_\_.*

*OR*

*Choice B: By the end, reading this story leads us to feel that the author has a negative outlook on life, because the author seems to believe that \_\_\_\_\_.*

3. A “**Worldview**” condition: Reading this story helps us see that the world can be a place where \_\_\_\_\_.

In each class, the conditions were counterbalanced so that about 1/3 of the students in each room were assigned to each condition.

Next, students were asked to respond to the following question: “What do you think are the four words in the story that are most important to your interpretation?” Finally, students were asked to use a Likert scale to respond to the question, “How much do you think you understood the story?”

### **Analysis and Key Findings**

The students’ written work was typed into an excel spreadsheet. The author and an independent rater, both blind to condition, first examined each student’s summary to ensure basic understanding of plot. The responses of three students who clearly did not understand the plot were removed to insure accuracy of the analyses, leaving the study with a total of 185 responses. This study hypothesized that using sentence stems that framed theme in terms of “worldviews” or emotions about life would support students’ construction of thematic inferences. To explore this hypothesis, student theme statements were first analyzed using an *a priori* code that characterized responses as either “literal” or “interpretive” in nature. In a “literal” response, students summarized the plot or restated details from the story. An interpretive response could include connotation, symbolic interpretation, local inferences about character, or more global inferences about human nature. These definitions and codes were adapted from other empirical research on readers’ interpretive responses (Lehr, 1988; Svensson, 1987).

The author and two raters, blind to the aims of the study, first normed their analyses with 20 responses not used in the study. Then, the two raters independently coded the thesis statements of all students who participated in the study (Cohen’s kappa .78). All disagreements were resolved through discussion.

Analysis revealed that students who used Worldview or Emotion sentence stems were more likely to construct interpretive thesis statements in response to “Prisoner’s Dilemma” than students in the Control condition,  $\chi^2(2, N = 185) = 21.80, p < .001$ . The effect size for this finding was moderate (Cramer’s V, .34) (Cohen, 1988). Marascuilo post hoc multiple comparisons showed that frequencies in the Control condition were different from those in the Worldview condition,  $\chi^2(2, N = 185) = 15.16, p < .001$  and from frequencies in the Emotion condition,  $\chi^2(2, N = 185) = 18.62, p < .001$ . However, there were no differences in frequencies between Worldview and Emotion conditions,  $\chi^2(2, N = 185) = .13, p = .938$ .

*Differences related to grade levels.* As described earlier, students from three different grade levels – 8<sup>th</sup>, 9<sup>th</sup>, and 11<sup>th</sup> – participated in the study. The 8<sup>th</sup> and 11<sup>th</sup> graders came from both honors and non-honors classes. Analyses indicated that there were no relationships between interpretive response and grade  $\chi^2(2, N = 185) = 1.09, p = .580$ , or interpretive response and academic track ( $\chi^2(1, N = 185) = 2.53, p = .128$ ).

The small number of 9<sup>th</sup> graders ( $n = 17$ ) meant that one of the assumptions of chi square was violated, and the test might not be reliable. In order to address this possibility, responses were analyzed with 9<sup>th</sup> graders excluded from the analysis and then combined with the 8<sup>th</sup> graders. In both cases, chi-squares were non-significant.

*Trends in responses within interpretive theme statements.* This study also explored the extent to which students' interpretive theme statements were more likely to account for multiple and perhaps conflicting features of the text, and hence less likely to be clichéd or “happiness bound.” Using a process of open coding, two independent raters read all responses previously coded as interpretive and found two clear patterns in the types of interpretations students were making. These patterns may best be described as differences in interpretation of the overall valence of the story. For example, one group of students framed “Prisoner’s Dilemma” as an encouraging life lesson. Students in this group wrote that the story showed the reader to “treasure every moment with your family,” or another approximation of that message. Quite often, these thesis statements took the form of clichés or aphorisms, such as “Life is short” or “Take nothing for granted.” These statements were coded as “positive valence.”

Another set of students found the story to be a grim recognition that life is hard and cold, and that “We are alone in the world.” These statements were coded as “negative valence.” While some students acknowledged both positive and negative effects of the story, in almost every case they completed their theme statements with a definitive move toward either positive or negative. For instance, one student interpreted the story to be ultimately optimistic in effect, writing, “Some of the themes of this story are that life is hard, but as long as you have family, you’ll be okay.” These statements were coded according to their final “move” toward positive or negative evaluation.

A chi-square test for condition and valence showed significant differences:  $\chi^2(2, N = 134) = 6.94, p = .031$ , with post-hoc comparisons showing that the differences existed between Control and Emotion conditions,  $\chi^2(2, N = 134) = 7.56, p = .023$  only. Hence, there were more “negative valence” responses coded in the Emotion condition (50.7%) than in the Control condition (31.3%). No significant differences emerged between Control and Worldview, and Worldview and Emotion groups.

*Important words.* In order to examine the students' choices for “most important” words in the story, the author listed every individual word chosen by any student as most important, and then counted the number of times other students in each condition also chose that word. Three of the four most popular choices “constellation,” “leaving,” and “cold” were the same across condition. However, the proportions of students choosing those words differed, with 40%

of the Emotion group identifying “cold” as salient, while 26% of the Worldview and 19% of the Control conditions selected “cold” as important.

*Understanding rating.* There were significant differences in the understanding rating scale,  $F(2, 182) = 7.79, p < .001$ . Post hoc comparisons showed that students in the Control condition ( $M = 3.30, SD = 1.28$ ) reported significantly better understanding of the story compared to students in the Emotion condition ( $M = 2.22, SD = 1.71$ ). Students in the Worldview condition reported similar understanding to both groups.

Practically speaking, the results suggest that these interpretive sentence stems, which are simple, easy to memorize, and broadly applicable, may not only help students adopt academic discourse, but may help move them from summary to interpretive inference. Further, the results suggest asking inexperienced readers to make an evaluation of the overall positive or negative impact of a text, they may be more likely to consider more of the features of a text in their interpretation, and less likely to offer clichés or “happiness-bound” morals. Groups’ choices for most important words may reveal similar effects. These results support previous READI and related work showing that affective evaluation can be a powerful tool for close reading and identifying interpretively salient details.

Interestingly, the Control group, which produced the fewest interpretive statements, felt that they understood the poem relatively well. The Emotion group, which had the highest number of interpretive statements, and the highest number of negative evaluations of the story, had a significantly lower understanding rating. This result suggests that as students begin to construct more nuanced interpretations, they may feel less confident about their grasp of a text. The result further suggests that part of the teacher’s job is to guide students toward more nuanced, and messier, interpretations, and help them equate complexity with deeper understanding. Overall, these results, along with further research in this area, contribute to building a model of interpretive response to literature that can inform further interventions and teacher strategies.

## **5. Affective Evaluation, Salience, and Literary Interpretation: Helping High School Students Read Like Experts.**

This study was a follow-up to prior READI research conducted by Levine (2014; Levine & Horton, 2013). It examined whether strategic use of an interpretive heuristic called “affective evaluation” can guide novice readers’ attention to the kinds of textual details that expert readers find salient, or that are consistent with the “rules of notice” to which experts often attend. As discussed in earlier work, attention to salient details is an important element of experienced readers’ practice. If affective evaluation supports novice readers in identifying salient literary details in a way that draws on their personal knowledge and emotional response, or helps them develop relevant and meaningful thematic inferences, it can be a doubly useful tool for teachers and students.

To explore these questions, we asked five experienced high school literature teachers to think aloud while reading an excerpt from the novel *Prisoner’s Dilemma* (Powers, 1996). We compared these expert responses to the pre- and post-intervention think-aloud protocols of five

high school novice readers – all participants in a READI affective evaluation intervention study – as they read the same excerpt. we used the think-aloud responses of the five expert readers as an *in vivo* instantiation of the models, on the assumption that these experienced teachers of English would readily access and put into practice modes of expert reading that closely reflect more formal models of interpretation. Our general interest, then, is whether novice readers would display similar patterns of attention to the text and modes of interpretation after having been taught to engage in affective evaluation.

Specifically, we explored the following questions:

Q1. To what degree does novice readers' attention to specific textual details align with that of experts, at both pre- and post-intervention?

Q2. When responding to highly salient details, are novice readers more likely to attend to characterization, symbolism, or other common literary conventions?

Q3. To what extent do responses to salient details seem to contribute to readers' overall thematic inferences?

## Methods and Procedure

The affective evaluation intervention took place at a large urban public high school in the Midwest. The intervention lasted 4½ weeks, for a total of 18 instructional days. During this time, the Senior English class was introduced to and practiced affective evaluation as they read, discussed, and wrote about a range of popular and canonical literary texts, all of which were loosely connected to the topics of gender and “coming-of-age.”

The process of evaluation involves asking two meta-cognitive steps during and after reading: 1) While reading, ask yourself which words, phrases, or events seem particularly positive, negative, or both, and 2) Explain why you made that affective evaluation.

Students practiced affective evaluation with a series of texts, beginning with simple and familiar texts, such as a version of “Cinderella,” and progressed to more complex poems and short stories by authors like Sandra Cisneros (1991), Richard Wright (2008), and Junot Diaz (1997). Students would discuss their choices of affect-laden words, and would sometimes argue about whether a particular word or phrase was, in context, positive, negative, or both. When they had finished reading and annotating stories in this way, they were taught to make an appraisal of the text as a whole.

For the think-aloud interviews, both students and expert readers read a one-page excerpt from the novel *Prisoner's Dilemma* (Powers, 1996) with a Flesch-Kincaid reading ease score of 70. Each student in the sample was interviewed independently, both before and after the instructional intervention. The five English teachers who acted as expert readers were interviewed only once. All participants were told that they would be reading a story out loud, and as they read, they should use any strategies they knew to make sense of the text *and* to think about its themes, interpretations, or big ideas.

After reading the excerpt, participants were asked to summarize the story and then articulate a theme statement. Readers were then asked to identify what they felt to be the five

most important words or phrases in the entire text – words or phrases that had helped them build their thematic interpretations of the text. Then they were asked to explain the reasons for their selections.

## Key Findings

*Q1: To what degree does novice readers' attention to specific textual details align with that of experts, at both pre- and post-intervention?*

All think-alouds were transcribed. Then the story was divided into 61 separate “propositions”; typically, these propositions were independent clauses. To arrive at an operational definition of interpretive salience, we identified those propositions that the majority of experts found important or interesting enough to comment on, as well as those propositions that the majority of experts did *not* comment upon. As a reflection of expert consensus, we coded propositions that received comments from 4 or 5 experts as “High salience” propositions, propositions that received comments from none or only 1 expert as “Low salience.” We then examined the degree to which students followed similar patterns of responding to high salience and low salience material, both before and after the affective evaluation intervention. Specifically, we coded as a “hit” each instance in which a given student commented on a High salience proposition as well as each instance in which a student *refrained* from commenting upon a Low salience proposition.

Given the categorical nature of this measure, we analyzed these data using logistic mixed effects regression (Bates et al., 2008), with proposition salience (high; low) and think-aloud session (pre; post), and their interaction as fixed effects, and both students and propositions as random effects. The dependent measure was the odds of producing a “hit” (a comment for high-salience propositions or no comment for low-salience propositions). This model revealed a significant effect of proposition salience ( $b = -1.70$ ,  $SE = 0.76$ ,  $p < .03$ ), with a greater likelihood of hits for low-salience propositions ( $M = 0.85$ ;  $SD = 0.35$ ) than for high-salience propositions ( $M = 0.61$ ;  $SD = 0.49$ ). Also, though there was no main effect of session ( $b = 0.15$ ,  $SE = 0.38$ ,  $p = .69$ ), the proposition salience by session interaction was significant ( $b = 1.57$ ,  $SE = 0.82$ ,  $p < .05$ ). To follow up this interaction, we created a pair of additional models that examined each proposition type separately. These models revealed a significant effect of session for High salience propositions ( $b = 1.23$ ,  $SE = 0.59$ ,  $p < .04$ ), but no effect of session for Low salience propositions ( $b = -0.48$ ,  $SE = 0.37$ ,  $p = 0.20$ ).

This analysis showed that while students generally did not comment on Low salience propositions as part of both their pre- and post-intervention think-aloud protocols, their tendency to comment upon High salience propositions increased significantly after the intervention. For example, none of the students attended to the title of *Prisoner's Dilemma* at pre-intervention, but all five did at post-intervention.

*Q2: When experts or novices find a particular detail to be salient, what is the nature of their responses?*



We developed a protocol coding scheme through an open-ended process that involved repeated passes through the data to categorize responses followed by a system of constant comparison whereby we constructed common labels for responses that seemed closely related (Andringa, 1990). Altogether, this procedure resulted in six primary coding categories, including a “Character” category, which included references to affect, comments about character personalities, goals, or relationships, and a “Rules of Notice” category, which included explicit reference to repetition, juxtaposition, and other rules of notice. In many cases, the contents of a given response were coded as belonging to more than one category.

All responses (N = 310) were coded by the first author, and half of the responses were also coded by a second independent rater (Cohen’s kappa ranged from .82 to .95). As might be expected, the experts spontaneously displayed a range of interpretive responses. Interestingly, the highest proportion of their responses (64%) was coded as relevant to rules of notice. The experts also frequently discussed characters’ relationships or motivations (61% of responses), commented upon particular symbolic devices and meanings (49% of responses). Prior to the affective evaluation intervention, the novice readers were mostly (78% of responses) summarizing literal details of the story, and occasionally (30% of responses) making observations about the story characters.

Logistic mixed effect regression models show that post-intervention, the novice readers were significantly more likely to comment upon affective content, story characters, and symbolic meanings. These post-intervention responses were at levels similar to those of expert readers. Also, there was a significant relationship between affective responses and character inferences and symbolic interpretations, suggesting that the affective evaluation strategy helped students move to interpretive sense-making.

*Q3: To what extent do salient details contribute to readers’ overall thematic inferences?*

To explore the relationship between readers’ attention to details and their overall thematic inferences, we asked all readers to identify five details they thought were most important to the overall themes of *Prisoner’s Dilemma*, and then to explain their choices. Although the experts’ thematic responses were in no way formulaic, all five identified juxtapositions between *dark and light* and *cold and warm* as most salient to their interpretations. Four experts identified propositions connected to *distance* and *emptiness*, and four identified language related to teaching and learning as salient to interpretation. All experts explained their choices by connecting the above motifs to emotions of love and loneliness that the children felt in the presence of their father. Further, the experts’ thematic inferences reflect a synthesis of those ideas about high-salience details, as well as an integration of the language of some of their choices of salient details.

The thematic statements of pre-intervention novices did not integrate their interpretations, perhaps simply because novice readers made only a few interpretations in their pre-intervention protocols. Interestingly, the pre-intervention students still drew on their choices of salient details when they constructed themes; however, those themes were related only superficially to those details.

Post-intervention, students added a set of details that were not present pre-intervention, and more importantly, students were able to explain their choices of salient details in interpretive terms. In particular, four of the five novices joined the expert readers in identifying words or phrases connected to darkness and cold as most salient to their interpretations. The students' thematic statements also were more clearly connected to their interpretive ideas.

One of the fundamental challenges of high school English instruction is helping students engage with literary texts in ways that lead to personally meaningful interpretations, and moving students toward independence in interpretive sense-making. The present work builds upon recent demonstrations of the effectiveness of leveraging students' everyday practices of affect-based evaluations (*Is this good? Is this bad? Why?*), and shows that when guided by affective evaluation, novice readers' attention to a text can mirror that of more expert readers in important ways. This kind of interpretive heuristic appears to be a useful route by which novice readers move toward expertise in their approaches to and engagement with literary texts.

Appendix A.

Descriptions of Types of Coding Applied to Think – Aloud Data of Adolescents and Undergraduates

Process Codes for Think - Aloud Statements

Category	Sub-Category	Definition	Example
Elaboration	Self Explanation	Meaningful integration of text content with prior knowledge or generation of a connection between text information that was not explicitly stated in the text.	“Brother Timothy is thinking out loud now-trying to understand why you miss mass, and I guess in this sort of congregation it's not really acceptable to miss mass.”
	Surface Text Connection	Relates the focal sentence to another sentence on the basis of surface features (usually makes a compound or complex sentence out of two separate sentences).	Sentence 1: It was the boy’s silence that was the enraging thing. Sentence 2: And what was even worse was how happy the boy seemed. Comment: “He was enraged that the boy was silent and happy.”
	Irrelevant Association	Use of prior knowledge that does not add to the meaning of the text.	“This reminds me of the butterfly cookie I brought for lunch today.”
	Prediction	Comments about information that the reader expects to follow.	“Maybe he’s going to talk next time someone comes in because he’s ready to go outside.”
Paraphrase		Repeating the gist of a sentence in the text without adding additional information.	Text Sentence: Once could see at a glance that he was angry. Comment: “He was mad.”
Evaluation		A positive or negative judgment about some aspect of the text.	“That’s a lot of information” “The author’s not adding anything new.”
Comprehension Problem		Comments that indicate a lack of understanding (can be either	"I don't understand" "What does mass mean?"

Comprehension Success	explicit or a question indicating a lack of comprehension). Comments that indicate an understanding of text	“That makes sense.”
-----------------------	--	---------------------

---

### Rules of Notice Coding for Think-Aloud Statements

Category	Definition	Example
Noticing Discrepancy	Noticing that information within the text contradicts or is not consistent with either with prior portions of the text or with prior knowledge.	“Wait, before it said he was angry, now he’s laughing- you can’t be angry and happy at the same time.”
Noticing Repetition	Mentions that information found within the text has been repeated.	“They keep bringing up the wooden door for some reason.”
Explicit Mention of Literary Device	Statements that explicitly mention a literary device (e.g., personification, alliteration, etc.)	“They’re using personification here because a caterpillar can’t really talk.”
Connection to Title	Any type of connection between a sentence or portion of the text and the title of the story.	“That’s why it’s called ‘The Butterfly’ because caterpillars grow into butterflies.”
Symbolism Interpretation	Attributing symbolic meaning to a part of the text.	Well he hasn't talked very much in the story, but what if his silence is a symbol of his innocence that he tried to do nothing wrong?
Other Interpretive Comment	Attributing meaning to a part of the text that extends beyond the situation of the text	The story has been based around this butterfly and how spirituality can keep you as a caterpillar, but going out and exploring the world can make you be more of a person expressing yourself and become a butterfly which is evolving and changing.

---