

Using affective appraisal to help readers construct literary interpretations

Sarah Levine and William S. Horton
Northwestern University

Students can readily engage in summary and literal sense-making when reading poems, short stories, and other literary texts, but are often unable to construct inferences and thematic interpretations of these works. This paper discusses the results of an instructional intervention built on an affect-based model of literary interpretation. Students in the intervention group spent four weeks reading and writing about popular and canonical texts, with a focus on poetry. As they read, they identified valence-laden language, made appraisals of valence, and then explained or justified their appraisals. Analyses of pre- and post-test results show that the intervention group made significant gains in the level of interpretive responses to poems compared to a control group of students who were not explicitly taught to engage in affective appraisal. This work sheds light on ways in which affect-based interpretive strategies can support novice readers' interpretive practices.

Keywords: literary interpretation, affect, novice readers, intervention, high school students, theme

Introduction

Many high school students find literature to be a puzzle that is divorced from personal experience. Oftentimes, they see literary interpretation as an “an occult process” (Harker, 1994, p. 202) with “secret meanings” (Graff, 2003, p. 52). They may readily summarize the literal events in a short story, identify the topic of a poem, or describe the characters in a novel, but they often have much more difficulty constructing connotation, theme, and other kinds of figurative inferences (Earthman, 1992; Graves & Frederiksen, 1991; Janssen, Braaksma, Rijlaarsdam, & van den Bergh, 2012). In the words of one student asked to interpret a short story: “What does the fish mean? What does the merry-go-round mean? It doesn't mean

anything to me” (Marshall, 1987, p. 41). Despite literature’s deep connection to feeling, literary interpretation is a spiritless chore for many students, and even the most emotionally compelling text may leave a novice reader feeling nothing at all.

If literary interpretation were simply a task that needed to be completed to receive a good grade in school, it probably would not matter very much. But the stakes are higher than that. Scholes (1986) reminds us that our worlds are made up of texts that persuade, criticize, and enlighten. Modeling the processes of literary interpretation is a fundamental way that teachers can help their students open themselves to — or defend themselves from — the world of texts around them.

In this paper, we argue that *feeling* can play a critical role in helping novice literary readers construct connotation and figurative interpretation. This argument is based on the assumption that students — even those who are novices at literary interpretation — have ready access to affect-based interpretive practices. That is, even though students may not immediately know what a text “means,” they can often evaluate the affective components of valence, mood, and tone. We propose that calling attention to these components can help students identify important authorial choices and, in doing so, help them shift from purely literal to thematic readings of texts. In an intervention study conducted with high school English literature students, we explore how an *affective appraisal* strategy — i.e., identifying the valence of words and phrases, and the moods and tones of whole texts, and then supporting those interpretations — can be an accessible first step for students in constructing thematic interpretations of poetry.

To understand how this strategy is intended to work, consider this example from our instructional intervention. In one of the first lessons, the teacher asked her 12th grade students to reflect upon Edward Hopper’s painting “Nighthawks,” which depicts a city diner at night. Specifically, the teacher asked the students to:

1. Appraise the valence of the painting, considering whether its overall impact was positive, negative, or both.
2. Look for details in the painting that give rise to these affective appraisals.
3. Explain why each detail seemed negative, positive, or both.

In response, many students appraised the valence of the painting to be fairly negative. To justify their responses, they pointed to details such as a dark window and the lack of people on the street. When asked to explain why those details seemed negative, one student said that while a window could ordinarily be a sign of openness and friendliness, the dark window in the painting seemed negative because it suggested loneliness. Other students said that the absence of people on the street was negative because it showed emptiness and isolation. As this example shows, students used affective appraisal to identify concrete details in the painting and to construct abstract thematic inferences about meaning of those details.

We propose that equipping students with an affective lens can help them tease out the meanings of a variety of texts, and that this approach may be especially useful for novice readers.

Background

Novice readers of literature

In his examination of U.S. students' reading achievement, Purves (1984) presents a model of the movement from literal to interpretive sense-making. He focuses on a line from a narrative that is rich with figurative language: "*Astride a gleaming motorcycle, [a man] roared into a dawn filled with the hum and smog of Caracas traffic*" (p. 87). Purves argues that to construct a figurative interpretation of such a line, a student must identify words they believe are significant to the overall effects of the text, draw on their knowledge of the norms associated with those words, and then generate relevant abstract connotations. In this particular case, Purves says, a student might identify words like "astride" and "roared," and then derive or construct abstractions like "opulence" and "power" (p. 88) as part of their overall interpretation.

However, as Purves points out, this series of cognitive moves is not straightforward for many novice readers. While studies show that children and inexperienced readers have some facility for interpreting figurative language (Bowdle & Gentner, 2005; Gibbs, 1994; Katz & Ferretti, 2001; Winner, Rosenstiel, & Gardner, 1976), and while metaphor is arguably embedded in Western linguistic and conceptual practices (Lakoff & Johnson, 2003; Lakoff & Turner, 1989), when it comes to literary texts, novice readers often struggle to make figurative interpretations. Studies using think-aloud paradigms and expert-novice comparisons have found that novice literary readers (here defined as novices in literary interpretation, not simple decoding) are frequently less able to identify possible symbols or patterns of imagery in texts, and are instead more likely to focus on plot and surface features of a text (Earthman, 1992; Peskin, 1998). In a study of high school readers, nine students who had been identified by their teachers as "below average in reading literature" showed "a monotonous pattern of retelling" when reading poetry, in contrast to ten stronger readers who responded to poems with more associations and personal responses (Janssen, Braaksma, Rijlaarsdam, & van den Bergh, 2005). Novice readers are also less cognizant of the relationship between language and theme (Dorfman, 1996); less aware of the purposeful structure of a text (Graves & Frederiksen, 1991); and less aware of the contribution of language and style to literary quality (Dixon, Bortolussi, Twilley, & Leung, 1993). Harker (1994) found

that even students whose teachers identified them as strong interpreters still tended to summarize action or describe concrete imagery in two poems.

When students do move toward thematic inferences, it is generally in a highly constrained setting; for example, kindergarteners were able to choose a pair of thematically matched children's books when given several possible options (Lehr, 1988). However, novice readers generally have difficulty independently *constructing* themes (Johnson & Goldman, 1987; Narvaez, Bentley, Samuels, & Gleason, 1998). Since one of the primary goals of teaching is to help students develop independence in critical thinking, the ability to construct themes (as opposed to picking from a provided menu) is especially important.

How is literary interpretation taught?

Evidence suggests that high school teachers may have difficulty providing models or instructional supports for interpretation. Some teachers frame interpretation as a practice that readers should absorb through simple exposure (Holt-Reynolds, 2000). Others present interpretation as the product of a one-to-one mapping between literal and symbolic meanings, leading students toward one preconceived interpretation (Applebee, 1993; Hamel, 2003; Nystrand & Gamoran, 1991). Even teachers who advocate more flexible approaches to interpretation may still rely on lecture or narrow "yes/no" questions, rather than emphasizing strategies that "could make students independent of their teachers' questions" (Applebee, 1993, p. 140).

Textbooks and teaching guides may also fail to make interpretive strategies explicit (Hillocks & Ludlow, 1984; Wilhelm & Smith, 2001). A teacher's edition of a leading literature textbook instructs the teacher to have students track the color red in a short story. The teacher is instructed to ask, "What qualities does red seem to symbolize?" and then is told to expect the following student responses: "Danger, love, courage...death" (Beers & Odell, 2005, p. 347). Such instructions presuppose that readers already have strategies for moving from the concrete "red" to the abstract "danger" or "courage." The textbook does not explain *how* to move from literal to interpretative meaning.

Some instructional models, built on both cognitive and literary theory, have been more successful in teaching interpretation. These models draw on the premise of the "cognitive apprenticeship" (Collins, Brown, & Holum, 1991), which maintains that the tacit processes involved in many cognitive tasks need to be made visible to students. The models also draw on literary theorists' ideas about literary conventions, such as repeated images or changes in point of view, which expert readers use as guides to interpretation (Booth, 1975; Culler, 2002; Rabinowitz,

1998; Scholes, 1989). We can think of these instructional models as “interpretive apprenticeships,” because they assume that students need the curtain pulled back on the tacit interpretive processes that underlie literary sense-making.

Several studies of instructional interventions may act as examples of these “interpretive apprenticeships.” In one intervention, built on Booth’s (1983) discussion of literary cues for the presence of irony, students constructed explicit criteria for interpreting irony in poetry. Students in the experimental group made significant gains in their ability to identify ironic language (Smith, 1989). Similarly, an instructional intervention in the teaching of poetry (Peskin, Allen, & Wells-Jopling, 2010) presented symbolic archetypes common to Western poetry, based on Frye’s model (1964). This study showed similar gains in interpretation as compared to a control group that did not engage with this model.

Lee’s Cultural Modeling Framework (1995a, 1995b, 2007) builds on the notion of the interpretive apprenticeship by drawing on students’ everyday knowledge and interpretive practices, making those practices visible, and using them as scaffolds for literary interpretation. One example is the practice of signifying among speakers of African American English, where interpreting figuration is central: “Your mama’s so skinny she can do the hula hoop in a cheerio” (Lee, 2001, p. 122). In discussion and writing, students make public their intuitive reasoning about interpreting everyday figurative language. That reasoning then becomes a guide to formal interpretive moves with canonical literature. Lee’s instructional interventions using this framework showed significant pre-post assessment gains in students’ levels of interpretation.

The models described above suggest that two elements of successful teaching of interpretation might be: (1) asking students to examine their current and probably implicit interpretive practices, and (2) providing explicit cognitive models for literary interpretation. In the present work, we make the case that affective appraisal might be a connected and equally important element of this interpretive process, both because it is an everyday interpretive practice and because it can be explicitly modeled with regard to literary texts.

The role of affective appraisal in literary processing

Affective appraisal is fundamental to interpreting everyday experience across a variety of contexts. Empirical and theoretical work supports the idea that we perceive our world through a paradigm framed by benefit on one side and harm on the other (Bargh, Chaiken, Raymond, & Hymes, 1996; Johnson-Laird & Oatley, 1992; Robinson, 2005; Zajonc, 1984). By the time we are teenagers, most of us have had a wealth of experience making affective appraisals of everything from facial

expressions to films. We appraise the positive valence in a compliment or the negative valence of a tragic story. We appraise the “good” or “bad” of an experience or event, and then reason about why we have done so.

Our experience with affective appraisal in the world may translate well to our experience with literary texts. Booth (1983) asserts that literature is built on feeling, and empirical research suggests that we respond more emotionally to literature than non-literary texts (Mar & Oatley, 2008; Palencik, 2008; Zwaan, 1999). It follows that affective appraisal may be a useful starting point for literary sense-making, where students can build on their everyday experience — this time appraising valence, mood, and tone in a text, and then supporting, explaining, and warranting those appraisals. For example, Williams et al. (2002) showed that asking elementary school students to evaluate the outcome of a fable as negative or positive helped them construct basic thematic inferences. Students in an intervention group who were asked, “Was what happened good or bad? Why was it good or bad?” (p. 237) made significant gains in their construction of thematic statements for thematically similar fables compared to a control group, who were simply asked, “What happened?” However, as was true in other studies, the intervention group did not successfully transfer their skills of thematic inference to thematically dissimilar fables — clearly a problem for teachers who wish to help their students become independent interpreters of literature.

Obviously, individual readers might make different appraisals of a text’s valence depending a variety of factors. For example, a particular reader’s current mood might affect his or her judgments of a word’s valence (Bower, 1981, 1992). Likewise, positive feelings for a character might make a reader more resistant to processing that character’s failure (Gerrig & Rapp, 2004). Readers also have affective responses independent of the mood or tone of a passage; for example, a reader might have a negative response when struggling to comprehend a complicated text (Kuiken, Miall, & Sikora, 2004). In the present work, however, we focus on the affect that readers *ascribe* to a text — that is, their evaluations of the valence, moods, and tones of a text (Brendl & Higgins, 1996).

In particular, we wish to consider what Dixon et al. (1993) call “common effects” — in other words, interpretations that are likely to be made by many readers in a given population. Evidence suggests that there are “common effects” in texts when it comes to valence, mood, and tone. Readers agree on positive or negative valence of stories (1993) and make common inferences about emotional states of characters, even when those emotional states change over the course of a story (Miall, 1988). Readers may even use very similar language to describe the moods of a story (Sadoski, Goetz, & Kangiser, 1988). Furthermore, readers are able to identify a text’s common effects even if they themselves do not partake of those effects. For example, in one study, a think-aloud participant read a fictional passage

about a tea party, and remarked sarcastically, “How exciting” (Smith, 1991, p. 266). The reader did not feel positive about tea parties, but she was able to assess the tone of the text as positive. For our purposes, then, readers’ affective appraisals are their evaluations of a text’s valence based on the features of that text. Our instructional intervention is designed to help readers identify these affective features as a route toward successful interpretation.

The study

To explore the hypothesis that affective appraisal might support literary interpretation for novice readers, we carried out a study with high school students that included pre- and post-tests of literary interpretation, a four-week instruction period, and think-aloud interviews. Both intervention and control groups completed a pre-test asking for thematic interpretation of a poem. Then, the intervention group engaged in a series of lessons in which they practiced the affective appraisal strategy, while a control group engaged in more conventional instruction in literary interpretation, without affective appraisal. Finally, both groups took a post-test that again asked for thematic interpretation of poetry. Representative students from both intervention and control groups also participated in clinical think-aloud interviews; here, however, we focus on the results of the poem interpretation tasks that came before and after the instructional period.

Participants

The study took place at a large, urban public high school in the Midwest. At this school, 77% of students are Latino, 13% are African American, 5% are Asian, and 4% are Caucasian (including many Polish immigrants). Approximately 86% are from low-income households. At the time of the study, 33% of the students met state reading standards, and 1% exceeded standards (“Illinois interactive report card,” 2012). English literature classes are required for all four years of high school, and meet every day for 45 minutes.

Two teachers volunteered to participate in this study in response to a general call to the school’s English department, and each was given a small stipend for participation. Both had master’s degrees and had taught for 4 to 8 years; both planned month-long units on literary interpretation with similar start and end points, and both wished to use a combination of popular and canonical texts in their units. In addition, both teachers’ instructional goals included helping students understand how literary devices and language influenced thematic meaning.

The control group consisted of consented students¹ from three untracked 12th grade classes (N = 42), all taught by one of the volunteer teachers (the “control teacher”). The intervention group consisted of students from one untracked 12th grade class (N = 17), as well as a class of 10th graders in an honors English class (N = 20). Honors classes included students who had earned high grades in the previous year, students recommended by teachers regardless of grade, and students who requested registration in honors classes because of interest or attention to college readiness. Untracked classes included students whose schedules could not accommodate honors classes, students who had high grades previously but decided against honors classes for one reason or another, and students who earned lower grades in previous classes. Both intervention classes were taught by the other volunteer teacher (the “intervention teacher”) and co-taught by the first author during the study period. The intervention teacher had been using exactly the same curricula with both the untracked and honors classes since the start of the school year, so it was decided to include both sets of students in the intervention group.

Clearly, honors and untracked students could be different from one another in ways that might affect their capacity to apply an affective appraisal strategy. We will return to this possibility in the Results section, where we examine potential differences across sub-groups.

Materials

All students completed a written poem interpretation task both before and after the multi-week instructional unit. The pre-test asked students to respond to *one* of two poems: Randall Jarrell’s “The Death of the Ball Turret Gunner,” about a soldier’s dehumanizing death in war; or Juan Delgado’s “The Lame Boy Returns,” about the narrator’s regret for childhood cruelties to a disabled boy (see Appendix for both poems). In the post-test, students were asked to respond to *both* poems. The poets’ names were removed from the assessments. Neither the intervention nor the control group read or discussed these poems before or during the study period.

Although it can be difficult to “match” literary texts, we attempted to do so in choosing test poems. These poems were selected because they were relatively short (52 and 101 words, respectively) and judged by the English teachers involved in the study to be comparably accessible to students. The poems measured similarly on the Flesch-Kincaid reading ease scale (“The Death of the Ball Turret Gunner” measured at the 9th grade level, with a reading ease score of 66; “The Lame Boy Returns” measured at the 8th grade level, with reading ease score of 57). A few words in each poem were defined in side notes. Additionally, when we compared

the student responses to each poem at pre-test, we found no significant differences in levels of interpretation, suggesting that the students found the poems to be similarly accessible.

The poems were followed by two written prompts: First, the students were given a prompt adapted from the College Board Advanced Placement English Literature exam (“Collegeboard.com,” 2009), which read, “Using what you know about understanding poetry, please discuss the meaning of the poem as a whole. Please write at least two sentences.” Second, the students were asked, “How well do you feel you understand the above poem?” For this self-report measure, students responded on a 6-point Likert scale, where “1” represented “a little” and “6” represented “a lot.”

Procedure

Pre- and post-test assessments

One day prior to the start of both groups’ instructional units, each teacher gave her students one of the two poems and the written prompt. The pre-test poems were counterbalanced so that half of each group received the first poem, and the other half received the second poem. The students were asked to read the poem to themselves and respond to the prompt. The teachers advised the students that they could annotate the poems if they wished. The students were given 22 minutes to complete the pre-test.

The day after the instructional unit was completed, each teacher instructed all of her students to read both poems (one after the other) and respond to the same prompt for each. The students were given 22 minutes to respond to each text. This design allowed comparison of pre- and post-test responses to the same poem, along with generalization to a new poem. Because the study used one pre-test and two post-tests, we obtained twice as many post-test responses as pre-test responses across students.

Instructional design, intervention group

The intervention unit was structured to allow students to practice an affective appraisal strategy over time with a sequence of literary texts organized by levels of difficulty. Introductory lessons drew on Lee’s Cultural Modeling Framework by interrogating everyday texts in multiple media, where indicators of high affective valence provided a lens for extrapolating connotation and theme. Almost every lesson involved individual or small group work, as well as whole class discussion focused on students making their thinking visible. Each lesson stretched across several days of 45-minute class periods; for example, a lesson involving

interpretation of song lyrics might take two days. The unit included 17 class periods in all. Below is a brief description of each lesson in the intervention, in sequence:

Lesson 1 – Interpreting word choice (3 periods): The teacher introduced the unit using what Lee (2007) calls “cultural data sets” — that is, texts that students might encounter in their everyday, non-academic lives, and which might help make explicit the interpretive practices that students already use. The teacher led a discussion of comparative word choices in several texts, including drafts of a Barack Obama campaign poster and a list of both the given and adopted names of celebrities. For example, students compared the valence of the name of actor Jamie Foxx with that of his given name, Eric Bishop. Most students ascribed a positive valence to both names, but explained that the name “Foxx” seemed positive because it suggested cleverness and sexuality, whereas “Bishop” seemed positive because it suggested purity. The teacher then explicitly presented students with the steps of the affective appraisal strategy, explaining that students may, without knowing it, already use these steps when interpreting the effects of language:

1. Appraise the valence of the text, considering whether the overall impact is positive, negative, or both.
2. Look for details in the text that led to the affective appraisals.
3. Explain why each detail seemed negative, positive, or both.

Lesson 2 – Interpreting overall effects of artwork (3 periods): Students analyzed paintings by Edward Hopper, Jacob Lawrence, and Frida Kahlo. Students used the affective appraisal strategy to make interpretive readings of the paintings, as discussed earlier.

Lesson 3 – Using new vocabulary and sentence stems (1 period): The teacher gave students lists of positive and negative descriptors to offer them an expanded vocabulary for discussing interpretations. The teacher also gave students a set of sentence stems designed to frontload the affective appraisal strategy. Sentence stems included “This image creates a mood of...” and “This text condemns a world in which...” Students used the vocabulary and sentence stems when they practiced affective appraisal on new texts in subsequent lessons. These stems provided additional scaffolding for use of the affective appraisal strategy as well as examples of academic language to convey interpretive arguments.

Lesson 4 – Interpreting messages of ads, writing argument (5 periods): The teacher presented students with two car advertisements with contrasting tones. The students practiced the affective appraisal strategy with those ads and then wrote one-paragraph arguments about the effects of one of the ads.

Lesson 5 – Interpreting songs (2 periods): The teacher moved to a comparative analysis of several love songs with different affective impacts. Students practiced the affective appraisal strategy with these more complex texts.

Lesson 6 – Interpreting poetry (3 periods): The students chose and made a detailed interpretive reading of one of several poems, using the affective appraisal strategy to appraise the valence of words and phrases in context as well as the cumulative moods and tones of the poems.

Instructional design, control group

During the same four-week period, which included 17 class periods in all, the control group teacher implemented her planned unit on literary interpretation. The teacher stated that her goal for the unit was to teach students to identify and construct connotation and thematic interpretations of literary texts. Like the intervention unit, the control unit included popular and canonical texts of several genres, sequenced in order of difficulty. Instruction included guiding students' attention to individual details and patterns in texts, and to literary devices. The control group teacher also asked open-ended questions about students' responses to literature, including questions about students' emotional responses to what they read. However, the teacher did not use the affective appraisal heuristic during instruction. As with the intervention group, the control students worked on their own, in small groups, and in whole class discussion throughout the unit. The first author observed and recorded five class periods over the course of the four-week study. The control group teacher also provided detailed daily lesson plans for her unit of instruction. Below is a brief description of each lesson in the control group's unit:

Lesson 1 – Interpreting allusions (2 periods): The teacher presented current magazine advertisements that used allusions to popular television shows to make their points. Students identified the allusions and constructed associations for them. For example, a credit card company used a picture of the character Sue from the television show "Glee" to advertise the card's connection to a national charity. The ad suggested that its audience would end up like Sue if they didn't use the credit card. The students noted that Sue was portrayed as mean and manipulative, and then developed their interpretation of the message of the ad.

Lesson 2 – Interpreting songs (3 periods): Students read and listened to several hip-hop songs by Drake and Li'l Wayne. Students identified allusions and concrete images in the songs. They identified the sources of the allusions, along with associations suggested by the allusions. Then students discussed effects of the allusions on song meaning.

Lesson 3 – Interpreting songs, writing argument (4 periods): The teacher modeled a structure for argumentative writing that included a main idea, evidence, and a statement that linked the evidence to the main idea. She asked students to make and support claims about which of the two hip-hop songs had the best lyrics. The students first defined “best lyrics” to mean lyrics that were poetic, concise, and helped convey a message, and then looked for lyrics that fit that definition. Then students drafted arguments, reviewed other students’ writing, and revised their own work based on criteria modeled in class.

Lesson 4 – Interpreting poetry (3 periods): Students studied poems by Maya Angelou (“Still I Rise”), Marge Piercy (“Barbie Doll”), and Kim Addonizio (“What Do Women Want?”) that addressed issues of oppression and gender. The teacher asked questions about the effects of the poem, including affect-based questions like “How does this line make you feel?” However, the teacher did not use the affective appraisal strategy; that is, students were not asked to ascribe valence, tone, or mood to the text, but instead were asked to evaluate their own affective responses as they read. Students constructed arguments about the effects of literary devices like metaphor and imagery, including statements about the poems’ themes.

Lesson 5 – Imitating style and syntax (1 period): The students wrote their own lines of descriptive prose, in which they mimicked the style and syntax of famous authors. The goal of this lesson was to help students understand the relationship between literary technique and effect. The teacher read these sentences to the class and commented on their effects.

Lesson 6 – Interpreting short stories (4 periods): Students read the prose poem “Girl” by Jamaica Kincaid, which addresses issues of gender inequality. The class discussed expectations for women and men in contemporary society, and wrote about their own understandings of gender roles. The teacher asked students to find symbols, sensory imagery, and allusions in the text, and then create associations for those literary devices. She then asked the students to think about what the author was trying to say. Several students wrote about the way the story expressed the unfair expectations society had for women. Then, students read “2 B R N 0 2 B” by Kurt Vonnegut, which criticizes, among other things, the devaluing of human life. They followed the same process of discussing the messages of the story, identifying symbol, imagery, and allusions in the texts and creating associations for them, and then writing about what they felt the author was trying to say.

Coding the pre- and post-tests

The open-ended pre- and post-test responses were coded for levels of thematic inference in the students' interpretations of the poems. See Table 1 for a detailed explanation of these levels and representative examples of student responses for "The Death of the Ball-Turret Gunner." The coding system was adapted from Svensson's (1987) work on the development of symbolic interpretation in poetry and Lehr's (1988) thematic scale, both of which attempt to articulate levels of inference and abstraction in participants' responses to literature. These scales describe a continuum of literary sense-making, anchored on one side by local, literal descriptions of action in the text and on the other by global inferences about theme. Here, we borrow from Kurtz and Schober (2001) in defining theme as "a reader's notion of [a] main idea, message, or central meaning of the text...best expressed as a generalized declarative statement or proposition. Theme grows out of the particulars of a text but moves beyond specifics to comment on generalities about culture and humanity" (p. 140).

For the current analysis, we adopted six specific interpretive codes, organized into three categories:

1. In a *literal descriptive* response, a student summarizes part or all of the text. Such simple responses can be *unsupported by the features of the text*, (e.g., "In *Romeo and Juliet*, Romeo lived") or *supported* (e.g., "In *Romeo and Juliet*, Romeo died").
2. In a *literal interpretive* response, a student moves beyond local summary or description, and provides local evaluation or inference about action or character, or mood or tone. Such responses can be *unsupported by the features of the text* ("Romeo is always thoughtful and deliberative") or *supported* ("Romeo is consistently desperate and impetuous"). An unsupported interpretive reading would alter the overall effects of the text in significant ways.
3. In a *thematic* response, a student moves beyond local evaluation of an aspect of the text, and achieves global interpretation of the text as a whole. Such responses can be *unsupported by the features of the text* ("*Romeo and Juliet* celebrates the delightful way that people always live in harmony") or *supported* ("*Romeo and Juliet* condemns those who would destroy true love"). To arrive at an understanding of some of the themes for each poem, three independent expert readers, all with advanced degrees in English literature, were asked to contribute interpretations of the themes of both poems. Any student response aligned with the experts' themes, the raters' themes, or any other "equally valid alternative interpretations" supported by the features of the text (Kurtz & Schober, 2001; Williams, 1993) was rated as *thematic supported*.

Table 1. Coding categories and sample student responses to “The Death of the Ball Turret Gunner.”

Category	Definition and guidelines for scoring	Sample response	Explanation of example	Point value
Literal descriptive unsupported	Summary or paraphrase only; no interpretation. Misrepresents the action or characters in a way that significantly alters the text’s overall literal meaning.	“A man is at war. He takes a shower to wash out his fright.”	No interpretation beyond text’s literal action, and the summary is unsupported: Mistaking death for a shower significantly alters the text’s meaning.	1
Literal descriptive supported	Summary or paraphrase only. May be incomplete, or include small inaccuracies that do not alter the text’s meaning.	“A soldier was using a war plane, the enemies shot him, he died.”	Summary. Even though it is incomplete (not mentioning being washed out of the turret, for example), it still qualifies as literal and supported.	2
Literal interpretive unsupported	Local interpretation not supported by the text so that meaning or effects are significantly altered. May include summary.	“The man in the poem is proud and happy to be at war.”	Makes local interpretation about poem (inferring pride and happiness). However, interpretation is unsupported by the text.	3
Literal interpretive supported	Local interpretation that is supported by the text. May include summary.	“He could be so easily replaced that he was of no importance.”	Includes interpretation of the last line, inferring worthlessness of soldier. Interpretation remains local.	4
Thematic unsupported	Universal statement not supported by the text, or so vague that it could apply to many texts with opposing themes. May also include summary.	“In this text, the author says we need to be careful when making decisions.”	Includes universal statement about human nature or society. Is “unsupported” because the poem stresses <i>lack</i> of control over decisions, not ability to control decisions. Interpretation is also vague.	5
Thematic supported	Global or universal statement about the main ideas, messages, or central meanings in text. May also include summary.	“Soldiers are not valued as people, just as fighters to wash away.”	Interpretation includes universal statement about human nature or society. It is supported by the features of the text.	6

As in similar research (Lehr, 1988; Williams et al., 2002), we assigned a point value to each of the above categories, intended to capture the degree to which students' responses contained evidence of thematic interpretation. We used a 6-point scale, ranging from "literal descriptive unsupported" at 1 point, to "thematic supported" at 6 points. Assigning point values in this way required judgments about the value of textually supported responses on the one hand, and depth of interpretation on the other. What is more "valuable" in terms of literary response: an accurate literal summary of a poem, or a figurative interpretation that is not supported by the features of the text? Since we are interested in the degree to which students move from literal to interpretive sense-making, we chose to assign a higher point value to an unsupported interpretive response than to a supported literal response. We readily acknowledge, though, that these values are somewhat arbitrary; they are mainly intended to capture the intuition that students' responses can vary from purely literal to richly interpretative. Table 1 shows the coding scheme and sample responses for the poem "The Death of the Ball Turret Gunner." An equivalent set of examples was developed for the other poem, "The Lame Boy Returns."

Identifying student annotations

We examined student annotations on the pre- and post-test poems to determine if the intervention group had made use of the affective appraisal strategy as they read, and to look for other possible trends in annotations in either group. If students made any notes at all on their test papers that related to the task at hand (as opposed to graffiti, for example), we marked the papers as "annotated." We further looked for evidence of the affective appraisal strategy (for example, words like "positive" or "negative," positive/negative signs (+ -), or arrows (↑↓)). This shorthand for affective appraisal was modeled during the intervention, and students were encouraged to practice it as they read.

In both groups, we looked at other types of notes; for example, summary, questions about lines or phrases, inferences about characters, or other personal responses to the text.

Results

Using the coding scheme described above, the first author and an independent rater coded 234 student responses — one response from each of the 79 students on the pre-test, and two responses from each student on the post-test (minus three students who only wrote one post-test response). Interrater agreement was 82.5%

(weighted Cohen's $\kappa = 0.84$). Cases of disagreement were resolved by discussion. The raters also examined the annotations of the poems to identify attention to affective appraisal, as well as other possible patterns in student notes.

First, it seemed necessary to examine the extent to which intervention students might actually have used the affective appraisal strategy as they read. One way to do this was by examining student annotations of test poems. In the control group, although the teacher did not instruct her students to do so, about 45% of students annotated their poems at pre-test, and about 60% annotated at post-test. The types of annotations varied, but most were notes about literal meaning, inferences about character (e.g. "The soldier is confused") or students' own emotional responses (e.g. "So sad"). Summary was the dominant kind of annotation in the control group's pre- and post-test papers.

The intervention teacher also did not instruct her students to annotate their poems. In the intervention group pre-test, about 50% of students did so. As with the control group, the majority of pre-test annotations involved summary of phrases or lines. At post-test, 100% of students annotated their poems. The post-test papers show that almost all students (90%) used the affective appraisal heuristic in their annotations, often marking arrows ($\uparrow\downarrow$) or positive/negative signs (+ -) next to lines and phrases in the poems to indicate appraisal of valence, and then adding comments to explain those judgments. These annotations support our claim that the affective appraisal was an important part of the intervention students' interpretive sense-making. Additionally, as we will show, the actual post-test responses also indicate that intervention students used the affective appraisal strategy.

The critical question posed by this study is: To what extent is the affective appraisal strategy associated with gains in levels of interpretive or thematic inference? Table 2 presents the frequencies and proportions of responses in each group that were coded as literal, interpretive, or thematic (both supported and unsupported) at pre-test and post-test. From this table, we see that at pre-test, the majority of responses of both groups (59% for the control group and 75% for the intervention group) were coded as "literal descriptive," meaning that they focused on summary and surface features of the text. At post-test, a substantial proportion of the responses from control group students (45%) were still coded as "literal descriptive," while 33% of control group post-tests were coded as "thematic." In contrast, only 1% of the intervention group's post-test responses were coded as "literal descriptive," while a majority of the intervention post-test responses (66%) were coded as thematic. This pattern of results indicates that the affective appraisal intervention was successful in helping students in the intervention group construct thematic interpretations of texts.

Table 2. Frequencies and proportions of student responses from the poem interpretation task that were assigned to each coding category, by group condition and test point.

		Literal descriptive		Literal interpretive		Thematic		Total ^a
		Unsup-ported (1)	Sup-ported (2)	Unsup-ported (3)	Sup-ported (4)	Unsup-ported (5)	Sup-ported (6)	
Control group (n = 42)	Pre-test	6	19	1	11	1	4	42
		0.14	0.45	0.02	0.26	0.02	0.10	1.00
	Post-test	6	31	2	16	9	18	82
		0.07	0.38	0.02	0.20	0.11	0.22	1.00
Intervention group (n = 37)	Pre-test	6	22	2	5	1	1	37
		0.16	0.59	0.05	0.14	0.03	0.03	1.00
	Post-test	0	1	3	21	8	40	73
		0.00	0.01	0.04	0.29	0.11	0.55	1.00

Note. The numbers in parentheses are the numeric values associated with the response code categories.

^a Each student responded to one poem at the pre-test and two poems at the post-test. There were three missing post-test responses, two in the control group and one in the intervention group.

To provide support for this conclusion, we examined the average response code values for the intervention and control groups at pre- and post-test, as shown in Figure 1². To confirm that students in the intervention group were providing more interpretive responses at the post-test, we submitted students' response scores at post-test to an ANCOVA with condition (control; intervention) as the independent variable and pre-test score as the covariate. Controlling for pre-test performance, this analysis revealed a significant main effect of condition ($F(1,76) = 43.05$, $MSE = 1.36$, $p < .001$, $\eta_p^2 = .36$), with higher response scores in the intervention group than control group. Although follow-up comparisons revealed that both the control group ($t(41) = 2.99$, $p < .01$, $d = 0.43$) and intervention group ($t(36) = 11.46$, $p < .001$, $d = 1.78$) showed significant gains from pre-test to post-test, the size of the pre-test/post-test gain was much larger (2.8 points on the 6-point scale) in the intervention group than the control group (0.61 points on the 6-point scale).

This pattern strongly indicates that the affective appraisal intervention facilitated the intervention students' ability to generate thematic interpretations of poetry. Recall, however, that at the pre-test each student was asked to interpret one of two poems, either "The Death of the Ball Turret Gunner" or "The Lame Boy Returns," while at the post-test, each student was asked to respond to both poems. It is possible that previous exposure to one of the poems allowed students to generate more successful interpretations at post-test for the poem that they had seen before. To examine this possibility, we separated the mean response ratings

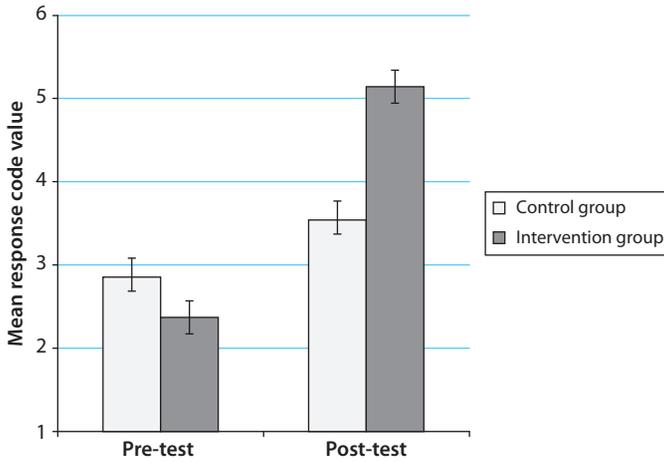


Figure 1. Mean student response code values, by group condition and test point. Error bars represent the standard errors of the means across subjects.

at post-test into two categories: responses for the poem students had seen before (familiar poem), and responses for the poem that students had not seen (new poem). The average post-test response ratings for each poem are shown in Figure 2, along with the average pre-test ratings shown previously. Although the groups neither saw nor discussed either poem between pre and post-test, the control group showed significant gains from pre- to post-test on the familiar poem ($t(41) = 4.11, p < .001, d = 0.64$). However, the control group did not make significant gains on the new poem ($t(39) = 1.11, p = .28, d = 0.17$). In contrast, the intervention group made significant gains on both the familiar poem ($t(36) = 10.53, p < .001, d = 1.74$) and the new poem ($t(35) = 11.38, p < .001, d = 1.85$).

Another question that is important to ask is whether the observed gains in interpretive levels occurred on a student-by-student basis, or whether these gains only occurred in the aggregate. That is, did individual students improve? To assess this, we examined the frequency with which individual students' response scores rose, declined, or stayed the same across the test points. When interpreting the familiar poem, 20 of 42 students in the control group made gains in their responses, and in the intervention group, 34 of 37 students made gains. When interpreting a new poem, only 16 of 42 control group students showed gains, while 35 of 37 intervention students showed gains. These patterns indicate that the improvement across test points seen in the overall analysis also occurred on a student-by-student basis — especially for those students in the intervention group.

Recall, though, that the intervention group was comprised of both 10th grade honors students and 12th grade “untracked” students, while the control group was

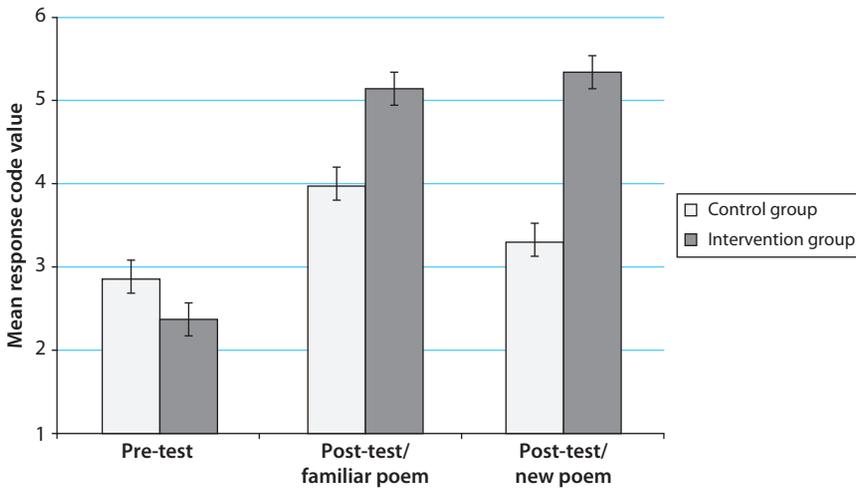


Figure 2. Mean student response code values by group condition and test point. Post-test scores have been split according to whether the interpretation task involved the same poem as the pre-test, or a new poem. Error bars represent standard errors of the means across subjects.

comprised solely of 12th grade untracked students. With these differences in mind, we conducted two additional follow-up analyses to explore how this might have impacted our results. First, within the intervention group, we compared the honors students against the untracked students and found that mean response scores on the poem interpretation task were generally higher in the 12th grade untracked class than the 10th grade honors class, at both pre-test (12th grade untracked: $M = 2.82$; 10th grade honors: $M = 1.95$; $t(35) = 2.53$, $p < .03$, $d = 0.86$) and at post-test (12th grade untracked: $M = 5.44$; 10th grade honors: $M = 4.85$; $t(35) = 2.08$, $p < .05$, $d = 0.70$). Thus, it appears that the 12th grade students in the intervention group were more successful at generating interpretive responses, even at pre-test, and this advantage carried over into the post-test (although both sub-groups showed significant gains in interpretation from pre-test to post-test). If anything, then, the inclusion of the 10th grade honors class reduced the degree to which the intervention group as a whole showed evidence of thematic interpretation. Next, we compared the subset of 12th grade untracked students from the intervention group against the control group, which was entirely comprised of 12th grade untracked students. Here, we found that the mean performance across these two groups was nearly identical at pre-test (12th grade intervention: $M = 2.82$; control: $M = 2.86$; $t(57) = .079$, $p = .94$, $d = 0.02$) but diverged at post-test (12th grade intervention: $M = 5.44$; control: $M = 3.54$; $t(57) = 5.08$, $p < .001$, $d = 1.35$). So, when we limit our analysis to include only students who were most comparable in terms of age and/

or experience, we obtain the same pattern as in the overall analysis, with more robust thematic interpretation at post-test by students in the intervention group.

Next, we were interested to know whether students made accurate assessments of their own understanding of the texts. Thus, we examined students' responses to the question "How well do you feel you understand the above poem?" One student failed to respond to this question at pre-test, and nine students failed to respond to either one or both questions at post-test. Average self-reported poem comprehension ratings are shown in Table 3. In general, students reported moderate understanding of the poems, with no overall difference in self-reported understanding between the control and intervention groups ($t(77) = 1.23, p = .22, d = 0.28$). In Table 3, we can see that control students were more confident at the pre-test in their assessments of understanding than the intervention group, but this difference went away at the post-test. To compare the ratings provided by students in the intervention vs. control groups at post-test, we conducted an ANCOVA on the post-test ratings, controlling for students' ratings at pre-test. This analysis revealed a marginal effect of condition ($F(1,72) = 3.03, MSE = .714, p = .08, \eta_p^2 = .04$), suggesting that, when pre-test differences in these self-report ratings are taken into account, students in the intervention group were relatively more confident in their understanding at post-test. Follow-up comparisons showed that the level of poem understanding reported by the intervention group increased from pre-test to post-test ($t(33) = 5.12, p < .01, d = 0.74$), while the level of understanding reported by the control group did not change ($t(40) = 1.43, p = .16, d = 0.21$).³

Table 3. Mean ratings in response to the question "How well do you feel you understand the above poem?" by group condition, test point, and (for the post-test only) whether the poem was familiar or new. Ratings were provided on a 6-point scale that ranged from 1 = "a little" to 6 = "a lot." Standard deviations are in parentheses.

	Pre-test	Post-test		
		Familiar poem	New poem	Total
Control Group	3.56 (1.12)	4.00 (1.10)	3.60 (1.14)	3.80 (1.13)
Intervention Group	2.93 (1.06)	3.81 (1.18)	3.83 (1.47)	3.82 (1.32)

Finally, we computed correlations between each group's self-reported understanding ratings and their poem interpretation response scores, for the pre-test and post-test separately. For the control group, there was no evidence of any correlation between their understanding ratings and actual interpretation scores at either pre-test ($r(41) = -.07, p = .64$) or post-test ($r(81) = .09, p = .43$). For the intervention group, there was no correlation between their understanding ratings and actual scores at pre-test ($r(37) = -.09, p = .58$), but these two measures were significantly

and positively correlated at post-test ($r(64) = 0.28, p < .03$). Higher interpretation scores at post-test in the intervention group were associated with higher ratings of self-reported understanding. This suggests that students in the intervention group had more accurate intuitions of their interpretive success after the affective appraisal intervention.

Discussion

In examining the responses to the poem interpretation task, we found that at pre-test, both the control and intervention groups tended to make literal responses to the test poems. This indicates that, at the beginning of the study period, these novice readers were primarily providing descriptions of the poem content. At post-test, the control group made significant gains in interpretive levels, but only when they responded to the familiar test poem, not the new test poem. In other words, if a control student read “The Lame Boy Returns” at pre-test, he or she was likely to show gains in response to “The Lame Boy Returns” at post-test, but *not* to “The Death of the Ball Turret Gunner.” This was true regardless of the poem read first. On the other hand, the intervention students made significant gains in interpretive and thematic response from pre-test to post-test regardless of poem, suggesting more independence in interpretive skills. Moreover, these interpretive gains were accompanied by an increase in the degree to which students in the intervention group indicated they understood the poem at post-test. These students appeared to have a better-calibrated sense of their own interpretative processes.

Representative responses

To gain insight into these general findings, and to study the possible role played by the affective appraisal strategy in this interpretation task, we now consider in more detail representative responses from the control and intervention groups at pre- and post-test.

Control responses

In the control group, it was frequently the case that students produced responses scored as “literal descriptive” at both pre- and post-tests. For example, one student’s pre-test response to “The Death of the Ball Turret Gunner” read, “The person came into the world from his or her mom. He laid in her belly until he was born. He died in a war and woke to see that he was died. The fighters that killed him he called nightmare fighters.”⁴ The response arguably misreads a few details of the

literal action of the poem, but is overall an accurate summary, and was therefore coded as “literal descriptive.” The same student’s post-test response to the familiar poem read, “He became a turret gunner in his home state. He got into a fighter plane and flew in the air. He got killed in a explosion and woke up to blackness. They washed his body out the turret.” This post-test response also summarizes the poem without adding inference or thematic interpretation, and was also coded as literal descriptive. The second post-test, in response to the new poem, was similarly focused on surface features of the text.

Interestingly, when control group students did move to thematic inference, their responses often took the form of morals or lessons, as in “The Ball Turret Gunner shows that you should treat other people with respect,” or adopted commonplace sayings that were generally supported by the features of the text, as in, “The meaning of this poem is that life can end anywhere, anytime.” Other responses coded as “thematic” resembled literal responses, but included universal language or a single inferential statement that seemed to synthesize or interpret the effects of the poem; for example, “Many people die when they’re in war and are forgotten.”

A few of the control group’s thematic responses created much richer interpretations; such as this control student response: “From reading this poem, I can see that the world is a cold place. To the world we are only useful as long as we are strong and alive and full of energy and health.” However, more of the control post-test thematic responses took the form of clichés or “universalized” summary.

Intervention responses

The most frequent responses in the intervention group were coded as literal at pre-test, but thematic at post-test. This shift occurred regardless of poem. For example, one student from the intervention group read “The Death of the Ball Turret Gunner” at pre-test and wrote, “Its about a soldier that is in a battlefield and is being shoot at and witness bombs blowing up as well as a flashback of how he was in his mothers womb and woke up to a bomb that killed him.” This response was coded as literal, since it focuses entirely on the action and surface features of the poem. After the intervention, the student wrote this response: “This author condemns a world in witch there will always be war, even in a dream or a flashback. You will always be a soldier even if its a horrifying experience.” This response was coded as thematic because it identifies global themes that were aligned with those put forward by the expert readers in our study. The student’s second post-test, written in response to the poem “The Lame Boy Returns,” was also coded as thematic.

This student’s response is representative of the intervention group at post-test in that it captures the predominant features of the poem and makes inferences about mood, tone, and general worldviews of the text. In addition, and important

to our exploration of the affective appraisal heuristic, the student's response is also representative in that it uses affect-laden language, as well as justifications of affective appraisal, to make its case. For example, the student above ascribed a negative valence to the text as a whole by using a sentence stem introduced during the intervention ("The text condemns a world in which..."). The student includes an affective evaluation of the soldier's experience ("horrible") and also justifies the negative appraisal by emphasizing the inescapable nature of war, pointing out that in this poem, there will "*always* be war" and "you will *always* be a soldier" (italics ours).

Generally speaking, post-test thematic responses by the intervention group seemed to capture and synthesize more salient details of the poems, while control responses tended toward broader statements. For instance, a control student's thematic response read, "The meaning of this poem is nobody wins when it comes to war." In contrast, an intervention response read, "This text condemns a world where fighters die and seem to be spit on afterwards." The intervention group's more precise focus, as well as integration of salient features of the poems, suggests that the affective appraisal strategy may help students build richer representations of texts.

Affective appraisal as sense-maker

How might an affective lens facilitate the intervention students' conceptualizations and interpretations of these poems? Research suggests that an affective lens may facilitate the processing and connecting of details with similar valence, thereby supporting students' ability to form a "cohort" of details and accompanying inferences (Bargh et al., 1996; de Vega, 1996; van den Broek, Rapp, & Kendeou, 2005). The affective appraisal strategy therefore might help a student connect negative images of war and death with negative images of a soldier being "washed out... with a hose." Further, research suggests that literary devices such as symbol or imagery – often used in literature to help construct or highlight important concepts – may be more valence-laden or simply more apt to elicit emotional responses than other kinds of language (Goetz, Sadoski, Stowe, Fetsco, & Kemp, 1993; Miall & Kuiken, 1994). Reading with the affective appraisal strategy, then, may have guided students to attend to the kind of literary language that is particularly salient to a text's meanings and messages.

If we return to the intervention student who moved from a literal to thematic reading of "Death of the Ball Turret Gunner," we may see affective appraisal at work as a literary sense-maker. The student's pre-test response read, "Its about a soldier that is in a battlefield and is being shoot at and witness bombs blowing up as well as a flashback of how he was in his mothers womb and woke up to a bomb that killed him." This response shows awareness of the poem's imagery of war, but simply

indexes those images, referencing “soldier,” “battlefield,” “shoot,” and “bombs.” At post-test, the student goes beyond listing the images by articulating a connection between them that seems to be affect-based. The student wrote, “This author condemns a world in which there will always be war, even in a dream or a flashback. You will always be a soldier even if it’s a horrifying experience.” It may be that the student’s appraisal of the negative valence of the poem and subsequent search for supporting details guided him or her to move from a listing of war imagery to a synthesis of that imagery (“there will always be war”) and an interpretation of the worldview of the poem.

Literary schema

A noteworthy difference between the two groups studied here is that the intervention group made gains on both post-test poems, while the control group only made gains on the poem that was familiar to them. Peskin (1998) surmises that novice readers may not make interpretive moves because they lack the “rich stock of schemata” that expert readers call upon when identifying effective literary devices or constructing thematic inferences. For example, an experienced reader generally expects that literary authors make purposeful choices for particular effects (Graves & Frederiksen, 1991; Vipond & Hunt, 1984). Novice readers are less able to identify possible symbols (Earthman, 1992), less cognizant of the simplicity or complexity of language and its relationship to theme (Dorfman, 1996), and less aware of the contribution of language and style to literary quality (Dixon et al., 1993). We hypothesize that our intervention may have helped students construct an additional schema for understanding literary texts. The affective appraisal strategy described here may function not just as a set of questions for students to ask and answer, but also as a general approach to literary reasoning. The focus on appraising textual valence (“Is it positive? Is it negative? Where? Why?”) may remind students that texts are designed to evoke feeling, and that one of the roles of the reader is to construct an interpretation of that feeling based on the features of the text. In contrast, the control unit of study did not explicitly address ways that texts operate, or basic strategies that readers can use to interact with texts, and therefore may not have added to students’ literary schemas.

Further considerations

There are several aspects of the current study that deserve mention as possible limitations and/or avenues for future research. First, as mentioned above, the intervention group was given an explicit reading strategy to practice as they read.

Additionally, this particular strategy drew on students' everyday, affect-based interpretive practices, and was therefore perhaps more accessible than other strategies. We would argue that these qualities are part of the strategy's unique strengths. On the other hand, while the control group was exposed to literary reading strategies and even instructed to look at details of a text to see what the text is "trying to say," the control teacher never explicitly outlined the processes underlying these strategies, nor did she connect strategies to everyday reasoning. In this way, control students may have been "prepared to learn interpretations, not make them" (Thompson & Wolff, 1994, p. 9). It may be, then, that the difference between these two groups arises simply from the use of any explicit strategy, and not particularly the affective appraisal strategy. A next step in this study is to compare the use of the affect-based strategy with another strategy for which students are given explicit instruction.

Another element that may have played a role in more successful thematic interpretations is the intervention classroom's use of sentence stems (e.g., "This text condemns a world in which..."). The sentence stems were designed to accompany and support the general use of affective appraisal by prompting students to identify what is positive or negative in a text, and justify that appraisal in abstract or universal terms. However, it is possible that the sentence stems alone — or other types of sentence stems — might have led to increased facility in connotation or thematic inference. Therefore, a next step in this research is to explore this possibility.

We can also ask whether a "positive/negative" frame is the best or only frame for assessing literary texts. For example, dimensions of the semantic differential scale (e.g. an "active/passive" comparison) could be used instead (Heise, 1970; Osgood, 1957, 1969). It seems likely that evaluation of texts on other scales would also lead readers to move from the literal to the abstract, for very much the same reasons that the affective appraisal strategy does: by providing a filter through which to organize concepts, establish coherence, and construct abstractions. However, a strong case can be made for focusing on negative and positive poles. Osgood, who created and tested the semantic differential scale to measure the quality and meaning of language, asserts that most of the scales along which one might measure meaning (such as a scale anchored by *weak* and *strong*, or *active* and *passive*) are ultimately grounded in affective measures (1969). Miall (1988) found the same return to affect when he asked students to evaluate poems on other scales. Additionally, other scales or lenses would necessarily alter a reader's interpretive processes and thematic inferences. For example, a Marxist reading might evaluate a text using a construct of power and powerlessness, or a feminist reading might evaluate a text in terms of its relation to gender inequality. Such readings would be fascinating, but perhaps beyond the kinds of initial interpretive moves we discuss in this study.

Another limit of this study concerns its application to longer texts. It is one thing to assess the valence of a painting and yet another to appraise the shifting and contradictory tones of a novel or play. There is some evidence that students used the strategy flexibly, allowing for multiple and even concurrent appraisals of different moods and tones. For example, students ascribed both positive and negative valence to the Edward Hopper painting. However, the strategy has not yet been tested on short stories or novels, where thousands of details or chunks of text might signal strong valence, and where it might become much more difficult to organize and synthesize interpretations. It seems likely that the affective appraisal strategy works better with shorter texts such as poems, which often act like mood pieces or emotional “snapshots.” On the other hand, it is also possible that an appraisal of the overall valence, tones, or moods of a longer text might help students develop a richer representation of that text, by guiding them to valence-laden — and possibly more significant — passages. Further research with other genres is required to explore the limits of the strategy.

Yet another possibility to consider is that some of the gains from pre- to post-test in this study result from students’ improved understanding of the intent of the test prompt. That is, perhaps at pre-test, more students misunderstood the assignment to be asking only for a literal summary. If this were the case, however, we should see equal gains in both groups, as both focused on literary interpretation and theme statements during the intervening four-week unit, and neither teacher used the prompt for written exercises during the course of their unit. Additionally, other studies of poetic interpretation have used similar prompts (e.g., “What do you think this poem is about?” (Eva-Wood, 2004; Peskin & Wells-Jopling, 2012; Svensson, 1987). Still, to isolate this variable, further studies should include a question about summary as well as theme so that students might clearly distinguish the two.

Finally, a clear limit of this study is that the intervention classes were co-taught by the intervention teacher and the first author, while all control classes were taught by the control teacher. For the intervention group, the presence of an outside teacher may have led to higher student excitement or motivation. However, the first author visited the control class and videotaped it five times during the study period. She introduced herself to the students, explained the nature of the study, and conversed with small groups and individual students as they engaged in class work. During these visits, the control teacher told her students several times that she wanted them to “do their best for the college researcher” and “make themselves look good” for the video camera. Our hope is that such actions may have led to greater student interest or motivation in the control group. However, there is still the distinct possibility that the two groups felt different levels of engagement during the study period.

Conclusion

In general, the findings from this study add to the existing body of research showing that “interpretive apprenticeships,” in which interpretive processes are made visible, can help students move from literal to thematic readings of literary texts. Additionally, this study suggests that the lens of affective appraisal, which readers naturally use in their everyday sense-making, is an interpretive process that can be leveraged and made explicit for novice readers. By learning to attend to and justify their appraisals of valence, tone, and mood in literature, students can take one more step towards independence when constructing interpretations of the world of texts that surrounds them.

Notes

1. Although all students in all participating classes received the same instruction, only data from consented students will be examined in this study.
2. We also compared the pre-test and post-test responses separately for each of the poems, “Death of the Ball Turret Gunner” and “The Lame Boy Returns,” in order to examine whether students found one poem easier to interpret. We found that there were no significant differences in responses to the different poems within or across groups. Thus, our analyses collapse across poems.
3. As can be seen in Table 3, the patterns of self-reported understanding for the familiar and new poems at post-test mirrored the patterns seen in the interpretation ratings. That is, compared to the pre-test, control group students reported an increase in understanding for the familiar poem ($t(39) = 2.98, p < .01, d = 0.49$) but not the new poem ($t(39) = 0.16, p = .88, d = 0.02$), while intervention group students reported increased understanding for both poems (pre-test to familiar: $t(31) = 4.02, p < .001, d = 0.70$; pre-test to new: $t(31) = 3.71, p < .01, d = 0.67$).
4. All student responses are presented without corrections to spelling or grammar.

Authors' note

This work was supported by a Northwestern University Cognitive Science Advanced Research Fellowship awarded to Sarah Levine. The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305F100007 to University of Illinois at Chicago (Goldman, 2009). The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education. We wish to thank Carol Lee, Andrew Ortony, David Rapp, and Adam Schwartz for their comments and suggestions during the development and writing of this project. We also wish to thank Michael Barber, Timothy Bagwell, and Kevin Bond for acting as expert readers of poetry, and Travis White-Schwoch for serving as the second rater for the students' poem interpretations. Finally, we extend our greatest appreciation to the teachers and students who participated in this study.

References

- Applebee, A. N. (1993). *Literature in the secondary school: Studies of curriculum and instruction in the United States*. Urbana Illinois: National Council of Teachers of English.
- Bargh, J. A., Chaiken, S., Raymond, P., & Hymes, C. (1996). The automatic evaluation effect: Unconditional automatic attitude activation with a pronunciation task. *Journal of Experimental Social Psychology*, 32(1), 104–128. doi:10.1006/jesp.1996.0005
- Beers, & Odell. (2005). *Elements of literature: third course – Illinois teacher's edition*. Austin Texas: Holt, Rinehart and Winston.
- Booth, W. C. (1975). *A rhetoric of irony*. University Of Chicago Press.
- Booth, W. C. (1983). *The rhetoric of fiction*. University of Chicago Press.
- Bowdle, B. F., & Gentner, D. (2005). The career of metaphor. *Psychological Review*, 112(1), 193–216. doi:10.1037/0033-295X.112.1.193
- Bower, G. (1981). Mood and memory. *American Psychologist*, 36(2), 129–148.
- Bower, G. (1992). How might emotions affect learning? In *The handbook of emotion and memory*. Hillsdale, New Jersey: L. Erlbaum Associates.
- Brendl, C., & Higgins, E. (1996). Principles of judging valence: What makes events positive or negative? In *Advances in Experimental Social Psychology* (Vol. 28, pp. 95–160). Elsevier. Collegeboard.com. (2009, October 21). Retrieved from <http://collegeboard.com>
- Collins, A., Brown, J. S., & Holum, A. (1991). Cognitive apprenticeship: Making thinking visible.
- Culler, J. D. (2002). *Structuralist poetics: Structuralism, linguistics and the study of literature*. Psychology Press.
- De Vega, M. (1996). The representation of changing emotions in reading comprehension. *Cognition & Emotion*, 10(3), 303. doi:10.1080/026999396380268
- Dixon, P., Bortolussi, M., Twilley, L., & Leung, A. (1993). Literary processing and interpretation: Towards empirical foundations. *Poetics*, 22(1–2), 5–33.
- Dorfman, M. H. (1996). Evaluating the interpretive community: Evidence from expert and novice readers. *Poetics*, 23(6), 453–470. doi:10.1016/0304-422X(96)00004-6
- Earthman, E. (1992). Creating the virtual work: Readers' processes in understanding literary texts. *Research in the Teaching of English*, 26(4), 351–384.
- Eva-Wood, A. (2004). Thinking and feeling poetry: Exploring meanings aloud. *Journal of Educational Psychology*, 96(1), 182–191.
- Frye, N. (1964). *The educated imagination*. Indiana University Press.
- Gerrig, R., & Rapp, D. N. (2004). Psychological processes underlying literary impact. *Poetics Today*, 25(2), 265–281.
- Gibbs, R. W. (1994). *The poetics of mind: Figurative thought, language, and understanding*. Cambridge University Press.
- Goetz, E. T., Sadoski, M., Stowe, M. L., Fetsco, T. G., & Kemp, S. G. (1993). Imagery and emotional response in reading literary text: Quantitative and qualitative analyses. *Poetics*, 22(1–2), 35–49. doi:10.1016/0304-422X(93)90019-D
- Graff, G. (2003). *Clueless in academe: How schooling obscures the life of the mind*. New Haven: Yale University Press.
- Graves, B., & Frederiksen, C. (1991). Literary expertise in the description of a fictional narrative. *Poetics*, 20, 1–26.
- Hamel, F. L. (2003). Teacher understanding of student understanding: Revising the gap between teacher conceptions and students' ways with literature. *Research in the Teaching of English*, 38(1), 49–84.

- Harker, W. (1994). "Plain sense" and "poetic significance": Tenth-grade readers reading two poems. *Poetics*, 22(3), 199–218.
- Heise, D. (1970). In G. Summers (Ed.), *Attitude measurement*. Rand McNally.
- Hillocks, G., & Ludlow, L. H. (1984). A taxonomy of skills in reading and interpreting fiction. *American Educational Research Journal*, 21(1), 7–24.
- Holt-Reynolds, D. (2000). What does the teacher do? Constructivist pedagogies and prospective teachers' beliefs about the role of a teacher. *Teaching and Teacher Education*, 16(1), 21–32.
- Illinois interactive report card. (2012, February 28). Northern Illinois University. Retrieved from http://iirc.niu.edu/School.aspx?source=Test_Results&source2=PSAE&schoolID=150162990250617
- Janssen, T., Braaksma, M., Rijlaarsdam, G., & Van den Bergh, H. (2005). Flexibility in reading literary texts: Differences between weak and strong adolescent readers. In *EARLI conference*. Presented at the EARLI, Cyprus.
- Janssen, T., Braaksma, M., Rijlaarsdam, G., & Van den Bergh, H. (2012). Flexibility in reading literature: Differences between good and poor adolescent readers. *Scientific Study of Literature*, 2(1), 83–107. doi:10.1075/ssol.2.1.05jan
- Johnson, D., & Goldman, S. (1987). Children's recognition and use of rules of moral conduct in stories. *The American Journal of Psychology*, 100(2), 205–224. doi:10.2307/1422404
- Johnson-Laird, P. N., & Oatley, K. (1992). Basic emotions, rationality, and folk theory. *Cognition & Emotion*, 6(3), 201–223.
- Katz, A. N., & Ferretti, T. R. (2001). Moment-by-moment reading of proverbs in literal and nonliteral contexts. *Metaphor and Symbol*, 16(3–4), 193–221. doi:10.1080/10926488.2001.9678895
- Kuiken, D., Miall, D. S., & Sikora, S. (2004). Forms of self-implication in literary reading. *Poetics Today*, 25(2), 171–203.
- Kurtz, V., & Schober, M. F. (2001). Readers' varying interpretations of theme in short fiction. *Poetics*, 29(3), 139–166. doi:10.1016/S0304-422X(01)00040-7
- Lakoff, G., & Johnson, M. (2003). *Metaphors we live by* (2nd ed.). University of Chicago Press.
- Lakoff, G., & Turner, M. (1989). *More than cool reason: A field guide to poetic metaphor* (First Edition.). The University of Chicago Press.
- Lee, C. D. (1995a). Signifying as a scaffold for literary interpretation. *Journal of Black Psychology*, 21(4), 357–381.
- Lee, C. D. (1995b). A culturally based cognitive apprenticeship: Teaching African American high school students skills in literary interpretation. *Reading Research Quarterly*, 30(4), 608–630.
- Lee, C. D. (2001). Is October Brown Chinese? A cultural modeling activity system for underachieving students. *American Educational Research Journal*, 38(1), 97–141. doi:10.3102/00028312038001097
- Lee, C. D. (2007). *Culture, literacy, & learning: Taking bloom in the midst of the whirlwind*. New York: Teachers College Press.
- Lehr, S. (1988). The child's developing sense of theme as a response to literature. *Reading Research Quarterly*, 23(3), 337–357.
- Mar, R. A., & Oatley. (2008). The function of fiction is the abstraction and simulation of social experience. *Perspectives on Psychological Science*, 3(3), 173–192.
- Marshall, J. (1987). The effects of writing on students' understanding of literary texts. *Research in the Teaching of English*, 21(1), 30–63.
- Miall, D. (1988). Affect and narrative: A model of response to stories. *Poetics*, 17(3), 259–272. doi:10.1016/0304-422X(88)90034-4

- Miall, D., & Kuiken, D. (1994). Foregrounding, defamiliarization, and affect: Response to literary stories. *Poetics*, 22, 389–407.
- Narvaez, D., Bentley, J., Samuels, J., & Gleason, T. (1998). Moral theme comprehension in third graders, fifth graders, and college students. *Reading Psychology*, 19(2), 217–41.
- Nystrand, M., & Gamoran, A. (1991). Instructional discourse, student engagement, and literature achievement. *Research in the Teaching of English*, 25(3), 261–290.
- Osgood, C. (1957). *The measurement of meaning*. Urbana: University of Illinois Press.
- Osgood, C. (1969). On the whys and wherefores of E, P, and A. *Journal of Personality and Social Psychology*, 12(3), 194–199.
- Palencik. (2008). Emotion and the force of fiction. *Philosophy and Literature*, 32(2), 258–277.
- Peskin, J. (1998). Constructing meaning when reading poetry: An expert-novice study. *Cognition and Instruction*, 16.
- Peskin, J., Allen, G., & Wells-Jopling, R. (2010). “The educated imagination”: Applying instructional research to the teaching of symbolic interpretation of poetry. *Journal of Adolescent & Adult Literacy*, 53(6), 498–507. doi:10.1598/JAAL.53.6.6
- Peskin, J., & Wells-Jopling, R. (2012). Fostering symbolic interpretation during adolescence. *Journal of Applied Developmental Psychology*, 33(1), 13–23. doi:10.1016/j.appdev.2011.08.002
- Purves, A. C. (1984). The potential and real achievement of U.S. students in school reading. *American Journal of Education*, 93(1), 82–106.
- Rabinowitz, P. (1998). *Before reading: Narrative conventions and the politics of interpretation*. Columbus: Ohio State University Press.
- Robinson, J. (2005). *Deeper than reason: Emotion and its role in literature, music, and art*. Oxford, England: Oxford University Press.
- Sadoski, M., Goetz, E. T., & Kangiser, S. (1988). Imagination in story response: Relationships between imagery, affect, and structural importance. *Reading Research Quarterly*, 23(3), 320–336. doi:10.2307/748045
- Scholes, R. (1986). *Textual power: Literary theory and the teaching of English*. Yale University Press.
- Scholes, R. (1989). *Protocols of reading*. New Haven; London: Yale University Press.
- Smith, M. W. (1989). Teaching the interpretation of irony in poetry. *Research in the Teaching of English*, 23(3), 254–272.
- Smith, M. W. (1991). Constructing meaning from text: An analysis of ninth-grade reader responses. *The Journal of Educational Research*, 84(5), 263–271.
- Svensson, C. (1987). The construction of poetic meaning: A developmental study of symbolic and non-symbolic strategies in the interpretation of contemporary poetry. *Poetics*, 16(6), 471–503. doi:10.1016/0304-422X(87)90014-3
- Thompson, G., & Wolff, J. (1994). All literature is argument. Retrieved from <http://www.eric.ed.gov/ERICWebPortal/contentdelivery/servlet/ERICServlet?accno=ED375396>
- Van den Broek, P., Rapp, D. N., & Kendeou, P. (2005). Integrating memory-based and constructionist processes in accounts of reading comprehension. *Discourse Processes: A Multidisciplinary Journal*, 39, 299–316.
- Vipond, D., & Hunt, R. A. (1984). Point-driven understanding: Pragmatic and cognitive dimensions of literary reading. *Poetics*, 13(3), 261–277. doi:10.1016/0304-422X(84)90005-6
- Wilhelm, J. D., & Smith, M. W. (2001). Literacy in the lives of young men: Findings from an American study. *English in Australia*, 2001(132), 17–28.
- Williams, J. P. (1993). Comprehension of students with and without learning disabilities: Identification of narrative themes and idiosyncratic text representations. *Journal of Educational Psychology*, 85(4), 631–641. doi:10.1037/0022-0663.85.4.631

- Williams, J. P., Lauer, K. D., Hall, K. M., Lord, K. M., Gugga, S., Bak, S.-J., ... deCani, J. S. (2002). Teaching elementary school students to identify story themes. *Journal of Educational Psychology, 94*(2), 235–48.
- Winner, E., Rosenstiel, A. K., & Gardner, H. (1976). The development of metaphoric understanding. *Developmental Psychology, 12*(4), 289–297. doi:10.1037/0012-1649.12.4.289
- Zajonc, R. B. (1984). On the primacy of affect. *American Psychologist, 39*(2), 117–123.
- Zwaan, R. (1999). Situation models: The mental leap into imagined worlds. *Current Directions in Psychological Science, 8*(1), 15–18.

Appendix

The Death of the Ball Turret Gunner

Randall Jarrell

From my mother's sleep I fell into the state
 And I hunched in its belly till my wet fur froze
 Six miles from earth, loosed from the dream of life
 I woke to black flak and the nightmare fighters.
 When I died they washed me out of the turret with a hose.

The Lame Boy Returns

Juan Delgado

My smile mocked your speed –
 Your nicknames dragging behind.
 I can't recall your name.
 Has it been that long?

You bobbed and planted your crutch,
 A re-carved bedpost, stumping a trail.
 Your other hand balanced your stride
 And waved, but I ran to my friends.
 You heard our jokes and still played along,
 Chasing our laughter through the street
 Finding me in a circle of flaring faces
 Planning to run even faster from you.

Your leg did not fade
 Like a childhood fear
 Like the creaking of my dark house.

Your limp is more than flesh,
 Casting a larger shadow now.

Corresponding author

Sarah Levine
School of Education and Social Policy
Walter Annenberg Hall
2120 Campus Drive
Northwestern University
Evanston, Illinois 60208
sarahlevine2013@u.northwestern.edu

Second author

William S. Horton
Weinberg College of Arts and Sciences
Department of Psychology
Northwestern University
Swift Hall 102
2029 Sheridan Road
Evanston, IL 60208-2710
whorton@northwestern.edu