Assessing Literary Reasoning: Text and Task Complexities

Carol D. Lee[1] and Susan R. Goldman[2]

[1]Northwestern University

[2]University of Illinois, Chicago

cdlee@northwestern.edu

sgoldman@uic.edu

Acknowledgments

To appear in *Theory into Practice.*

Assessing Literary Reasoning: Text and Task Complexities

Carol D. Lee and Susan R. Goldman

Citizens of the 21st century society need to be able to critically read and evaluate the burgeoning information resources relevant to personal, academic, and professional life. Yet, there is ample evidence that the literacy levels of the majority of graduating high school students are inadequate for these purposes (Carnegie Council on Advancing Adolescent Literacy (CCAAL), 2010; NAEP, 2009; OECD, 2013). Responses to this gap call for greater emphasis on analysis and critical reading in the context of complex argumentation and problem solving tasks (Achieve, 2013; Council of Chief State School Officers, 2010). Many argue that these tasks are discipline-specific and there must be more attention paid to disciplinary literacies (e.g., Goldman, 2012; Lee & Sprately, 2010; Moje, 2008). As great as the challenges are for instruction that creates opportunities to develop competency in these areas, the challenges of assessing them are even greater. These challenges include adequate conceptualization of the knowledge and skills that underlie critical reading to accomplish authentic disciplinary tasks, developmental progressions in them, and unattended to aspects of text complexity (Goldman & Lee, 2014). Attention to these challenges is critical if we are to develop assessments that are authentic and reliable.

Our goal in this article is to discuss these challenges, illustrating them using the specific case of literary reading. To frame our discussion, we use the assessment triangle conceptual approach (Pellegrino, Chudowsky, & Glaser, 2001) and the Evidence-

Centered Design (ECD) approach to assessment design (Mislevy, Steinberg, & Almond, 2003).

The Assessment Triangle and the Evidence Centered Design Process

Pellegrino et al. (2001) defined assessment as a process of determining what students know relative to what we want them to know. The process involves gathering evidence for purposes of making claims about what students can do and the knowledge, skills, practices, and dispositions that underlie what they do. However, the knowledge and processes that underlie observable performance are often not visible. Pellegrino, et al. (2001) proposed that three components of the assessment reasoning process have to be aligned: (1) a model of student *cognition* and learning in the domain being assessed; (2) assumptions about the types of *observations* that can serve as evidence of what students know; and (3) a process of *interpreting* the evidence given the form of the observation and the purpose of the assessment. For quality assessments, these three components of have to be aligned. ECD (Mislevy, et al., 2003) provides a systematic approach to accomplishing this.

The ECD process has three major components that define a Domain Model: the Student Model, the Task Model, and the Evidence Model. The Student Model defines the cognition - the knowledge, skills, and practices that constitute competence in the domain. These are expressed in claim – evidence statements. Claim statements articulate what students should know and be able to do (e.g., *The student can distinguish ironic from non-ironic cartoons.*). Evidence statements linked to specific claims indicate what observables students would need to produce in support of inferences about competencies (e.g., *The student can explain what juxtaposition of elements makes a cartoon ironic.*).

Task Models articulate the characteristics of assessment situations that are used to generate observations in the form of student work (e.g., types of cartoons, written response). The Evidence Model specifies how to interpret the observations with respect to claims about students.

In this essay we focus on the challenges of specifying Domain Models for critical reading, reasoning, and argumentation in the disciplines, using as our primary lens literary reading. (See Goldman, et al. (2012) on use of ECD for multiple source comprehension assessment in history and science.)

Domain Analysis

Developing a domain model typically begins with an analysis of the knowledge, skills, and practices that define competent performance in the domain. Domain analysis draws on theory and research that identifies what those highly experienced or expert in the domain know and can do, and the knowledge, skills, strategies, and practices that underlie them. Examples of such research include studies of document-based inquiry by historians or history teachers (e.g., Leinhardt & Young, 1996; Wineburg, 1991), think-aloud studies of literary experts interpreting a story or poem (e.g., Graves & Frederiksen, 1991; Warren, 2011), or the reading behavior of chemists and physicists (e.g, Bazerman, 1985; Shanahan & Shanahan, 2008). In the IES funded Project READI (Goldman, et al., 2009), we drew on such literature to articulate domain analyses for literary interpretation, historical reasoning, and text-based scientific reasoning. The domain analyses are similar in that each uses the same five dimensions to capture the multi-dimensionality of the knowledge, skills, and practices involved in reading and reasoning in these disciplines. The dimensions of the READI domain analysis framework are (1) epistemology, (2)

inquiry strategies (3) key concepts and frameworks, (4) types of texts, and (5) discourse and ways of using language. Epistemology refers to conceptions of what forms of reasoning are valued, what is considered worth knowing, and how new knowledge is generated. Inquiry strategies are the processes and ways of thinking about information within and across sources. Key concepts include the big ideas, disciplinary core ideas, and forms of knowledge upon which readers in each discipline draw in making sense and evaluating texts. Types of texts include the prototypical genres within each discipline, including static and dynamic visuals. Discourse and language include the features of language that are valued in the discipline. In Table 1 below, we offer a very truncated articulation of these dimensions in literature. The specification of these dimensions differs for history or science (cf. Goldman, 2012).

Insert Table 1

The Student Model

The student model is guided by the domain analysis and indicates what learners are expected to know and be able to do at different levels of progression. One challenge in moving from the domain analysis to claim – evidence statements is the multi-dimensionality of literary reasoning and thus the interconnectedness of the dimensions in the framework. Consider the claim *The student can interpret symbolic language in a story to make inferences about a character and how that character does or does not instantiate a prototypical character type.* Such inferences depend on aspects of the key concepts, in particular moral and philosophical orientations, as well as knowledge of language, types of texts and types of characters.

Further compounding this challenge is that literary epistemology values argumentation over "right" interpretations (Rosenblatt, 1978; Tompkins, 1980). The epistemological orientation and inquiry strategies that readers draw on in literary interpretation can vary widely leading to many degrees of freedom in the range of warrantable interpretations (Scholes, 1985). This contrasts with other disciplines where the inquiry strategies and criteria for valid evidence result in a more restricted range of valid interpretations (Goldman, 2012). For literature, even multiple choice tasks with more than one "right" answer are problematic because an expert reader can likely justify a response not available or argue against the "right" response. Evidentiary supports can come from different sources, depending on the reader's interpretive community, an aspect of Key Concepts (e.g. appeals internal to the text – New Criticism, Structuralism -, to the historical/social/political context of the setting or publication date, to critical moral or ideological commitments – Black Criticism, Feminist Criticism, Marxist Criticism, or to the personal beliefs, interests and commitments of the reader – Reader Response Theory). Appleman (2000) and others have argued that high schools pay too little attention to helping students develop a wide repertoire of approaches to interpreting literary works (Applebee, Burroughs, & Stevens, 2000). In short, Student Models of literary reasoning need to incorporate competence with multiple ways of approaching the interpretation of literary works (Lee, 2011). This requirement of rigorous assessments of literary reading complicates the Task and Evidence Models.

There are also complications in specifying the student model for literature that come from the way in which developmental progressions in reading are typically described using the same generic skills (e.g., summarizing) but applied to increasingly

complex texts (Goldman & Lee, 2014). There are three issues raised by this account of developmental progressions in reading. First, there has been little attention to how a generic skill like summarizing might differ in complexity depending on the purpose for reading or the discipline in which the task is situated. For example, creating a summary of the chronology of events in a narrative might be quite different from a summary focused on critical incidents contributing to the transformation of a protagonist. Indeed these two purposes would define different claim – evidence statements and be informative about different aspects of students' domain knowledge in literature. Likewise, creating a summary of key events in a historical account, in addition to being a different task than summarizing the plot of a literary work would likely require attention to different aspects of the discourse and language of the text. . In short, both the text and the task need to be considered in determining complexity (Valencia, Wixson, & Pearson, 2014).

The second issue is that traditional measures of text complexity are limited because they only account for surface features of syntax and vocabulary (Goldman & Lee, 2014). But even for vocabulary, these measures obfuscate complexity. For example, in literary texts, simple concrete words can serve figurative pragmatic functions. Coh-Metrix, a quantitative system for assessing text complexity does pay attention to coherence and this can be very useful with expository texts in history and science (Graesser, McNamara, & Kulikowich, 2011). However, representations of coherence can look quite different in literature. For example, Coh-Metrix indices of higher narrativity indicate easier comprehension (Graesser, et al., 2011). However, this claim is based on measures typically derived from short texts or short selections from

novels. The fact that narrative structures can change within longer works such as novels is not captured by Coh-Metrix. Yet such a change can increase the complexity of a literary work. From an ECD perspective, text complexity considerations would be part of the task model. Our argument is that important sources of text complexity are inherent to features of literary works including plot structure, figuration, character complexity, and point of view (Lee, 2011). As such they are important foci for instruction and not simply parameters in a task model.

Text – task considerations aside, other dimensions of the domain analysis are entailed in learning progressions for disciplinary interpretation and reasoning. In particular, disciplinary Epistemological Orientations and Key Constructs/Frameworks are influenced by maturation over time. For example, there are dimensions of understanding historical time and political dilemmas in history (Coles, 2000), moral dilemmas interrogated in literature (Kohlberg, 1981), and scientific processes not subject to direct observation in science to which 6th graders and 11th graders will bring very different resources and abilities. And even then, children's and adolescents' abilities to wrestle with these will be influenced by differences in life experiences (Kunda 1999). Consider the novel *The Grapes of Wrath*, which has a lexile level of 680, suggesting that a 6th grader reading on or above grade level can read the words on the page. However, it is likely that such a 6th grader will have some difficulties interrogating the moral, social, political and economic dilemmas posed in the novel compared to a 10th or 11th grader, reading at the 5th or 6th grade level who is likely to have more mature social cognition; or a 6th grader who is reading at the 5th grade level but is from a migrant working family and who can bring life experiences that his or her middle class peer is not likely to have.

Thus we argue that the domain model for literary reasoning is a complex problem space. Tackling it requires that we draw from and coordinate across multiple disciplines: cognitive work in the nature of knowledge, reasoning, and problem solving in each discipline (Achieve, 2013; Bransford, et al., 2006, Lee 2007; Wineburg, 2001); processes of reading comprehension (Kintsch, 1998, RAND Reading Study Group 2002); linguistics (Halliday & Hasan 1976; Webster 2007); text complexity (Goldman & Lee 2014; Pearson & Hiebert 2014; Valencia, et al., 2014); and human development around moral development and social cognition (Flavell & Miller 1998; Kohlberg, 1981; Spencer, 1985). Reading comprehension always involves complex and dynamic processes contingent on text and task, but also resources the reader brings, including motivation, goals, knowledge of content, text structure, vocabulary, syntax, strategies, relevant constructs, broader rhetorical patterns, as well as technical and social supports.

The Task and Evidence Models

The Task Model articulates the conditions for collecting observations and the Evidence Model for how these observations are measured and interpreted relative to a claim about what the student knows. Leaving aside the issues of task and text complexity, another central challenge in Task and Evidence Models for assessments of critical reading, evaluation and argumentation in the disciplines, is how to "observe" thinking – how to make thinking visible. Creating external representations of internal reading comprehension processes poses distinct challenges for all disciplines, but especially for literature. This became evident through our experiences in attempting to design task situations to externalize thinking processes of literary interpretation in Project READI.

The Project READI Literature team designed assessment task situations that were aligned with the logic of the READI instructional model and the domain analysis shown in Table 1: explicit teaching of heuristics and strategies for interpreting problems central to the discipline (e.g. symbolism, irony, satire, unreliable narration), articulating criteria for making judgments (e.g. warranting evidence around claims about internal states of characters, themes, and the interpretive problems we have identified) (Lee, 2007; 2011). Articulating the Evidence Model - what to observe and how to measure it – in relation to the claim-evidence statements required careful consideration of several key requirements of the tasks and the texts that would be used in them: text complexity in pre-post assessments; potential counter explanations for weaknesses in students' reasoning in essays that could reflect writing skills rather than interpretive skills; measuring the nature of the reasoning students employ. We will briefly describe these challenges and how we have tried to address them.

Creating task situations with texts of equivalent difficulty

Beyond grade level formulas, we matched texts on the basis of common themes, complexity of characterization, rhetorical features, and interpretive problems. For example, for a task that would produce evidence relative to claims about symbolic and thematic interpretation, we developed a text set focused on a coming of age theme and symbolism at a high level of abstraction. To address development across grades we designed the same assessments for middle grades, early and late high school. In the first iteration, we could not really equate texts, finding that lexiles and grade level readability often completely obfuscated the complexity of the text for the focus on theme and symbolism. Anticipating this problem, we counter balanced the order of text sets within

classrooms so that we could test for impact of text set. We found that the order of text sets mattered for middle and early high school, but not for late high school, although there were some younger students for whom this did not matter and some older students for whom it did. The complexity of the texts was contextualized with regard to the maturity of the age cohort.

Externalizing the internal

The initial assessments included only essays. We then faced the challenge of distinguishing whether the problem was one of comprehension versus composition. We used these findings to re-design both instructional units and assessments. Most recently, we introduced two additional tasks. One is an ordering task where students identify plot elements and their order for a crude indicator that basic comprehension of what is going on is not the source of the problem. The second is a graphic organizer to structure an argument around symbolic meanings in each of the two stories. The purpose of the graphic organizer is two-fold: a non-essay representation to show they have claims, what is the evidence, and what reasoning they have employed to argue that we should believe their evidence; a structure to organize thinking about the complex rhetorical problem of addressing similarities and differences regarding symbolic meanings in two texts.

Analytic approach to assessing literary interpretive arguments

The third challenge involved developing rubrics to disentangle: (1) the quality of the claims, (2) the nature of evidence and the degree to which it was supportable, minimally by the text, (3) whether and how students made explicit how they reasoned from claims to evidence, (4) the ability to organize the essay around similarities and differences, and (5) general issues of clarity of communication. We have been using

11

rubric score points for each of these dimensions of literary interpretive arguments at the same time that we are attempting to extract a holistic measure, although we are not certain a holistic measure would be useful in our analysis. We have developed several iterations of this rubric in efforts to deal with these challenges. While rubrics are routinely used to assess writing, even in large-scale assessments, these rubrics often do not address the specific demands of writing arguments from texts in a particular domain.

New Directions in Assessments

We are not alone in efforts to deal with the challenges of assessing critical reading and evaluation of texts in various disciplines. For example, in two cases with which we are familiar – the Educational Testing Service's scenario-based comprehension assessments (Bruce, Halderman, O'Reilly, & Sabatini, 2014) and the PARCC assessments, student thinking is assessed using multiple response formats that have different balances of recognition and production demands (e.g., multiple choice and constructed responses). ETS is also attempting to take into account previously backgrounded dimensions of comprehension situations, including motivation, goals, prior content knowledge and social context.

At the same time, both the ETS and the PARCC assessments have some significant challenges. For example, although students write on-demand essays to argue for a claim about a multiple text set, students are not required to articulate criteria (e.g. warrants) for why the evidence presented (typically only evidence from the texts) supports the claim (Toulmin, Rieke, & Janik, 1984). Furthermore, it remains to be seen how well these assessments address the multiple dimensions of disciplinary

comprehension. These overall limitations are important because assessments, particularly those used for accountability, can influence instruction on the ground.

Conclusion

The need for students to critically read and evaluate information from multiple texts and use it to both understand and generate arguments to address disciplinary problems introduces new assessment challenges. These challenges reflect the technical difficulties of assessing text complexity, especially in literature, and of creating external representations of internal thinking processes. As well there is the broader political contexts of instituting standards and rigorous assessments widely. We saw this problem with the New Standards Project, with efforts to institute at scale performance based assessments, and we revisit these problems with the implementation of Common Core and the emerging assessments across states. We need to find ways of bringing these challenges – conceptual, technical, and political – into the forums (publications, funded research, research conferences, public reports, research syntheses) we have available to begin to wrestle with new frontiers. If we are to address the equity challenges of equalizing opportunity to learn for all, we need to be able to push the bounds of how we conceptualize assessments of reading in the disciplines, particularly because reading is a fundamental resource for learning in the content areas (Moss, Pullin, Gee, Haertel, & Young, 2008).

References (1132 words)

Achieve (2013). *The Next Generation Science Standards.* Washington, DC: Achieve.

Applebee, A., R. Burroughs and A. Stevens (2000). Creating continuity and coherence in high school literature curricula. *Research in the Teaching of English, 34 (*3), 396-429.

Appleman, D. (2000). *Critical encounters in high school English : teaching literary theory to adolescents.* NY, Teachers College Press.

Bazerman, C. (1985). Physicists reading physics: Schema-laden purposes and purpose-laden schema. *Written Communication, 2,* 3-23.

Bransford, J., R. Stevens, D. Schwartz, A. Meltzoff, R. D. Pea, J. Roschelle, N. Vye, P. Kuhl, P. Bell, B. Barron, B. Reeves and N. Sabelli (2006). Learning theories and education: Toward a decade of synergy. In P. A. Alexander and P. H. Winne (Eds.), *Handbok of Educational Psychology* (pp. 209-244). Mahway, NJ, Lawrence Erlbaum Associates

Bruce, K. M., Halderman, L., O'Reilly, T., & Sabatini, J. (2014, April). *Developing Theory-Based Literacy Assessments for Pre-K-12 Students: Challenges and Opportunities*. Paper presented at American Educational Research Association, Philadelphia, PA.

Carnegie Council on Advancing Adolescent Literacy. (2010). *Time to act: An agenda for advancing adolescent literacy for college and career success*. NY: Carnegie Corporation of New York.

Coles, R. (2000). *The political life of children*. NY, Atlantic Monthly Press.

Council of Chief State School Officers (CCSSO) and the National Governors Association (NGA) (2010). *The Common Core State Standards for English Language Arts & Literacy in History/Social Studies, Science, and Technical Subjects*. [Retrieved July, 16, 2010: http://www.corestandards.org/the-standards/english-language-arts-standards]

Flavell, J. H. & Miller, P. H. (1998). Social Cognition. In W. E.I.C Damon, D. Kuhn, & R. V. E. Siegler (Eds.) *Handbook of Child Psychology Vol. 2,* (pp. 851-898). NY, Wiley.

Goldman, S. R., et al. (2009). *Reading for understanding across grades 6 through 12: Evidence-based argumentation for disciplinary learning.* Funded July, 2010 by the Institute of Education Sciences, U.S. Department of Education, Grant # R305F100007.

Goldman, S. R. (2012). Adolescent literacy: Learning and understanding content. *Future of Children, 22*, 89 – 116.

Goldman, S. R. & Lee, C. D. (2014). Text Complexity: State of the Art and the Conundrums It Raises. *Elementary School Journal.*

Goldman, S. R., Lawless, K. A., Pellegrino, J. W., Braasch, J. L.G., Manning, F. H., & Gomez, K. (2012). A technology for assessing multiple source comprehension: An essential skill of the 21st century. In M. Mayrath, J. Clarke-Midura, & D. H. Robinson (Eds.). *Technology-Based assessments for 21st century skills: Theoretical and practical implications from modern research* (pp. 171 - 207). Charlotte, NC: Information Age Publishing.

Graesser, A.C., McNamara, D.S., & Kulikowich, J. (2011). Coh-Metrix: Providing multilevel analyses of text characteristics. *Educational Researcher, 40,* 223-234.

Graves, B. & Frederiksen, C. H. (1996). A cognitive study of literary expertise. In R. J. Kruez & M. S. MacNealy (Eds.), *Empirical Approaches to Literature and Aesthetics* (pp. 397-418). Norwood, NJ: Ablex.

Halliday, M. A. K. & Hasan, R. (1976). *Cohesion in English.* London: Longman.

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

Kohlberg, L. (1981). *Essays on Moral Development, Vol. I: The Philosophy of Moral Development.* San Francisco, CA, Harper & Row.

Kunda, Z. (1999). *Social cognition: making sense of people.* Cambridge, MA: MIT Press.

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

Lee, C. D. (2007). *Culture, Literacy and Learning: Taking Bloom in the Midst of the Whirlwind.* NY, NY: Teachers College Press.

Lee, C. D. , & Spratley, A. (2010). *Reading in the Disciplines and the Challenges of Adolescent Literacy.* NY, NY: Carnegie Foundation of New York.

Leinhardt, G., & Young, K. M. (1996). Two texts, three readers: Distance and expertise in reading history. *Cognition and Instruction*, *14*, 441–486.

Mislevy, R. J., Steinberg, L., & Almond, R (2003). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspective, 1*, 3-67.

Moje, E. B. (2008). Foregrounding the disciplines in secondary literacy teaching and learning: A call for change. *Journal of Adolescent & Adult Literacy, 52*, 96-107.

Morrison, T. (1970). *The bluest eye*. NY, NY: Alfred A. Knopf.

Moss, P., Pullin, D., Gee, J. P., Haertel, E., L. J. Young, L. J. (Eds.) (2008). *Assessment, Equity and Opportunity to Learn.* NY, NY: Cambridge University Press.

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

Organization of Economic and Cultural Development (2013). *PISA 2012: Results in focus.* Paris: OECD.

Pearson, P. D. & Hiebert, E. H. (2014). The State of the Field: Qualitative Analyses of Text Complexity. *Elementary School Journal*

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (2001). *Knowing what students know: The science and design of educational assessment.* Washington, DC: National Academy Press.

RAND Reading Study Group (2002). *Reading for understanding: Toward an R&D program in reading comprehension*. Santa Monica, CA: RAND.

Rosenblatt, L. (1978). *The Reader, The Text, The Poem: The Transactional Theory of the Literary Work.* Carbondale, IL: Southern Illinois University Press.

Spencer, M. B. (1985). Cultural cognition and social cognition as identity factors in Black children's personal-social growth. In M. Spencer, G. K. Brookins & W. Allen *Beginnings: The social and affective development of Black children* (pp. 59 - 72). Hillsdale, NJ, Lawrence Erlbaum**.**

Tompkins, J. (1980). *Reader-response criticism: From formalism to post structuralism.* Baltimore, MD, Johns Hopkins University Press.

Toulmin, S., Rieke, R., & Janik, A. (1984). *An introduction to reasoning.* NY, NY: Macmillan Publishing Company.

Valencia, S. W., Wixson, K., & Pearson, P. D. (2014). Putting text complexity in context: Refocusing on comprehension of complex text. *Elementary School Journal.*

Warren, J. E. (2011). "Generic" and "specific" expertise: An expert/expert study in poetry interpretation and academic Argument. *Cognition & Instruction, 29,* 349-374.

Webster, J. J., Ed. (2007). *Language and Education: M. A.K. Halliday.* NY, NY: Continuum.

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

Wineburg, S. (2001). *Historical thinking and other unusual acts: Charting the future of teaching the past.* Philadelphia, PA: Temple University Press.

Table 1

READI Framework of Dimensions of Knowledge Informing Literary Reading

Dimensions of Knowledge, Skills, Practices	Literature
Epistemology	• Interrogating human experience • Open to dialogue between readers and texts • Focus on content and functions of form
Inquiry Strategies	• Infer details (stated & implied) regarding plot, characters, themes, and structural patterns) • Evaluate how rhetorical strategies shape reader's response • Draw on prior knowledge (social milieu of the text, moral & philosophical precepts, other texts, author, rhetorical traditions, reader's experiences)
Key Concepts/Frameworks	• Moral & philosophical content (e.g. archetypal themes) • Historical contexts • Critical theories • Intertextuality
Types of Texts	• Plot structures (e.g. coming of age, myth, allegory) • Poetic forms (e.g. sonnets, ballads) • Character types (e.g. epic hero, trickster, tragic hero)
Discourse & Ways of Using Language	• Imagery • Figuration (e.g. symbolism, irony, satire) • Problems of point of view • Rhetorical strategies & patterns (e.g. parallelism, repetition, exaggeration, etc.)