Research Report

ETS RR–11-14

Writing Assessment and Cognition

Paul Deane

April 2011

Writing Assessment and Cognition

Paul Deane

ETS, Princeton, New Jersey

April 2011

Technical Review Editor: Dan Eignor

Abstract

This paper presents a socio-cognitive framework for connecting writing pedagogy and writing assessment with modern social and cognitive theories of writing. It focuses on providing a general framework that highlights the connections between writing competency and other literacy skills; identifies key connections between literacy instruction, writing assessment, and activity and genre theories; and presents a specific proposal about how writing assessment can be organized to promote best practices in writing instruction.

Key words: writing, assessment, CBAL, cognitive, competency model, evidence-centered design, learning progressions, reading, literacy

Acknowledgments

Table of Contents

List of Tables

List of Figures

More than anything else, this paper is about connections:

- Connections between writing and reading

- Connections between writing and critical thinking

- Connections between writing and its social context

- Connections between how writing is tested and how writing is taught

The context is an ongoing effort at ETS to develop a new approach to K–12 writing assessment in which these connections are not only respected but also deeply embedded into the very design of the assessment. Writing is not an isolated skill. It builds upon a broad foundation of prerequisite literacy skills, both supports and requires the development of critical thinking skills, and requires the writer to solve a complicated array of rhetorical, conceptual, and linguistic problems.

None of these themes are new in and of themselves. To point out a few of the more salient discussions, Shanahan (2006) examined complex interconnections and interdependencies among reading, writing, and oral language. Applebee (1984) reviewed older literatures connecting writing to the development of critical thought, while Hillocks (1987; 1995; 2002; 2003b) emphasized the importance of inquiry in writing, noting that students need above all to learn strategies that will enable them to think about the subject matter of their writing (Hillocks, 2003a). And the literature on the social aspects of writing is even more extensive, so that the comments that follow can do little more than indicate major themes.

In recent years a number of themes have been emphasized. Literacy is a complex, varied, highly nuanced class of social practices in which school literacy has a privileged but specialized position in our society. Students who may do poorly on literacy tasks in a school setting may yet display considerable sophistication on related skills embedded in well-defined, socially significant practices (Hull & Schultz, 2001). Reading and writing are not monolithic entities but complex skill sets deployed in historically contingent contexts; that is, the choices of forms and genres available to the author, and the modes of communication and interaction with which they are associated, have evolved and are evolving under the influence of social and technological factors (Bazerman & Rogers, 2008; Bolter, 2001; Foster & Purves, 2001; Heath, 1991; Holland, 2008; Murray, 2009; Street, 2003; Venezky, 1991). Education in reading and writing should be viewed not simply as the inculcation of a skill set, but as socialization into literate communities,

and therefore as learning how to participate in a specific set of concrete and socially valued practices (Barab & Duffy, 1998; Barton & Hamilton, 1998; Barton, Hamilton, & Ivanic, 2000; Carter, 2007; Englert, Mariage, & Dunsmore, 2006; Lave & Wengler, 1991; Marsh & Millard, 2000; Reder, 1994; Resnick, 1991). There is broad consensus that writing skill is most effectively acquired in a context that makes writing meaningful, both in relation to its content and to the social context within which writing takes place (Alverman, 2002; Graham & Perin, 2007; Langer, 2001).

Criticisms of particular methods of writing assessment often revolve around the contrast between the testing situation and the situation in which writers ordinarily write. For instance, in a timed impromptu essay examination, the writer may have

- no control over the topic, and often little knowledge or interest in it;

- no access to any source of information about the topic;

- little time to think deeply about the topic; and

- considerable incentive to focus on surface form (since the scoring rubric may penalize grammatical mistakes or favor those students who produce the standard five-paragraph essay).

And yet this list of flaws (from the writer's point of view) can readily be transformed into a list of virtues (from a test administrator's point of view), such as fairness, uniformity of testing conditions, objectivity and consistency of scoring, and efficiency. In short, progress in writing assessment requires us to reconcile the twin virtues of validity and cost, which are often in tension, and which may lead to fundamentally different solutions, with fundamentally different implications for instruction.

Assessment constitutes a social context in its own right. It holds a central place in our educational institutions and has a powerful impact upon instruction, not always for the better. What teachers teach is strongly influenced by what is on the test and even by seemingly minor details of test format. Frederiksen (1984) discussed a variety of ways in which the format of a test and the implicit link between instruction and assessment can have unintended consequences. As Frederiksen put it:

The "real test bias" in my title has to do with the influence of tests on teaching and learning. Efficient tests tend to drive out less efficient tests, leaving many important

2

abilities untested—and untaught. An important task for educators and psychologists is to develop instruments that will better reflect the whole domain of educational goals and to find ways to use them in improving the educational process. (p. 201)

Responses to this issue have gradually led toward broader use of performance-based assessments in writing. As Yancey (1999) noted, the general trend from the 1950s to the 1970s was to assess writing indirectly with multiple-choice tests, with direct writing assessment and then portfolio-based assessment gradually entering the picture (Elliott, 2005; White, 2004). A landmark of direct writing assessment, Ed White's *Teaching and Assessing Writing* (1985) established holistic direct writing assessment as the norm; and White (2005) demonstrates a continuing focus on developing effective methods of writing assessment—in this case, methods of portfolio assessment that connect portfolio contents to curricular goals via student reflective writing. Yet considerable room exists for improvement, particularly if connections are taken into account—connections that make it almost impossible to assess writing meaningfully if it is viewed merely as an isolated skill.

In 1984, Norman Frederiksen made the following observation:

Over the past 25 years or so, cognitive psychologists have been investigating the mental processes that are involved in such tasks as reading, writing, solving puzzles, playing chess, and solving mathematical problems. The result is a theory of information processing that has important implications for teaching… Some of the cognitive processes that have been identified have to do with the development of internal representations of problems, the organization of information in long-term memory for efficient retrieval, the acquisition of pattern cognition and automatic-processing skills, use of strategic and heuristic procedures in problem solving, and how to compensate for the limited capacity of working memory. Such skills are not explicitly taught in schools today, but we are at a point where cognitive psychology can make substantial contributions to the improvement of instruction in such areas. (1984, p. 200)

Frederiksen postulated that this class of skills can most readily be tested with situational tests (that is, with tests that simulate the typical conditions under which such skills are used) and suggested the following:

Perhaps an adventuresome consortium of schools, cognitive scientists, and testing agencies could carry out demonstration projects to test the feasibility of systematically using tests to influence the behaviors of teachers and learners and to provide the large amount of practice needed to make the skills automatic. (p. 200)

The past 25 years have seen further progress in modeling the cognitive foundations of reading, writing, and other intellectual skills, and even greater progress in building socially as well as cognitively sophisticated models of instruction. But thus far, nothing like Frederiksen's vision has been realized, not least because it requires synthesis and coordination across several disciplines, and the solution of a wide range of practical and technical problems.

The nature of the problem can be measured in part by the kinds of difficulties encountered by the performance assessment and authentic assessment movements (Haertel, 1999; Hamilton, 2005): It can be very difficult to make an assessment more closely resemble real-life performance, or bring it more closely into alignment with best practices in instruction and curriculum, while meeting all of the other constraints intrinsic to summative assessment situations, including the powerful constraints of cost and the way that testing is budgeted in particular institutional settings. Instruction and curriculum are variable, as is practical performance outside a school setting, and both are dependent on context in ways that can make performances difficult to assess and compare. It is not easy to devise an assessment system that delivers good measurement, models the kinds of tasks teachers should prepare students to perform, and supports instruction. However, Bennett and Gitomer (2009) sketched out one possible strategy for dealing with these issues involving coordinated development of summative assessments, classroom assessments, and professional support materials. Bennett and Gitomer set as their goal an integrated assessment that did more than fulfill a simple accountability function. They advocated a form of assessment intended simultaneously to document student achievement (*assessment* of *learning*), support instructional planning (*assessment* for *learning*), and engage students and teachers in worthwhile educational experiences during the testing experience (*assessment* as *learning*). They argued that these goals could be achieved by leveraging advances in cognitive science, psychometrics, and technology to build much richer assessment experiences.

In 2009, the National Academy of Education issued a white paper on standards, assessments and accountability that endorsed a similar set of goals. The academy recommended

a series of summative assessment reforms in which modified test designs are based upon a strong cognitive foundation and coordinated systematically with support systems for classroom teachers (including professional development and support systems, parallel formative assessments, and other supports for classroom instruction).

The research reported in this paper applied Bennett and Gitomer's (2009) ideas to writing assessment in primary and secondary grades. It focused on three aspects of the overall vision:

- Understanding the cognitive basis for effective writing instruction

- Designing formative and summative writing assessment designs that meet Bennett and Gitomer's goal for assessment designs that use richer, more meaningful tasks, provide effective support for instruction, and constitute valuable learning experiences in their own right

- Conceptualizing an approach to essay scoring that maintains a strong rhetorical focus while using automated methods to assess key component skills.

These three topics will define the three main sections of this paper. Section 1 will document a cognitive framework for writing assessment. Section 2 will describe pilot assessment designs that instantiate this framework. Section 3 will sketch an innovative approach to essay scoring intended to make effective use of automated essay scoring techniques without substituting automated scores for human judgment about content and critical thinking.

A key conceptual element of the analysis to be presented derives from *activity theory* (Engestrom, Miettinen, & Punamaki, 1999), which treats interactions among people in a social environment as the fundamental unit of analysis. Particular institutions, the tools skills that enable people to participate in those systems, and the social conventions that govern interaction are all part of *activity systems* in which people act to accomplish goals that emerge from and are partially defined by the roles and situations in which they are participating. Activity theory leads directly to a constructivist view of learning, in which learning a skill emerges naturally from participating in the activities for which the skill is intended (Hung & Chen, 2002; Jonassen & Rohrer-Murphy, 1999). The fundamental goal of the research outlined in this paper is to help redefine writing assessment so that it more directly supports learning and helps to engage novice writers in appropriate communities and practices. The availability of online, computerized assessment and instructional tools presents an important opportunity to achieve this goal.

1. Writing as a Complex Cognitive Skill

1.1. Connections and Disconnections Among Writing, Reading, and Critical Thinking

Classical cognitive models of writing may disagree in points of detail but they agree in several common themes. One theme is that expert writing clearly involves at least the following elements:

- A set of *expressive* skills that enable fluent text production. In Hayes and Flower (1980) this was identified as the *translating* process. In Hayes (1996) it was *text production*. In Bereiter and Scardamalia (1987) it was the *knowledge-telling* process.

- A set of *receptive* skills that support self-monitoring and revision. In Hayes and Flower (1980) this was called the *reading* process. In Hayes (1996) it was *text interpretation*. In Bereiter and Scardamalia (1987) it was largely kept in the background except in Chapter 9, which argued for significant parallels between reading and writing processes, and Chapter 11, which presupposed self-reading as part of the feedback loop necessary to revision.

- A set of *reflective* skills that support strategic planning and evaluation. In Hayes and Flower (1980) reflective skills were distributed among the planning, monitoring, and editing processes. In Hayes (1996) these elements were unified into a single category labeled *reflection*. In Bereiter and Scardamalia (1987) the knowledge-transforming model was intended to capture strategic, reflective thought. It differed from the Hayes and Flower model by postulating distinct rhetorical and conceptual problem spaces and subjecting both to problem analysis and goal-setting processes.

Normally, given the nature of literacy as an integrated process of communication, one would expect to find parallel expressive, receptive, and reflective skills across tasks with similar domains in play. These are different modes of thought, but they invoke the same mental representations. A reader may start with letters on the page and end up with ideas. A writer may start with ideas and end up with letters on the page. A thinker may deal simultaneously with letters and words, sentences, paragraphs, documents, ideas, and rhetorical goals.

Classical models of writing also distinguish several forms of representation that play critical roles in the cognition of composition:

- *Social and rhetorical elements* are among the most complex aspects of writing skill, requiring the writer to be consciously aware of and able explicitly to model personal interactions (specifically rhetorical transactions between author and audience) and to respond strategically to social and institutional expectations. While this aspect of writing is somewhat backgrounded in Hayes and Flower (1980), it is foregrounded in Bereiter and Scardamalia (1987) in the form of the *rhetorical problem space* and a major theme in sociocultural accounts of the writing process, as discussed above.

- *Conceptual elements* (representations of knowledge and reasoning) are also critical in the classical cognitive models of writing. Bereiter and Scardamalia (1987) represented this aspect of writing skill as the *conceptual problem space*. By definition, the process of planning and evaluating writing must address its content, and as Hillocks (1987) and Graham and Perrin (2007) indicated, few things are more necessary to the writer than to have effective strategies for dealing with the subject matter that they wish to address.

- *Textual elements* (representations of document structure) also play a key role in all models of writing. From Hayes and Flower (1980) onward, document planning is largely a matter of deciding how to produce a coherent, well-structured text.

- *Verbal elements* (linguistic representations of sentences and the propositions they encode) are the essential targets of text production in every model of writing. While control of verbal elements is as much a part of oral language as writing, writing depends first and foremost upon fluency of verbal production (McCutchen, 2000).

- *Lexical/orthographic elements* (representations of how verbal units are instantiated in specific media such as written text) obviously also play a role in writing, though they are not in focus in the major cognitive accounts discussed above. See Berninger (2005).

Therefore, it is appropriate to conceptualize skills relevant to writing by modes of thought (receptive, expressive, or reflective) and by types of cognitive representation (social, conceptual, textual, verbal, or orthographic). Figure 1 presents a visualization of writing skills that embodies this understanding. It is possible to interpret Figure 1 as a list of competencies or skills, viewed

in an entirely cognitive mode, but a richer interpretation is also available. Figure 1 can be viewed as a kind of cross-section of cognitive processes likely to be taking place in close coordination during any act of writing. It can also be viewed as an inventory of the types of *activities* in which literate individuals commonly engage, and thus viewed as part of the definition of activity systems relevant to writing. The advantages to viewing Figure 1 in these ways will be explained later.

Note that Figure 1 presents these skills by providing a single action verb such as inquire, structure, or phrase, which is intended to name the intended activity (and to indicate indirectly what skills are therefore critical). Each layer of the model—social, conceptual, textual, verbal, and lexical/orthographic—covers a range of phenomena including those elements listed in Table 1, which helps to clarify the kinds of tasks and thought processes to which each mode of representation applies.

	Interpretation	Deliberation	Expression
Social Model	Situate	Reflect	Engage
Conceptual Model	Enrich	Rethink	Conceptualize
Discourse Model	Integrate	Plan/Organize	Structure
Verbal Model	Parse	Edit	Phrase
Print Model	Decode	Monitor/Correct	Transcribe

Figure 1. **Modes of thought and modes of representation in the literacy processes.**

8

Table 1

Activity/Skill Categories Relevant to the Writing Process

Level of representation	Range of activities and skills
Social	Intentionality (genre, role, purpose) Perspective (point of view, bias, voice) Affect (stance, evaluation, tone)
Conceptual	Exploration (review, reflection, description) Explication (generalization, definition, analysis) Modeling (synthesis, application, hypothesis-formation, experimentation) Judgment (evaluation, justification, criticism)
Textual	Document structure (organization, rearrangement) Cohesion (relevance, focus/emphasis, given/new, transitions, textual inference) Development (topics, elaboration)
Verbal	Vocabulary (word familiarity, word choice, paraphrase) Sentence structure (sentence complexity, sentence variety, sentence combining) Ambiguity/figures of speech (creative word use, semantic flexibility, clarification)
Lexical/ orthographic	Grammar & usage (standard English) Spelling & mechanics (conventional written form) Word-formation (inflection, derivation, word families) Code-switching (register, dialect)

The major headings in Table 1 can briefly be defined as follows:

- Social Skills

  - *Empathize*—The ability to interpret documents or other forms of communication in a rich, socially perceptive fashion that takes into account the motivations, perspectives, and attitudes of author, intended audience, and individuals referenced in the text. This heading involves forms of inference based upon social skills and the ability to model human interaction.

  - *Engage*—The ability to communicate with an audience in a disciplined and effective way, focusing on achieving a particular purpose, and maintaining a voice and tone appropriate to that purpose

  - *Collaborate*—The ability think reflectively while working collaboratively in the full range of social practices common to highly literate communities (such as critical interpretation of text, presentation of research results, and reasoned

9

argumentation) with full sensitivity to the social, cognitive, and emotional transactions that such social practices may entail, including choice of register and genre to suit the social situation, and rhetorical purpose, choice of stance, and sensitivity to multiple perspectives.

- Conceptual Skills

  - *Infer*—The ability to subject a document or a set of documents to *close reading*, in which the reader goes beyond literal meaning to engage the ideas presented and integrate them deeply with prior knowledge. This involves the kinds of inference typically referred to as *bridging inference* and more active forms of text interpretation requiring close attention to conceptual content.

  - *Inquire*—The ability to develop ideas in an organized and systematic way such that they can be presented clearly and convincingly to someone who does not already understand or believe them

  - *Rethink*—The ability to evaluate, critique, and modify one's own or another's ideas using evidence and logical reasoning

- Textual Skills

  - *Integrate*—The ability to read a document and build a mental model of its content and structure. This definition is intended to include what current reading theories refer to as the construction of the *text base*. What reading theories refer to as the *situation model* requires mobilization of conceptual and social inferencing, which can go well beyond information directly available in the text.

  - *Structure*—The ability to produce a written document that follows an outline or some other well-structured textual pattern.

  - *Plan/Revise*—The ability to conceive a document structure that does not exist and plan that structure to serve a rhetorical purpose, or conversely, upon determining the structure of an existing document, to evaluate how well it organizes and presents its content, and rework it accordingly.

- Verbal Skills

    - *Understand*—The ability to understand texts written in standard English; that is, the ability to extract literal meaning from a sequence of sentences. This element (in combination with the ability to handle complex document and textual structures) is critical in constructing a literal understanding of a document (or *text-base*), though success at understanding phrases and sentences does not guarantee an adequate understanding of a complex text.

    - *Phrase*—The ability to express oneself in standard English; that is, the ability to find the right words and phrasings to convey one's intended meaning

    - *Edit*—The ability to identify places in a text where word choice and phrasing do not convey the intended meaning clearly and accurately, and then to come up with alternative phrasings that work better in context

- Orthographic Skills

    - *Read*—The ability to take printed matter and read it either aloud or silently; that is, the ability to convert characters on the page into mental representations of words and sentences

    - *Inscribe*—The ability to take words and sentences and convert them into printed matter; that is, the cognitive and motor abilities needed to produce words and sentences in written form

    - *Proofread*—The ability to examine printed materials, identify nonstandard patterns and errors, and modify them so that they conform to the norms of standard English grammar and orthography

Cognitive models also highlight aspects of writing skill that depend upon more general features of cognition (Bransford, Brown, & Cocking, 1999). The role of short-term memory and long-term memory, for instance, can hardly be neglected (Kellogg, 1996, 1999, 2001). And yet accounts of reading and writing processes emphasize trade-offs between automated and strategic processes (McCutchen, 1988, 1996, 2006). Skilled writers combine *efficient* receptive and expressive skills with *appropriate and effective* reflective strategies.

1.2. Connections and Parallelisms Among Writing, Reading, and Critical Thinking Skills

One advantage of the kind of analysis presented above is that it highlights the extent to which complex verbal skills draw upon the same underlying capacities. Figure 1 can be read simultaneously as a specification of skills that underlie writing and as a broad inventory of literacy skills. One set of arrows followed out from the center, from orthographic to social, closely tracks skills that would be highlighted in a model of reading competency: the abilities to decode written text, apply basic verbal skills, build up a literal interpretation of the document, and then create a situation model reflecting a conceptual model of document content and a rhetorical understanding of the writer's purpose. Another set of arrows followed inward from social to orthographic, closely tracks skills that are highlighted in writing assessment: the abilities to assess the rhetorical situation, understand the concepts to be communicated, plan a document that will communicate particular concepts and achieve particular rhetorical purposes, convert that plan into phrases and sentences, and express them in written form. The third set of arrows, followed either inward or outward, deals in the outer layers with skills normally highlighted in accounts of critical thinking and in the inner layers with revision, editing, and proofreading, textual skills closely associated with the critical evaluation of texts.

It would be possible simply to equate reading with receptive skills, writing with expressive skills, and critical thinking with reflective skills, but that would be an oversimplification. For instance, reading skill is often taken to include all the activities that support effective comprehension, which may include writing notes, asking reflective questions, and participating in a range of other activities that are not reading activities in and of themselves but which are being used to support reading. In the same way, writing skill includes a whole range of skills that involve reading and critical thinking, particularly during revision. And it is fairly clear that skilled critical thinkers (at least in a literate society) will deploy a variety of reading and writing activities in support of reasoning.

In other words, reading, writing, and critical thinking appear to be mutually supporting and highly entangled. Every skill noted in Figure 1 matters for writing. But the same skills appear to matter for reading, too, with a different emphasis. The skills that are most important for reading play a supporting role in writing competency; but conversely, skills that are critical for writing play supporting roles in enhancing reading comprehension.

Reading, writing, and critical thinking can thus reasonably be viewed as different but complementary *activity types* that share a common *underlying skill set.* They have complementary purposes (such as comprehension, explanation, and negotiation of common ground) but combine in specific ways to define the practices of a literary community. In activity theory terms, the literacy skill set—that is, the elements listed in Figure 1—can be viewed as activities that function as Vygotskian tools for members of a literate community of practice. Novice writers may have to learn some of the skills in the toolkit, but above all they have to learn how to coordinate them in the ways that enable them to create effective written texts. The difference between reading, writing, and critical thinking is defined by the final goal of activity, but in the course of accomplishing that goal, a writer may call upon any skills drawn from any of the categories in Figure 1 and may combine them in strategic ways.

Aligning reading and writing with critical thinking: The Paul-Elder frameworks. The observations made thus far suggest that it should be possible, in general, to align specific critical thinking skills with specific reading and writing skills. This hypothesis appears to be correct. The relationship can most readily be expounded by taking one popular model of critical thinking—the Paul-Elder model (Paul & Elder, 2005)—and showing how it lines up with the skills outlined in Table 1. While the Paul-Elder model is not the only model of critical thinking (Ennis, 1987; King & Kitchener, 1994; Kuhn, 1999), it is widely accepted and provides a useful standard of comparison since it was designed as an explication of critical thinking appropriate to support instruction.

The Paul-Elder model distinguishes several *elements of thought* and provides a list of several partially corresponding standards for evaluating the quality of thought. The elements of thought comprise the following (see Elder & Paul, 2007)

- **Purpose**—"all reasoning has a purpose." Effective critical thinking aims to accomplish clear, meaningful, and realistic purposes. The corresponding standard is *relevance* ("relating to the matter at hand").

- **Question at Issue**—"all reasoning is an attempt to figure something out, to settle some question, to solve some problem." Effective critical thinking identifies the question at issue, clarifies its meaning, and explores its ramifications. The corresponding standard is also relevance.

- **Point of View**—"all reasoning is done from some point of view." Effective critical thinking is aware of its own point of view, considers alternative points of view, and avoids egocentric and bias. The corresponding standard is *breadth* ("encompassing multiple viewpoints").

- **Assumptions**—"all reasoning is based on assumptions." Effective critical thinking is aware of its own assumptions, recognizes their consequences, and is willing to question them. The corresponding standard is *fairness* ("justifiable, not self-serving or one-sided")

- **Concepts**—"all reasoning is expressed through, and shaped by, concepts and ideas." Effective critical thinking defines its concepts fully. The relevant standards are *clarity* ("understandable, the meaning can be grasped")*, precision* ("exact to the necessary level of detail"), and *depth* ("containing complexities and multiple interrelationships")

- **Information**—"all reasoning is based on data, information, and evidence." Effective critical thinking bases its conclusions on accurate information that fully justifies the conclusions drawn. The corresponding standard is *accuracy*: whether the information is "free from errors or distortions; true."

- **Interpretation and Inference**—"all reasoning contains inferences or interpretations by which we draw conclusions and give meaning to data." Effective critical thinking is aware of the difference between inferences and direct evidence and is open to alternative interpretations. The relevant standard is *logic* ("the parts make sense together, no contradictions")

- **Implications and Consequences**—"all reasoning leads somewhere or has implications and consequences." Effective critical thinking explores and takes responsibility for the consequences of its own conclusions. The relevant standard is *significance* ("focusing on the important, not trivial").

These can be set in approximate parallel with elements in our own model, as shown in Table 2, though the two models are not identical. One difference worth noting in passing is that the Paul-Elder model does not distinguish between concepts and their expression; thus, three standards

14

Table 2

Mapping Between Skills Mentioned in Table 1 and the Paul-Elder Critical Thinking Model

Cognitive level	Specific skill categories	Paul-Elder model--elements of thought	Paul-Elder model—standards
Social	Intentionality	Purpose Question at issue	Relevance—relating to the matter at hand
Social	Perspective	Point of view	Breadth—encompassing multiple viewpoints
Social	Affect	Assumptions	Fairness—justifiable; not self-serving or one-sided
Conceptual	Exploration	Concepts	*(Clarity)* *(Precision)* *(Depth)*
Conceptual	Explication	Interpretation and inference	Logic—the parts make sense together, no contradictions
Conceptual	Modeling	Information	Accuracy—free from errors or distortions, true
Conceptual	Judgment	Implications and consequences	Significance—focusing on the most important, not trivial
Textual	Document structure	*[Expression of concepts]*	Depth—containing complexities and multiple interrelationships
Textual	Cohesion	*[Expression of concepts]*	Clarity—understandable; the meaning can be grasped
Textual	Development	*[Expression of concepts]*	Precision—exact to the necessary level of detail

apply to concepts, though they also map more or less transparently onto three distinct aspects of the textual level in our framework. However, the most important difference is that the Paul-Elder model does not draw a distinction between the social and conceptual elements of their model, a

difference that connects rather strongly to their emphasis on critical thought, rather than literacy construed more broadly.

This parallel display helps clarify the idea that reading, writing, and critical thinking are distinct activity systems founded upon common underlying skills. One can have critical thinking without reading or writing (for there is no requirement that reflective thought be expressed in written form). Writing can take place without deep reflection, for there is no guarantee that the thoughts expressed in a written text will be significant, relevant, fair, clear, precise, complex, accurate, or logical. Yet the whole point of skilled writing is to mobilize all of the resources available to the writer to achieve meaningful goals. The expert writer knows when to apply reflective thinking to writing tasks, just as the expert thinker knows when to use writing as a tool for reflection. The skills are not the same, but they mobilize similar underlying abilities.

These points can be elaborated a bit further by considering how Table 2 brings the Paul-Elder model into alignment with Figure 1 and Table 1. The parallels are not exact, but they are highly suggestive. Table 1 isolates three major elements that play a crucial role in social understandings of communication: intentionality, perspective, and affect. Table 2 illustrates how certain aspects of the Paul-Elder model are essentially parallel. Let us examine these aspects one piece at a time, starting with the social model, then proceeding to the conceptual and textual models.

Social aspects. The literacy model presented in this paper selects intentionality, perspective and affect as broad subject headings capturing some of the distinctive elements of socially-focused thought. The Paul-Elder model does not make the same distinction, so Table 2 identifies elements of that model that correspond to ours, not identifying the two models. It seems unexceptional to claim that the concepts of purpose, of the question at issue, and relevance are all related to intentionality, or to claim that point of view and breadth address issues of perspective. Table 2 sets up a parallel between the affective elements in our model and two Paul-Elder elements: assumptions and fairness. This is more questionable, since assumptions to a significant extent are related to point of view. It seems reasonable to place it parallel to affect, since the affective element of the social model includes commitments and stances toward ideas, which is what usually biases people not to notice their own assumptions or to treat the perspectives of others dismissively.

Conceptual aspects. Table 1 outlines four general types of activities (exploration, explication, modeling, and judgment) that constitute major types of conceptual thought. These general tyupes map onto much more specific types of activities, and families of strategies that go with them, which are outlined in more depth in an appendix at the end of this paper (Table 5). The parallel to elements of the Paul-Elder model is not exact, but it is informative. The Paul-Elder model distinguishes concepts, interpretation/inference, inference, and implications/consequences, and sets forth standards for intellectual quality focusing on clarity, precision, depth, logic, accuracy, and significance. Practically speaking, it is impossible to perform any sort of thinking activity without being concerned with all of these elements, but it is reasonable to postulate that

- *exploration* activities are first and foremost concerned with identifying, understanding, and/or explaining concepts clearly, precisely, and in depth;

- *explication* activities are first and foremost concerned with making explicit the inferences and interpretations necessary to understand a subject or a text, and thus in bringing out the underlying logic of the conceptual system being addressed;

- *modeling* activities are first and foremost concerned with providing an accurate model that captures all of the important facts about the subject being modeled; and

- *judgment* activities are first and foremost concerned with evaluating ideas in terms of their significance, implications, and consequences—though of course, evaluation implies critical attention to all aspects of conceptual structure.

These parallels highlight the presence of similar conceptual elements without necessarily organizing them in precisely the same way. In effect, the literacy model outlined in this paper identifies a range of activities in which conceptual thinking is prominent, while the Paul-Elder model seeks to identify aspects of conceptual thinking that help define its structure; the two organizations share important elements but are not by any means identical.

Textual aspects. The models in Figure 1 and Table 1 help highlight distinctions that are not so clear in the Paul-Elder model, and thus cannot clearly be explicated in Table 2. Do the Elder-Paul standards of clarity, precision, and depth represent standards for thought, or do they refer instead to the manner in which thoughts are verbally expressed? It is not entirely clear that this is a meaningful distinction, but at first blush it would seem that standards of clarity,

precision and depth apply much more directly to the textual presentation of a system of ideas than they are to unexpressed, purely mental conceptions that have not yet been put into a form that can be communicated to other people. It is hard to evaluate whether a system of ideas has depth unless the complexities and interrelationships it addresses have been laid out explicitly in textual form. An inextricable connection exists between precision of content and precision of phrasing, or between the clarity of thought and the ability to express it coherently. Table 2 expresses these parallels and connections by linking these standards both to the conceptual and to the textual models.

1.3. The Role of Reflective Strategies and Genres: Modeling Activity Systems in Instruction and Skill Development

Research on the acquisition of complex skills—including writing, reading, and critical thinking—emphasizes the importance of strategy instruction (Block & Parris, 2008; De La Paz & Graham, 2002; Graham & Harris, 2000; Graham & Harris, 2005; Graham, Harris, & Troia, 2000; Graham, MacArthur, Graham, & Fitzgerald, 2006; Pressley, 1990; Pressley, Harris, Alexander, & Winne, 2006; Souvignier & Mokhlesgerami, 2005; van Gelder, 2005; van Gelder, Bissett, & Gumming, 2004).

The typical path to mastery begins with explicit instruction in conscious strategies that support the learner in the early stages of skill acquisition. Over time, the new skill becomes routine and aspects of it are automatized, though the learner has the capacity to fall back on conscious strategies under conditions that stress or overwhelm automated capacities.

Given the arguments that this review has presented thus far, it would be reasonable to expect deep parallelisms among the kinds of strategies that support reading, writing, and critical thinking. An examination of the literature suggests that this is indeed the case.

Strategy families as modes of thought. An obvious connection is advanced between strategy instruction and the classification of educational objectives in Bloom (1956) and presented in revised form in Anderson and Krathwohl (Anderson et al., 2001). Strategies to support comprehension, composition, and critical thinking range from simple memory-based methods to complex forms of synthesis and evaluation. In terms of the high-level model in Table 1, such strategies are ways to *rethink* what one already knows by clarifying what one does not fully understand, synthesizing and hypothesizing new ideas, and criticizing old ones. These kinds of strategies tend to fall into a relatively small range of families. Space is not available here to

elaborate on these families, though Table 5 in the appendix presents a taxonomy of conceptual strategies that appear (often in slightly different guises) sometimes as reading strategies, sometimes as writing strategies, and sometimes as more general conceptual, critical thinking, or inquiry strategies. By way of illustration, two such families we be considered. A first example is a family of strategies that include freewriting (a writing strategy) and its close cousin, self-explanation (a reading strategy); a second example is outlining, which can be deployed strategically either as a tool to support planning (a writing strategy) or to improve global text comprehension (a reading strategy).

Freewriting vs. self-explanation. *Freewriting* is a common strategy recommended when writers are beginning to develop their ideas. The technique requires the writer to forget about strategic control and planning and just put words to the page, letting one idea lead to another, giving the writer every chance to express himself or herself without worrying (yet) how those ideas will fit into a rhetorical plan (Elbow, 1987). After freewriting has taken place, the text produced can be subjected to analysis, which may help the writer identify what is really significant and important, and to identify what really needs to be said (Elbow, 1994; Yi, 2007).

Self-explanation is a strategy recommended when readers need to deepen their understanding of a text. Readers write down what they understand the text to mean, worrying only about expressing their current understanding without worrying about how closely the self-explanation tracks all details of the text. Afterward, the reader can compare the original text to the self-explanation and perhaps discover aspects of the text that are not yet fully understood (Chi, 2000; Chi, Bassok, Lewis, Reimann, & Glaser, 1989; McNamara & Magliano, 2009). The parallelism between the two techniques is worth noting. Both involve the use of expressive skills to force a clarification of ideas and involve a temporary suppression of evaluation in order to facilitate the process. Under the proper circumstances, both techniques can enable reflection and thus support critical thinking.

Outlining for comprehension vs. outlining as text planning. *Outlining* is the use of a graphic organizer or other explicit hierarchical structure to represent how a document is organized. Creating an outline is often recommended as a strategy to support reading comprehension (Jiang & Grabe, 2007). While a skilled reader may be able to organize document content implicitly, without recourse to outlining, the reflective act of creating an outline forces readers to identify main ideas and supporting details specifically and requires them to encode

19

explicitly how different parts of the outline are related. A graphic organizer reduces the load on short-term memory by offloading some of the organizational effort into a visual encoding. Of course, outlining has the same advantages when recruited as a planning tool, which makes it one of the few planning strategies known to have a powerful positive effect on writing quality (Kellogg, 1988). Both forms of outlining instantiate a general class of strategies for reflective thought—the use of visual hierarchies to encode relevance and significance relations.

Genres of writing as purpose-driven activities. The general framework proposed in this review treats writing as being essentially purpose-driven. It is part of an activity system and is distinguished from other, closely related activities by its goal (producing a written text) and by the strategies it deploys to mobilize literacy skills to achieve that goal. Once writing is conceived of in this way, it extends logically to cover the concept of genre. A specific genre of writing is focused on achieving a particular type of goal. For instance, an argumentative essay is focused upon the goal of establishing the truth of a claim. Achieving this goal logically requires the writer of an argumentative essay to accomplish certain things, such as elaborating subclaims, providing supporting evidence, rebutting counterarguments, or exploring logical consequences. Some of the tasks that need to be accomplished will be similar from one genre to another, while others, such as those listed above for argumentation, form a constellation of tasks strongly linked to genre-specific goals. Genres typically adopt conventional patterns, including conventional patterns of organization and conventional stylistic features. If genres are viewed as conventionalized activities within a larger activity system, these conventions reflect strategies for solving genre-specific problems whose usefulness has led to repetition and ultimately to conventionalization.

There is nothing particularly surprising about any of the conclusions noted thus far—similar observations have been made by a variety of genre theorists (Bazerman, 2004; Russell, 1997)—but it does lead to an important conclusion for our purposes. It means that learning to write consists in large part of three things:

- Learning key strategies

- Learning how to assemble those strategies in meaningful ways to accomplish specific goals as part of purposeful activities

- Turning the resulting assemblies (i.e., complex activity plans) into routine, efficient procedures for handling ordinary problems

- A corollary is that writers are likely to be ill-served if they learn strategies piecemeal, without understanding how to connect them to meaningful purposes—and that they will be equally ill-served if they are taught narrow routines for achieving specific writing goals without ever learning how general-purpose strategies cohere with specific writing tasks in meaningful contexts.

Another way of making the same point is to consider how conceptual strategies map onto the genre categories that students need to have mastered by the time they reach college. Various studies of the kinds of writing required at the collegiate level have been conducted (Biber, 1980; Bruce, 2005; Gardner & Powell, 2006; Hale et al., 1996; Martin & Rose, 2006; Nesi & Gardner, 2006; Rosenfield, Courtney, & Fowles, 2004), as well as genre analyses of the types of reading and writing required in primary and secondary school (Kirsch & Jungeblut, 2002; Martin & Rose, 2006). If this information is collated to produce a reasonably complete list of genres that support academic work, and to determine which strategies are most central to each, it rapidly becomes clear that students need to master a wide range of conceptual strategies—and develop complex procedures supporting complex activities in a variety of genres—to achieve collegiate levels of performance. Historical analysis depends critically upon one kind of strategy—reconciling multiple sources—while literary analysis depends critically upon another—close reading. Scientific reports require a familiarity with hypothesis testing, while philosophical research is more strongly associated with definitional techniques going back to the Socratic method.

Obviously, students before college age will not need to perform at a more complex level, but sophistication in applying conceptual strategies does not emerge automatically; for instance, Kuhn's (1991) study of the development of argumentation skills demonstrated considerable range in skill even among adults. It thus follows that the effectiveness with which students will learn to write in a range of genres is critically dependent on their mastery of the conceptual strategies that will enable them to accomplish genre-specific purposes. Space does not permit a detailed explication of the range of genres that students need to acquire to perform well at a collegiate level (though see Table A2 and the associated discussion in the appendix for a

condensed presentation of associations between genres and conceptual strategies). But it is very clear that effective writers are able to handle a broad range of genres and, thus, that they must be able to mobilize many different varieties of strategic thought.

Developing skill in writing does, of course, involve developing discourse, verbal, and orthographic skills—but these considerations suggest that writing skill also depends upon strategy instruction for one very simple reason. Strategy instruction enables writers to selectively mobilize a wide range of social, rhetorical, and conceptual skills depending on their purpose in writing, and these skills are as necessary to high-level writing performance as general verbal fluency or generic understanding of document structure.

This view militates against any approach to writing instruction—or writing assessment—that treats writing as a skill to be taught or assessed in a vacuum, which would risk construct under-representation. For example, teaching students how to write a persuasive essay is unlikely to succeed unless students also develop critical reading and logical reasoning skills, and know how to deploy those skills in support of writing an essay. That additional development is likely to happen only if they also internalize all the elements of a community of practice in which argument and debate are normal activities, so that they acquire not only strategies but also a sense of their relevance, and internalize appropriate practices and norms.

1.4. Modeling Activity Systems: A Strategy for Assessment That Supports Learning

Having come this far in extending connections among cognition, literacy, and instruction, it is now possible to return to assessment—but with a much richer understanding of the construct to be assessed and a much clearer understanding of how assessment, as an activity, needs to be structured to reinforce the kinds of social learning that instruction should ideally support. As noted in the introduction, Bennett and Gitomer (2009) argued that educators should develop assessment systems that document what students have achieved, help identify how to plan instruction, and turn the testing situation into a worthwhile educational experience in and of itself. The analysis presented in this review suggests a very specific strategy for accomplishing these goals.

Expert writers can successfully pull together very complex performances that can ultimately be measured by the written product. But the final written product is in some sense the tip of the iceberg: It represents performance within a complex activity system and acquisition of procedures for producing texts in which many different skills have been coordinated

22

successfully. Less-skilled writers may lack critical skills—or they may have no idea what skills need to be mobilized or how they should be coordinated, and that fact means that far less information is to be obtained than one might wish from an analysis of the final written product.

Viewing the problem purely from an assessment point of view, therefore, it would be very helpful to find out whether writers have the skills they need to put the pieces of an activity system together, which means both mastery of a variety of specific procedures (in this case, genres) and mastery of appropriate procedural knowledge that will mobilize the skills they need to apply to accomplish their goals. Lacking that, there is a risk that the final written product will mask student difficulties due to compensatory relationships among skills. To take a fairly straightforward example, it is quite common on some writing examinations for students to memorize a *shell script*—a skeletal outline that contains all the elements that signify clear organization and effective transitions. Instead of developing an organic organization focused on the task, the student plugs reasonably relevant content into the shell. The resulting essay may provide much less useful information about the students' ability to construct arguments or to organize information than one might wish.

While this may be a relatively extreme example, the same point recurs. Given a complex task such as an argumentative essay, there are many construct-relevant skills about which the final essay provides less-than-direct evidence. Can students understand and summarize other people's arguments? Can they recognize useful evidence when they see it (much less use it consistently)? Do they understand the idea that arguments have to be supported and that the support may not work (or can be successfully attacked)? Given a high-quality argument, the answer to all these questions is an unqualified yes. But given an unsuccessful performance, the reason for the failure may be hard to determine.

It is, of course, true that everything tends to correlate with everything else—that is what you get when many different activities within the same activity system draw upon the same underlying pool of skills—but if test developers structure a test carefully, it should be possible to generate reasonable hypotheses about why particular students are falling short of ideal performance. For instructional purposes certainties are not required, only reasonable hypotheses that could help teachers focus their instructional goals, addressing such questions as:

- Whether students have the skills they need to apply appropriate strategies

- Whether their final performance demonstrates an ability to coordinate those strategies effectively

For example, an argumentative essay requires students to apply argument-building strategies. Some students may understand what an argument is yet fail to apply appropriate strategies. Others may function at a much more basic level. The difference matters a great deal for instructional purposes.

These considerations lead to the somewhat paradoxical suggestion that a writing test ought to test more than writing. Given a specific writing task, a specific set of activity systems can be identified that guide expert performance. These activity systems will include specific strategies applied by experts, and task sequences that model the kinds of things skilled writers do as they think about, plan, write, and revise that sort of text. Given that, it should be possible to identify reading, critical thinking, and smaller-scale writing tasks that measure the skills students need and that instantiate at least some of the strategies they ought to be applying. Moreover, there will be a bonus that attaches to tests with this type of structure: The test will actually model the kinds of strategies students need to use and will help to communicate how the writer's work fits within the larger activity system, which will make the test an educational experience in its own right. Perhaps it should not be called a bonus, since it is precisely what can make assessment fit organically into instruction rather than making is an alien mode of interaction superimposed upon a fundamentally different form of activity. As long as the purpose of each task is clear—as long as students can easily infer why each task has been included on the test and can see how that task helps them prepare for the final, integrated writing task—the test itself can become a meaningful experience and can be structured to model appropriate forms of strategy instruction.

Of course, the assessment strategy sketched in this review requires that each test should focus on a particular genre or category of writing. The strategies that support writing an argumentative essay will not be the same as those that support writing a research paper or a literary analysis. Not only will the strategies differ, they will need to be coordinated differently. This revelation is consistent with the vision advanced by Bennett and Gitomer (2009): One test is not enough, at least not if the purpose of the test is to represent something of the richness of writing tasks that students are expected to master. It may not be necessary to increase the number of tests vastly or to cover as wide a range of writing situations as might be covered in a portfolio

assessment. But the proposed test design, focused as it is on specific genres of writing, implies a richer array of assessments and a strategy that combines results across assessments to get a composite picture of writing skill.

In addition, the proposed test design is effectively a kind of scaffolding where the structure of the test partially guides students through the thinking they need to accomplish. This kind of design makes the most sense for the age ranges at which a writing task has been introduced but not yet mastered—helping to address students within the Vygotskian zone of proximal development (Vygotsky, 1978). That is, with a population consisting primarily of students who may have been introduced to the task but have not yet reduced it to a routine performance, a scaffolded assessment structure yields more information about partial learning while focusing instruction on making sure that students are able to apply the right strategies to the task. When a writing task has become routine, it is reasonable not to scaffold it, and scaffolding might interfere with the skills one wishes to measure. Thus it can be anticipated that at one grade, a task such as summarization might be the focus of an entire test, with a full array of lead-in tasks modeling appropriate summarization strategies. Then at a later grade level, summarization might be treated as a basic task and function as part of a supporting strategy for more complex forms of writing.

In effect, a concept of writing assessment is being proposed that involves the creation of a sequential family of assessments, with earlier assessments (appropriate for earlier grade levels) focused on simpler writing tasks and with later writing tasks incorporating earlier, simpler forms of writing as part of the scaffolding leading up to a more complex integrated task. The task of constructing such a sequence of assessments corresponds, in effect, to building a pedagogical sequence based upon empirical studies in which some genres are introduced before others and incorporated at higher grade levels as component activities in more complex forms of writing.

The task of constructing such a sequence presupposes a detailed analysis of the activity systems underlying literate discourse. As such, it entails an analysis of the ways in which different genres relate to one another and form meaningful patterns of activity. The current article cannot undertake such ananalysis in depth, but considerable prior literature focuses on this kind of issue and illustrates the kind of analysis from which this paper has drawn. Of particular note is work on specific academic communities of practice such as literature, science, history, and philosophy (Geisler, 1994; Graves, 1991, 1996; Hunt, 1996; Norris & Phillips, 1994, 2002;

Norris, Phillips, & Korpan, 2003; Rouet, Favart, Britt, & Perfetti, 1997; Vipond & Hunt, 1984, 1987; Vipond, Hunt, Jewitt, & Reither, 1990; Voss, Greene, Post, & Penner, 1983; Voss & Wiley, 2006; Wineburg, 1991a, 1991b, 1994, 1998; Zeits, 1994)

But the family of assessments envisaged here would go beyond genre analysis, because each genre would be placed in a well-designed pedagogical context. Each assessment would model the kinds of strategies critical to a particular genre, while the sequence of genres would carry students systematically toward more complex, more demanding tasks that depend on every-increasing sophistication in the use of task-appropriate strategies. Space precludes a detailed discussion of what such a sequence might look like (though see Table A3 in the appendix for an attempt to map out some rough estimates of when particular genres might usefully be taught and assessed). But the strategy at least is clear: At each grade level, the tests should be focused on forms of writing that depend on strategies students can reasonably be taught at that age. Since more variation is found within grades than between grades, one of the purposes of such an assessment would be to identify students who were in need of instruction at earlier and later stages of the sequence, while scaffolding learning for those students who were in the zone of proximal development.

The sections that follow will sketch out preliminary work on creating an assessment system in line with this vision. In particular, Section 3 will present a design focused on writing tasks appropriate for 8th and subsequent grades, and Section 4 will discuss some of the scoring issues that arise from these designs.

2. A Pilot 8th Grade Design

2.1. General Considerations

At this point the discussion must shift from a generic consideration of writing skill and focus instead on issues of test design. The list of skills in Table 1 can be understood as constituting a *competency model*—a specification of the skills needed to achieve the highest levels of skill as a writer—as long as it is understood that strong interconnections and interdependencies are present among the skills so the different competencies are not viewed as independent components but as strands within a larger, ultimately integrated set of skills. The general path of development appears to involve relatively early progress with the verbal and orthographic aspects of writing, transitioning to an emphasis on discourse and document

26

structure in the middle grades, with conceptual and social aspects of writing playing an ever more important role in middle and later grades (Applebee, 2000; Britton, Burgess, Martin, McLeod, & Rose, 1975; Langer, 1992; Langer & Applebee, 1986), though the picture is complex and varied when variations in social background, pedagogy, and genres of writing are taken into account.

The work to be reported here has focused upon 8th grade for several reasons. Eighth grade is one of the earliest grades at which students are expected to produce developed essays and other texts with complex internal structure. It is also the age at which persuasive writing, research, and exposition first come into focus—academic genres that require very different skills than the narrative-focused writing so common in the primary grades in the United States.(Duke, 2000, 2004). Eighth grade is thus an appropriate grade at which to examine the usefulness of scaffolded test structures focused on rhetorical purpose and critical thinking, while allowing the skills that underlie general writing fluency (e.g., verbal and orthographic skills) to be assessed without scaffolding.

2.2. Current Status

All of the 8th grade tests were developed in collaboration with the Portland, Maine, school district, which has three middle schools reflecting a mix of urban, suburban, and rural students, including English language learners, since Portland is a refugee resettlement site. The designs presented below represent several years of development. Test designs were thoroughly reviewed by Portland school district teachers and administrators and were revised and reworked multiple times in consultation with them.

Initial pilots were administered between 2007 and 2009 in Portland with relatively small numbers of students participating (between 125 and 200 per administration). Between October and December of 2009, the four test designs described later in this paper were administered in a large national sample.[1] Twenty four schools participated, representing a mix of urban, rural, and suburban districts from 12 states throughout the country (Alabama, Arizona Arkansas, California, Florida, Georgia, Kentucky, Louisiana, Massachusetts, Mississippi, Ohio, and Texas). A total of 2,564 eighth grade students participated. Each student was randomly assigned two different writing tests in a counter-balanced design; 1,978 students completed all sections of two tests; each of the four tests was therefore completed by more than 1,000 students. Answers were collected for all questions; background data was also collected, including No Child Left Behind

(NCLB) test scores and demographic data, and keystroke logs (records of the time course of student responses to the essay tasks). ETS has recently completed scoring these tests and has begun in-depth analysis, which will include psychometric studies appropriate for large-scale pilots, examining item functionality, dimensionality, equating across forms, and related issues. Forthcoming studies will also examine the extent to which automated scoring and automated collection of timing data can be used to extract instructionally useful information.

Since these analyses are not yet complete, the focus in this paper will be on the test design itself and will explicate the design decisions that underlie it.

2.3. Test Design

The following specification underlies the designs to be presented and helps make clear how that design maps onto the kinds of skills specified in Figure 1 and Table 1.

Individual forms. Each test form is administered on the computer, requires approximately 90 minutes, and has the following characteristics:

1. Embodies a realistic scenario in which a series of related tasks unfold within an appropriate social context. The scenario is clearly established at the beginning of the test form to give students a sense of what they will need to do and why. It thus connects to the social elements in Figure 1: *engage, empathize, collaborate*.

2. Contains a sustained writing task (30–45 minutes) that strongly exercises the ability to use critical thinking skills for writing, plan, and structure documents, to use formal written English, and to follow written conventions (thus exercising the expression elements in Figure 1: *engage, inquire, structure, phrase, inscribe*. This task may require students to write an essay, memorandum, letter, proposal, newspaper article, or other document form that they may encounter outside of school. The specific form will be determined by the scenario. The writing needs to be formal enough, and directed to a mature enough audience, so as to require written rather than oral vocabulary and style. These documents will be scored for the following:

   - Rhetorical effectiveness and conceptual quality (e.g., for success at *engaging* the rhetorical task and *inquiring* into the subject addressed).

- General quality of the document produced in terms of structure, phrasing, and language (e.g., for success at *structuring* the document, *phrasing* its ideas, and *inscribing* those ideas in standard written English).

3. Contains a series of lead-in and/or follow-up tasks, each relatively short (5–20 minutes), that require the student to think about the content to be addressed and to engage fruitfully with the overall critical-thinking and rhetorical requirements implied in the scenario (and thus involving elements of Figure 1 that cannot easily be addressed in a long, integrated writing task).

These tasks should also satisfy the following general criteria:

1. They introduce enough information, through reading materials or other sources, to enable students to write meaningfully about the subject (and thus may exercise the kinds of *interpretation* and *reflection* processes laid out in Figure 1).

2. They require students to demonstrate critical thinking skills that are necessary to perform well in the scenario modeled by the test (*inquire, infer, rethink*).

3. They are either short writing or selected-response tasks that most students can reasonably be expected to have mastered by the target grade, but are prerequisite to successful performance on the longer writing tasks.

4. Taken as a set, these tasks provide enough psychometric information to judge whether students have control of specific prerequisite reading, writing, and critical thinking skills needed to address the larger-scale writing task. For instance, if the final writing task focuses on building arguments, the lead-in tasks should do so also. Ideally they would address aspects of prewriting or revision that cannot easily be measured in the final written product.

5. Taken as a set, these tasks scaffold, and thus help model, what it means to perform well on the overall scenario and represent important stages of the thinking-and-writing process needed for successful performance. Ideally, the scenario should represent a longer writing task that would be difficult for many students at grade level to achieve without help but which most can achieve if guided through the process step by step with appropriate scaffolding.

6. The shorter tasks should contrast with the longer-writing task in important ways, exercising parts of the competency model not easily measured by an essay task alone, in particular:

- At least one task should be a critical reading task without a written response, to help disentangle the ability to reason critically about content from general writing and drafting skills, by measuring prerequisite *interpretation* skills.

- Where practical, at least one task should require students to demonstrate the ability to assess and modify documents (*revise, edit, proofread*).

- At least one task should allow students to write in a less formal style, addressing peers or younger students rather than elders, allowing them to demonstrate the ability to switch between a more formal and a more oral style, and more generally, giving them an ability to adapt what they write to purpose and audience.

7. They present grade-appropriate texts for students to read and think about. The purpose of these texts is not to assess reading skills but to give students content to consider (e.g., to summarize, to analyze, to synthesize, to evaluate) in preparation for writing, thus modeling the kinds of activity systems that the genre actually belongs to. The texts may be informative, persuasive, literary, research-based, or a part of any other genre relevant to the scenario and purpose for writing. The length of the texts must not exceed reasonable reading-time expectations for the target grade.

8. They support thinking and writing activities with resources such as guidelines, writers' checklists, evaluation criteria, tips for getting started, or other reference materials to help students as they progress through the composing process, thus helping to make the test experience more of an educational experience in its own right.

Each year's sequence of periodic accountability assessments. The sequence of assessments administered during any given year and grade level will be selected to exercise a broad variety of critical reasoning skills set within an equally broad array of rhetorical situations.

The focus and content of each such assessment (periodic accountability assessment) will be driven by critical thinking and rhetorical requirements, and not by surface form, in particular:

1. Each periodic accountability assessment will require the student to demonstrate control of a different type of critical thinking.

2. Each assessment will require students to demonstrate the ability to write in a particular genre or form for which that type of critical thinking is essential and has been targeted for instruction at that grade level.

3. The distribution of critical thinking skills across forms will reflect reasonable grade-level expectations about the type and range of critical thinking skills that students will be expected to demonstrate.

4. Each periodic accountability assessment should be self-contained. The order in which forms are administered should not matter, in order that test sequences can be adjusted to match curricular requirements.

Four periodic accountability assessments. Table 3 presents key conceptual features of four assessments developed to model key characteristics of different sorts of writing that students should be learning in 8th grade. None is a genre that 8th grade writers can be expected to have mastered, making a scaffolded structure appropriate. Table 3 specifies the kinds of critical thinking involved, the critical thinking strategies these specific tasks help to develop, the genre of the major writing task, and the kinds of reading materials included as part of the scaffolding for the longer, culminating writing task.

While space does not permit explication here, formative and teacher-support materials have also been developed, in two forms: parallel scenarios (with a richer array of tasks than could be included in the tests) and relatively independent formative assessment tasks designed to support skills that students need to master before they undertake the integrated writing tasks built into each assessment, such as summarization and thesis sentences. ETS is continuing to work with educators to build a model that is closely linked to grade-appropriate standards and which provides models of appropriate instructional practice.

2.4. Walkthrough of a Sample Test Design

At this point is will be useful to consider one test design in order to clarify the transition from theory to practice. What follows is a short tour through the final test design given in Table 3, which focuses on explication of a literary text. Figure 2 presents an early screen from this text, which explains the scenario.

The timings shown on this screen are provisional. ETS is also experimenting with longer, untimed administrations, but these are the timings built into the current pilot, which was administered in Fall 2009, and whose results are currently being analyzed. As this outline indicates, the full-scale writing task is last, with preliminary tasks supporting student understanding, while simultaneously measuring how well students perform on simpler versions of skills they must call upon to succeed at the integrated writing task.

The House on Mango Street

Your class is reading selections from *The House on Mango Street*. The author, Sandra Cisneros, draws upon her childhood experiences and Hispanic heritage to create characters who are often isolated from mainstream American culture.

In this book, the narrator is a girl named Esperanza, who is twelve years old when the story begins.

For this project, you will perform four tasks.

Task 1:	Support interpretations of the story with details from the text	15 minutes
Task 2:	Explain whether a character's attitude changes	15 minutes
Task 3:	Help another student interpret the text	15 minutes
Task 4:	Write an essay about the character	45 minutes

Figure 2. **Overview screen for a test focused on literary analysis.**

Note. CBAL = Cognitively Based Assessment of, for, and as Learning.

Table 3

Design for Four 8th Grade Writing Assessments

Genre	Key strategies	Skills in focus
Recommendation	Defining Appeal-building	Collaborate + infer (explication) Judge how well a persuasive letter meets a rubric Infer (explication) Judge how well proposed activities meet evaluation criteria Rethink (explication) Analyze how well alternative proposals satisfy evaluation criteria Inquire (judgment) + structure Recommend one alternative, and justify that choice in the form of a letter or memorandum
Report	Guiding questions Concept mapping Reconciliation	Infer (judgment) Evaluate sources of information Rethink (exploration) Formulate guiding questions Infer (exploration) + integrate Organize information in terms of guiding questions Inquire (exploration) + structure Explain this information using an appropriate set of major points or bullets in pamphlet form
Essay	Outlining Argument-building	Infer (judgment) Assess how well a student text meets standards for summarization Integrate + inquire (explicate) Summarize arguments on an issue Infer (judgment) Classify arguments as pro or con; assess whether evidence strengthens or weakens an argument Collaborate + rethink (judgment): Critique an argument containing errors in argumentation Inquire (judgment) + structure Justify a position on an issue in essay form
Interpretive review	Simulation/ roleplaying Close reading	Empathize + infer Make inferences about character intentions, perspectives & attitudes from details in the text Collaborate + rethink (modeling) Clarify inferences about the text in response to other attempts at interpretation Infer (modeling) Clarify difficult points in a text in light of global inferences and explanations Inquire (modeling) + structure Explain and justify an interpretation of a text in essay form

The screen shown in Figure 3 illustrates one item from the first set of tasks students are assigned, which could be viewed as a reading task but is part of the class of procedures students need to have mastered in order to justify an interpretation in written form. The test contains five

items of this type. The reasoning is that if students cannot identify specific places in a text that provide evidence to support an interpretation, they are very unlikely to be able to produce a written text that depends upon being able to accomplish the same task in verbal form, which adds all the complexities of text production to the basic analytical procedure. Several interpretive questions of this form are presented so as to be able to form a rough estimate of whether students are capable of performing this task in isolation.

Figure 3. **Interpretive questions: identifying textual support.**
Note. CBAL = Cognitively Based Assessment of, for, and as Learning..

Figure 4 shows the next question. The question simulates a blog-based classroom discussion comparing two selections from the source text, using previous student comments to identify an interpretive issue and focus those issues to encourage an appropriate student response. One of the key elements being assessed here is whether students will be able to focus on the interpretive issue and on identifying support for it. Both selections are available to students while they write this response. The choice of task is designed to create a situation in which students are allowed to use a voice comparable to what they might use in a class setting

addressing peers. The task is primarily scored for content. While students are told to use standard English, they will not be penalized for informal features in their response.

CBAL Writing - The House on Mango Street | Question # 1 of 1 | Timer 15 minutes | Back Next

Blog | House | Alicia and I

1.
Esperanza doesn't want to stay on Mango Street, but at least it's better than where she lived before. She was really embarrassed about living on Loomis Street, so she must think things are getting better.

--Sara

2.
I don't know. Esperanza's parents might look on the bright side. Maybe Alicia does, too. But does Esperanza? From the details in the story, like the way she talks to Alicia, I'm not so sure that she feels good about living on Mango Street.

-- Victor

Directions: The teacher told the class: "Write your comments comparing the two chapters. Your comments should focus on Esperanza's feelings about the house on Mango Street and build on the ideas that other people have posted. Don't just say that you agree or disagree with them, but add your own thoughts. As always, use standard English."

Type your comments here.

Figure 4. **Developing an interpretation: short response.**

Note. CBAL = Cognitively Based Assessment of, for, and as Learning..

The final set of preparatory tasks focus on a third selection from the source, one that presents some interpretive difficulties. In the initial screen shown in Figure 5, text is highlighted and interpretive questions are inserted in the margins. The questions partially help to scaffold students' understanding of the text (by explaining elements that might be too difficult for most students and by highlighting issues for them to think about). These questions are re-presented one at a time after this introductory screen, as shown in Figure 6. They are presented in multiple-choice form, but the difference among choices has to do with the quality of the explanation provided for an answer rather than with the answer itself. Once again, several such questions are presented so that there will be enough information to make a rough estimate whether students are able to handle this kind of analytic task. Note that this type of question is (quite intentionally) rather more difficult than the initial task where students only had to identify textual support for a predefined interpretation.

CBAL Writing - The House on Mango Street
Task 3

Question #
0 of 6

Timer
15 minutes

Back   Next

A student wants your help interpreting parts of the story. Read his questions in the boxes.

Who (or what) is the ghost? How does writing make it not ache so much?

If "Mango" is the street, how can it say goodbye?

What does this part of the story show about Esperanza?

Mango Says Goodbye Sometimes

I like to tell stories. I tell them inside my head. I tell them after the mailman says, Here's your mail.

Once again, text has been deleted to avoid copyright infringement

I am going to tell you a story about a girl who didn't want to belong.

I put it down on paper and then the ghost does not ache so much.   I write it down and Mango says goodbye sometimes. She does not hold me with both arms. She sets me free.

One day I will say goodbye to Mango. I am too strong for her to keep me here forever. One day I will go away.

I have gone away to come back. For the ones I left behind. For the ones who cannot out.

I'm not sure about this part. The girl in the story is Esperanza, but is she also telling the story?

OK, so "who cannot out" means "the people who can't get away from Mango Street." But what does it mean when Esperanza says that she has gone away so that she can come back for them?

Figure 5. **Preparatory screen for the third selection from the source.**
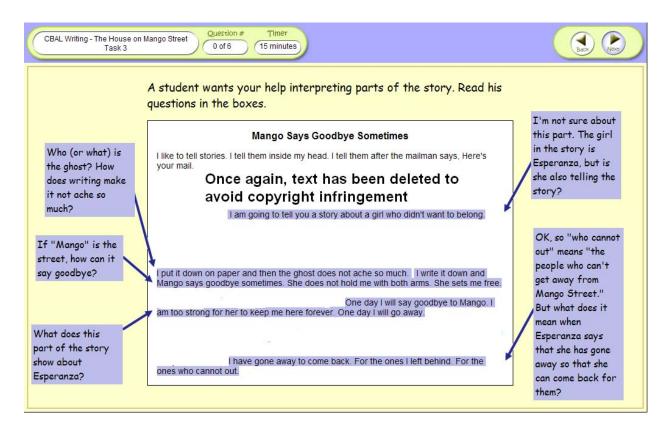
Note. CBAL = Cognitively Based Assessment of, for, and as Learning.

The final task is to write an essay addressing the development of the protagonist's feelings over the three selections. Evaluation of the essay focuses on whether a reasonable interpretation is presented and justified effectively using evidence from the text. The essay prompt is straightforward, as shown in Figure 7. All three selections are available to the student, and the final version of these tests includes various tools to assist the writer, such as planning tools, the use of which is not assessed.

The key point to note about this design is that it varies from a standard writing test by including a wide range of preparatory planning tasks. These tasks could variously be interpreted as reading tasks, critical thinking tasks, or short writing tasks—but in each case, the lead-in tasks help students prepare for the final full-scale writing task, test whether they have competencies necessary to successful performance of that particular task, and firmly embed the entire test into a particular activity system and a well-defined community of practice. In effect, the structure of

Directions: Read the highlighted text in the story and the student's question in the box below. Then choose the best answer to the student's question.

Student's question:

Who (or what) is the ghost? How does writing make it not ache so much?

○ The ghost is Esperanza's ancestor. It is trying to tell her what to write about, but she refuses to listen.

○ The ghost is an image of success. Esperanza dreams of becoming famous, but she knows that it will be hard to reach her goal.

○ The ghost symbolizes everything good and bad about books. Esperanza realizes that books are often hard to write but rewarding to read.

○ The ghost represents Esperanza's feelings about the place where she lives. Writing about it can sometimes ease her pain.

Figure 6. **Questions requiring selection of plausible explanations.**

the test—and the fact that there are multiple such tests, each examining a different genre of writing—helps to define writing as a richer construct than would otherwise be the case.

3. Issues Connected With Scoring

3.1. General Strategy

At this point it will be useful to take a step back from the details of the design and consider what information educators might wish to obtain from a writing test and how the testing approach being advocated can be used to serve educational needs. These concerns dovetail, in turn, with recent trends toward the use of automated scoring methods in writing assessment and with concerns that have been raised about their use. It is therefore incumbent upon us to consider how tests will be scored if they are designed along the lines presented above and to explore how

that can be done efficiently, providing full support to the rich construct they are intended to test while providing as much useful information to educators as possible.

Directions: Write an essay explaining how Esperanza's feelings about Mango Street develop over the three chapters.

Type your essay here.

Figure 7. **The literary explication prompt.**

The outline provided in Table 1 sets forth a comprehensive list of verbal skills that may be called upon to a greater or lesser extent in different writing tasks. It is obvious that some of these skills are more centrally writing tasks than others. In particular, the points in label 1 entitled engage, inquire, structure, phrase, and inscribe are central writing skills almost by definition, since together they comprise the ability to create an effective rhetorical plan, deal accurately with the subject matter to be addressed, and compose a well-structured, clear, and appropriately phrased document.

In scoring effort for the national pilot described earlier in this paper, a strategic decision was made to separate scoring for the "engage" and "inquire" competencies from from scoring for the "structure", "phrase", and "inscribe" competencies. The former are intimately connected to rhetorical purpose and strategic thinking, whereas the latter are intimately connected to the development of fluency in text production. It is thus possible to make a fairly clean separation between the two aspects of writing. The rhetorical and strategic aspects of writing cannot be separated from genre in any meaningful way. By contrast, the ability to produce a well-structured text, while connected to genre, can be assessed in ways that are far more comparable from one writing situation to the next. The simplest way to illustrate this strategy will be to consider two candidate rubrics, both developed for the persuasive essay test form. Table 4 presents a draft scoring guide focused on rhetorical argument building; Table 5, a draft scoring guide focused on fluent, accurate, well-structured text production.

It would be reasonable to expect, based both on theoretical grounds and upon initial analyses of our early pilots, that scores based on rhetorical success and scores based upon text structure will be closely linked. In cognitive models of writing, a tradeoff occurs where fluency of text production processes frees up cognitive resources for strategic planning and reflective evaluation. Thus, from the fundamental perspective presented in this study, it is very useful to provide a dual score, since that will encourage instruction that recognizes the importance of developing fluent text production while teaching appropriate writing and thinking strategies. A significant implication of this strategy is that it will involve development of quite distinct rhetorical evaluations for each genre. The centrality of genre to our assessments cannot be overemphasized, even though it is also important gain information on the more generic skill categories presented in Figure 1. It may be particularly instructionally useful for teachers to be able to identify students who are not following the usual trend where fluency and strategic thinking develop in close synchronization. These may reflect special cases, such as students with high verbal abilities in another language or students who need to be challenged to go beyond fluency to engage writing at a deeper level, although specific studies of these issues using pilot data are still underway. Rubrics for rhetorical success have been developed for each of the four genres in the 8th grade design, and their effectiveness and correlations with one another, with human scoring for text structure, and with automated scoring will be detailed in forthcoming studies.

Table 4

A Rhetorical Scoring Guide Focused on Argument-Building Strategies

Level	Scoring criteria
Exemplary (5)	An EXEMPLARY response meets <u>all</u> of the requirements for a score of 4 <u>and</u> <u>distinguishes itself</u> with such qualities as i*nsightful analysis—for example, recognizing the limits of an argument, identifying possible assumptions and implications of a particular position; intelligent use of claims and evidence to develop a strong argument—for example, including particularly well-chosen examples or a careful rebuttal of opposing points of view; or skillful use of rhetorical devices, phrasing, voice and tone to engage the reader and thus make the argument more persuasive or compelling.*
Clearly competent (4)	The response demonstrates a competent grasp of argument construction and the rhetorical demands of the task, by displaying all or most of the following characteristics: <u>Command of argument structure</u> *States a clear position on the issue* *Uses claims and evidence to build a case in support of that position* *May also consider and address obvious counterarguments* <u>Quality and development of argument</u> *Makes reasonable claims about the issue* *Supports claims by citing and explaining relevant reasons and/or examples* *Is generally accurate in its use of evidence* <u>Awareness of audience</u> *Focuses primarily on content that is appropriate for the target audience* *Expresses ideas in a tone that is appropriate for the audience and purpose for writing*
Developing high (3)	While a response in this category displays considerable competence, it <u>differs from</u> Clearly Competent responses in at least one important way, such as a *vague claim; somewhat unclear or undeveloped arguments; limited or occasionally inaccurate use of evidence; simplistic treatment of the issue; arguments not well suited to the audience; or an occasionally inappropriate tone.*
Developing low (2)	A response in this category <u>differs from</u> Developing High responses because it displays problems that seriously undermine the writer's argument, such as *a confusing claim, irrelevant or self-defeating evidence, an emphasis on opinions or unsupported generalizations rather than reasons and examples, or an inappropriate tone throughout much of the response.*
Minimal (1)	A response in this category <u>differs from</u> Developing Low responses in that it displays little or no ability to construct an argument. For example, there may be *no claim, no relevant reasons and examples, or little logical coherence throughout the response.*

3.2. Automated Scoring Technologies and Fluency

Table 5 focuses on aspects of text quality that reflect text production skills where fluency is a paramount consideration. In terms of Figure 1, it involves the ability to *structure* a

Table 5

A Scoring Guide Focused on the Ability to Produced Well-Structured Texts

Level	Scoring criteria
Exemplary (5)	An EXEMPLARY response meets <u>all</u> of the requirements for a score of 4 but distinguishes itself by skillful control of language and sentence structure and a well–thought out and effective organization, which work together to control the flow of ideas and enhance ease of comprehension

Clearly competent (4)	The response displays all or most of the of the following characteristics:
	♦ It is well structured.
	   That is, clusters of related ideas are grouped in separate paragraphs, the sequence of paragraphs follows an appropriate organizing principle, and transitions between discourse segments are easy to bridge or else are signaled by the use of transitional phrases and discourse connectives so that it is easy to recover the global structure of the text.
	♦ It is coherent.
	   That is, new ideas are introduced with appropriate preparation, so as not to confuse the reader and connections between ideas are obvious or else indicated explicitly, so that the sequence of sentences leads naturally from one idea to the next, without disorienting gaps or leaps or hard-to-follow shifts in focus.
	♦ It is well phrased.
	   In particular, ideas are expressed clearly and concisely; words are well chosen and demonstrate command of an adequate range of vocabulary; sentences are varied appropriately in length and structure to control focus and emphasis.
	♦ It is well formed.
	   In particular, grammar and usage consistently follow the patterns of Standard English; spelling, punctuation, and other orthographic elements follow standard written English conventions; the register is appropriate for the genre and avoids inappropriately oral, colloquial, or casual usage.

Developing high (3)	While a response in this category displays some competence, it <u>differs from</u> Clearly Competent responses in at least one important way, including inconsistencies in organization, occasional tangents, lack of explicit transitions, failure to break paragraphs appropriately, wordiness, occasionally confusing turns of phrase, little sentence variety, lapses into an inappropriate register, or several distracting errors in achieving standard English grammar, spelling, or punctuation.
Developing low (2)	A response in this category <u>differs from</u> Developing High responses because its displays problems that seriously interfere with meaning, such as disjointed or list-like organization, paragraphs that proceed in an additive or associative way without a clearly focused topic, lapses in cross-sentence coherence, unclear phrasing, excessively simple and repetitive sentence patterns, inaccurate word choices, an inappropriate and distracting choice of register, or errors in achieving standard English grammar, spelling, and punctuation that sometimes interfere with meaning.
Minimal (1)	A response in this category <u>differs from</u> Developing Low responses because of serious failures in control of document structure, phrasing, or standard written form, such as lack of multiple-paragraph structure, general incoherence, vague, confusing and often incomprehensible phrasing, or a written form that consistently fails to follow the conventions of standard English grammar, spelling, and punctuation.

document, *phrase* its content, and *inscribe* it following the conventions for written text. These processes have direct effects on the form of the text, which can therefore be measured both by humans and somewhat less directly using automated, natural language processing features.

It is therefore important to consider the connection between our writing assessment design and automated essay scoring systems, since such systems appear to provide fairly direct measurement of the fluency- and accuracy-focused construct outlined in Table 5. For instance, ETS has an automated essay scoring technology, *e-rater*®, that predicts human holistic scores on the basis of features calculated using natural language processing technologies (Attali & Burstein, 2006; Burstein, Chodorow, & Leacock, 2004; Burstein & Shermis, 2003; Chodorow & Burstein, 2004). This scoring method makes use of the following classes of features:

- Features measuring accuracy (adherence to convention) in the areas of grammar, usage, mechanics, and style

- Features measuring vocabulary level and (where appropriate) topic-specific vocabulary choices

- Features measuring the presence of discourse coherence and discourse structure

Publications on other automated scoring technologies suggest that similar constructs are being measured (Landauer, Laham, Foltz, Shermis, & Burstein, 2003; Page, 2003; Shermis, Burstein, & Bliss, 2004).[2]

It is not our purpose here to consider the case for or against automated scoring. Automated essay scoring systems often correlate about as well as human holistic scores as human holistic scores correlate with one another (Deane, 2006; Dikli, 2006). In addition, writing trait scores tend to correlate strongly with one another, reflecting a general tendency for all aspects of writing quality to advance together (cf., Diederich, French, & Sydell (1961), discussed in Elliot (Elliot, 2005, pp. 155–158); Huot (1990) or Weigle, Bachman, & Anderson (Weigle, Bachman, & Alderson, 2002, pp. 108–115). Thus it is possible that use of automated scoring for fluency-related constructs could free human scorers to focus on rhetorical success, conceptual content, and other features that cannot be measured well by machines, along the lines of the scoring guide presented in Table 4.

This possibility would be of particular interest if it could be shown that automated methods could be used for more narrowly defined purposes, such as identifying students

potentially at risk due to weak text production skills. It is thus important to note that advances in computer text processing also make it possible to collect data about the *process* of writing, not just the *product*. Research on writing processes has long suggested that skilled writers show very different patterns than novice writers and that their use of time in particular reflects fundamental differences in the strategies they use to address writing tasks (Chenoweth & Hayes, 2001; Flower & Hayes, 1981; Matsuhashi, 1987). Computer technology now makes it possible to collect detailed keystroke logs that capture every step in the composition and revision process and identify significant pauses, such as pauses within or between words and those at major breaks such as sentence or paragraph boundaries, and editing events such as cut-and-paste or backspacing.

Moreover, there is strong reason to believe that automated measurement of process features could provide direct evidence about important aspects of writing not currently captured in automated text analysis systems (Lindgren, 2005). In preliminary analyses of keystroke logs collected in small-scale initial pilots, patterns have been identified that suggest such connections; for instance, longer pauses within words appear to be connected to lower-performing writers, possibly due to inefficiencies in their text production skills, while certain editing behaviors are more characteristic of writers producing more highly valued texts. Keystroke logs have been collected for every essay produced in large pilots currently being administered, scored, and/or analyzed, so in future studies it should be possible to examine how well keystroke logs and other automated features can be used to identify patterns of performance that will support instructionally useful hypotheses about student performance.

4. Conclusions

This paper represents ideas that are still actively being researched. While the Cognitively Based Assessment of, for, and as Learning (CBAL) model is likely to have an impact on current ETS assessment development work, the goal is longer-term, focused on developing a coherent framework for the assessment of literacy skills, viewed broadly as skills that support reading, writing, and associated thought processes. This study is intended to explore the implications of cognitive research for writing assessment, particularly implications about how reading, writing, and thinking skills are interleaved. But the model developed in this study is also important because it motivates innovations in test design that bring assessment more closely in line with best classroom practices.

In particular, an approach is developed that has several key features reflecting the insight that writing is a socially driven skill that requires the integration of a wide range of specific capabilities. Our approach includes the following:

- Orients test design toward a sophisticated cultural theory of language and communication in which writing genres are social and rhetorical constructs

- Focuses on designing assessments that will help students internalize appropriate norms for each written genre

- Grounds writing assessment in an explicit cognitive framework that clearly delineates the array of skills drawn upon by expert writers

- Employs a scaffolded, scenario-based structure designed to link genres with writing and thinking strategies

- Measures both prerequisite skills and integrated writing performances

- Presupposes that writing assessment needs to take place periodically, over the course of the school year, in ways that will integrate with and support learning and instruction

A primary goal of the CBAL initiative is to create assessments that are learning experiences in their own right. This goal has driven much of the design work reported in this paper, and if successful, may lead to the creation of writing assessments that more strongly support and enhance instruction.

References

Alverman, D. E. (2002). Effective literacy instruction for adolescents. *Journal of Literacy Research, 34*, 189–208.

Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, J. R., . . . Wittrock, W. C. (Eds.). (2001). *A taxonomy for learning, teaching and assessing: A revision of Bloom's taxonomy of educational objectives*. New York, NY: Addison Wesley Longman.

Applebee, A. N. (1984). Writing and reasoning. *Review of Educational Research, 54*(4), 577.

Applebee, A. N. (2000). Alternative models of writing development. In R. Indrisano & J. R. Squire (Eds.), *Perspectives on writing research, theory, and practice* (pp. 90–110). Newark, DE: International Reading Association.

Attali, Y., & Burstein, J. (2006). Automated essay scoring with E-rater V. 2.0. *The Journal of Technology, Learning, and Assessment, 4*(3), 13–18.

Barab, S. A., & Duffy, T. (1998). *From practice fields to communities of practice*. Bloomington, IN: Center for Research on Learning and Technology, Indiana University.

Barton, D., & Hamilton, M. (1998). *Local literacies: Reading and writing in one community*. New York, NY: Routledge.

Barton, D., Hamilton, M., & Ivanic, R. (2000). *Situated literacies: Reading and writing in context*. London, England: Routledge.

Bazerman, C. (2004). Speech acts, genres, and activity systems. In C. Bazerman & P. A. Prior (Eds.), *What writing does and how it does it* (pp. 309–340). Mahwah, NJ: Lawrence Erlbaum Associates.

Bazerman, C., & Rogers, P. (2008). Writing and secular knowledge within modern European institutions. In C. Bazerman (Ed.), *Handbook of research on writing*. New York, NY: Lawrence Erlbaum Associates.

Bennett, R. E., & Gitomer, D. H. (2009). Transforming K–12 assessment: Integrating accountability testing, formative assessment and professional support. In C. Wyatt-Smith & J. J. Cumming (Eds.), *Educational assessment in the 21st century*.New York, NY: Springer.

Bereiter, C., & Scardamalia, M. (1987). *The psychology of written composition*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Berninger, V. W. (2005). Developmental skills related to writing and reading acquisition in the intermediate grades. *Reading and Writing, 6*(2), 161–196.

Biber, D. (1980). A typology of English texts. *Language, 27*, 3–43.

Block, C. C., & Parris, S. R. (2008). *Comprehension instruction: Research based best practices*. New York, NY: Guilford Press.

Bloom, B. S. (1956). *Taxonomy of educational objectives, handbook 1: The cognitive domain*. New York, NY: Addison Wesley.

Bolter, J. D. (2001). *Writing space: Computers, hypertext and the remediation of print* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (1999). *How people learn: Brain, mind, experience and school*. Washington, DC: National Academy Press.

Britton, J., Burgess, T., Martin, N., McLeod, A., & Rose, H. (1975). *The development of writing abilities*. London, England: Macmillan.

Bruce, I. (2005). Syllabus design for general EAP writing courses: A cognitive approach. *Journal of English for Adademic Purposes, 4*, 239–256.

Burstein, J., Chodorow, M., & Leacock, C. (2004). Automated essay evaluation: The criterion online writing service. *AI Magazine, 25*(3), 27–36.

Burstein, J., & Shermis, M. D. (2003). The e-rater scoring engine: Automated essay scoring with natural language processing. In M. D. Shermis & J. C. Burstein (Eds.), *Automated essay scoring: A cross-disciplinary perspective* (pp. 113–122). Mahwah, NJ: Lawrence Erlbaum Associates.

Carter, S. (2007). *Literacies in context*. Southlake, TX: Fountainhead Press.

Charney, D. (1984). The validity of using holistic scoring to evaluate writing: A critical overview. *Research in the Teaching of English, 18*(1), 65–81.

Chenoweth, N., & Hayes, J. R. (2001). Fluency in writing. *Written Communication, 18*(1), 80–98.

Chi, M. T. H. (2000). Self-explaining expository texts: The dual process of generating inferences and repairing mental models. In R. Glaser (Ed.), *Advances in instructional psychology*. Mahway, NJ: Lawrence Erlbaum Associates.

Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science, 13*, 145–182.

Chodorow, M., & Burstein, J. (2004). *Beyond essay length: Evaluating e-rater's® performance on TOEFL essays* (TOEFL Research Rep. No. TOEFL-RR-73). Princeton, NJ: ETS.

De La Paz, S., & Graham, S. (2002). Explicitly teaching strategies, skills, and knowledge: Writing instruction in middle school classrooms. *Journal of Educational Psychology, 94*(4), 687–698.

Deane, P. (2006). Linguistic assessment of textual responses. In D. M. Williamson, R. J. Mislevy, & I. I. Bejar (Eds.), *Automated scoring of complex tasks in computer-based testing* (pp. 313–372). Mahwah, NJ: Lawrence Erlbaum Associates.

Diederich, P. B., French, J. W., & Sydell, T. (1961). *Factors in judgments of writing ability* (ETS Research Bulletin No. RB-61-15). Princeton, NJ: ETS.

Dikli, S. (2006). An overview of automated scoring of essays. *Journal of Technology, Learning, and Assessment*, *5*. Retrieved from http://escholarship.bc.edu/cgi/viewcontent.cgi?article=1044&context=jtla

Donovan, C. A., & Smolkin, L. B. (2006). Children's understanding of genre and writing development. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 131–143). New York, NY: The Guilford Press.

Duke, N. K. (2000). 3.6 minutes per day: The scarcity of informational texts in first grade. *Reading Research Quarterly, 35*(2), 202–224.

Duke, N. K. (2004). The case for informational text. *Educational Leadership, 61*(6), 40–44.

Elbow, P. (1987). Closing my eyes as I speak: An argument for ignoring audience. *College English, 49*(1), 50–69.

Elbow, P. (1994). Teaching two kinds of thinking by teaching writing. In K. S. Walters (Ed.), *Re-thinking reason: New perspectives in critical thinking* (pp. 25–32). Albany, NY: State University of New York, Albany.

Elder, L., & Paul, R. (2007). *To analyze thinking we must identify and question its elemental structures* [interactive chart]. Retrieved from http://www.criticalthinking.org/CTmodel/CTModel1.cfm

Elliot, N. (2005). *On a scale: A social history of writing assessment in America*. New York, NY: Peter Lang.

Engestrom, Y., Miettinen, R., & Punamaki, R. (1999). *Perspectives on activity theory*. Cambridge, England: Cambridge University Press.

Englert, C. S., Mariage, T. V., & Dunsmore, K. (2006). Tenets of sociocultural theory in writing instruction research. In C. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research*. New York, NY, and London, England: The Guilford Press.

Ennis, R. H. (1987). A taxonomy of critical thinking dispositions and abilities. In R. H. Ennis, J. Boykoff, & R. J. Sternberg (Eds.), *Teaching thinking skills: Theory and practice*. New York, NY: WH Freeman.

Flower, L., & Hayes, J. (1981). A cognitive process theory of writing. *College Composition and Communication, 32*(4), 365–387.

Foster, P., & Purves, A. (2001). Literacy and society with particular reference to the non-western world. In R. Barr, M. L. Kamil, P. B. Mosenthal, & P. D. Pearson (Eds.), *Handbook of reading research* (Vol. II, pp. 26–45). New York, NY: Longman.

Frederiksen, N. (1984). The real test bias: Influences of testing on teaching and learning. *American Psychologist, 39*(3), 193–202.

Gardner, S., & Powell, L. (2006). *An investigation of genres of assessed writing in British higher education: A Warwick-Reading-Oxford Brookes project.* Paper presented at the annual research, scholarship and practice in the area of academic literacies seminar, University of Westminster, London, England.

Geisler, C. (1994). *Academic literacy and the nature of expertise: Reading, writing and knowing in academic philosophy*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Goldman, S. R., & Bisanz, G. (2002). Toward a functional analysis of scientific genres: Implications for understanding and learning processes. In J. Otero, J. A. Leon, & A. C. Graesser (Eds.), *The psychology of science text comprehension* (pp. 19–50). Mahwah, NJ: Lawrence Erlbaum Associates.

Graham, S., & Harris, K. (2005). *Writing better: Effective strategies for teaching students with learning difficulties*. Baltimore, MD: Brookes Publishing Company.

Graham, S., & Harris, K. R. (2000). The role of self-regulation and transcription skills in writing and writing development. *Educational Psychologist, 35*(1), 3–12.

Graham, S., Harris, K. R., & Troia, G. A. (2000). Self-regulated strategy development revisited: Teaching writing strategies to struggling writers. *Topics in Language Disorders, 20*(4), 1–14.

Graham, S., MacArthur, C. A., Graham, S., & Fitzgerald, J. (2006). Strategy instruction and the teaching of writing: A meta-analysis. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 187–207). New York, NY: The Guilford Press.

Graham, S., & Perin, D. (2007). *A report to Carnegie Corporation of New York. Writing next: Effective strategies to improve writing of adolescents in middle and high schools*. Washington, DC: Alliance for Excellent Education.

Graves, B. (1991). Literary expertise in the description of a fictional narrative. *Poetics, 20*, 1–26.

Graves, B. (1996). The study of literary expertise as a research strategy. *Poetics, 23*(6), 385–403.

Haertel, E. (1999). Performance assessment and education reform. *Phi Delta Kappan, 80*, 662–666.

Hale, G., Taylor, C., Bridgeman, B., Carson, J., Kroll, B., & Kantor, R. (1996). *A study of writing tasks assigned in academic degree programs* (TOEFL Research Rep. No. TOEFL-RR-54). Princeton, NJ: ETS.

Hamilton, L. (2005). Assessment as a policy tool. *Review of Research in Education, 27*, 25–68.

Hayes, J. R. (1996). A new framework for understanding cognition and affect in writing. In C. M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences, and applications* (pp. 1–27). Mahwah, NJ: Lawrence Erlbaum Associate.

Hayes, J. R., & Flower. L. (1980). Identifying the organization of writing processes. In L. Gregg & E. R. Steinberg (Eds.), *Cognitive processes in writing* (pp. 3–30). Hillsdale, NJ: Lawrence Erlbaum Associates.

Heath, S. B. (1991). The sense of being literate: Historical and cross-cultural features. In R. Barr, M. L. Kamil, P. Mosenthal, & P. D. Pearson (Eds.), *Handbook of reading research* (Vol. II, pp. 3–25). New York, NY: Longman.

Hillocks, G., Jr. (1987). Synthesis of research on teaching writing. *Educational Leadership, 44*(8), 71–76, 78, 80–82.

Hillocks, G., Jr. (1995). *Teaching writing as reflective practice*. New York, NY: Teachers College Press.

Hillocks, G., Jr. (2002). *The testing trap*. New York, NY: Teachers College Press.

Hillocks, G., Jr. (2003a). Fighting back: Assessing the assessments. *English Journal, 92*(4), 63.

Hillocks, G., Jr. (2003b). *Reconceptualizing writing curricula: What we know and can use*. Chicago, IL: University of Chicago.

Holland, U. (2008). History of writing in the community. In C. Bazerman (Ed.), *Handbook of research on writing*. New York, NY: Lawrence Erlbaum Associates.

Hull, G., & Schultz, K. (2001). Literacy and learning out of school: A review of theory and research. *Review of Educational Research, 71*, 575–611.

Hung, D., & Chen, V. (2002). Learning within the context of communities of practices: A re-conceptualization of the tools, rules and roles of the activity system. *Education Media International, 39*(3/4), 248–255.

Hunt, R. A. (1996). Literacy as dialogic involvement: Methodological implications for the empirical study of literary reading. In R. J. Kreuz & M. S. MacNealy (Eds.), *Empirical approaches to literature and aesthetics*. Norwood, NJ: Ablex.

Huot, B. (1990). Reliability, validity, and holistic scoring: What we know and what we need to know. *College Composition and Communication, 41*(2), 201–213.

Hyland, K. (2003). Genre-based pedagogies: A social response to process. *Journal of Second Language Writing, 12*, 17–29.

Jiang, X., & Grabe, W. (2007). Graphic organizers in reading instruction: Research findings and issues. *Reading in a Foreign Language, 19*(1), 34–55.

Jonassen, D. H., & Rohrer-Murphy, L. (1999). Activity theory as a framework for designing constructivist learning environments. *Educational Technology, Research and Development, 47*(1), 61–79.

Kellogg, R. T. (1988). Attentional overload and writing performance: Effects of rough draft and outline strategies. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*(2), 355–365.

Kellogg, R. T. (1996). A model of working memory in writing. In C. M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences, and applications* (pp. 57–71). Mahwah, NJ: Lawrence Erlbaum Associates.

Kellogg, R. T. (1999). Components of working memory in text production. In M. Torrance & G. Jeffrey (Eds.), *The cognitive demands of writing: Processing capacity and working*

memory effects in text production (pp. 143–161). Amsterdam, The Netherlands: Amsterdam University Press.

Kellogg, R. T. (2001). Long-term working memory in text production. *Memory & Cognition, 29*(1), 43–52.

King, P. M., & Kitchener, K. S. (1994). *Developing reflective judgment: Understanding and promoting intellectual growth and critical thinking in adolescents and adults*. Ann Arbor, MI: Jossey-Bass.

Kirsch, I., & Jungeblut, A. (2002). *Literacy: Profiles of America's young adults*. Princeton, NJ: ETS.

Kuhn, D. (1991). *The skills of argument.* Cambridge, England: Cambridge University Press.

Kuhn, D. (1999). A developmental model of critical thinking. *Educational Researcher, 28*(2), 16–46.

Landauer, T. K., Laham, D., Foltz, P. W., Shermis, M. D., & Burstein, J. (2003). Automated scoring and annotation of essays with the Intelligent Essay Assessor. In *Automated essay scoring: A cross-disciplinary perspective* (pp. 87–112). Mahwah, NJ: Lawrence Erlbaum Associates.

Langer, J. A. (1992). Reading, writing and genre development: Making connections. In M. A. Doyle & J. Irwin (Eds.), *Reading and writing connections* (pp. 32–54). Newark, DE: International Reading Association.

Langer, J. A. (2001). Beating the odds: Teaching middle and high school students to read and write well. *American Educational Research Journal, 38*(4), 837–880.

Langer, J. A., & Applebee, A. N. (1986). Reading and writing instruction: Toward a theory of teaching and learning. *Review of research in education, 13,* 171–194.

Lave, J., & Wengler, E. (1991). *Situated learning: Legitimate peripheral participation*. New York, NY: Cambridge University Press.

Lindgren, E. (2005). *Writing and revising: Didactic and methodological implications of keystroke logging*: Umeå, Sweden: Modern Languages.

Marsh, J., & Millard, E. (2000). *Literacy and popular culture: Using children's culture in the clasroom*. London, England: Paul Chapman.

Martin, J. R., & Rose, D. (2006). *Genre relations: Mapping culture*. London, England: Equinox.

Matsuhashi, A. (1987). Revising the plan and altering the text. In A. Matsuhashi (Ed.), *Writing in real time: Modeling production processes* (pp. 197–223). Norwood, NJ: Ablex.

McCutchen, D. (1988). "Functional automaticity" in children's writing: A problem of metacognitive control. *Written Communication, 5*(3), 306–324.

McCutchen, D. (1996). A capacity theory of writing: Working memory in composition. *Educational Psychology Review, 8*(3), 299–325.

McCutchen, D. (2000). Knowledge, processing, and working memory: Implications for a theory of writing. *Educational Psychologist, 35*(1), 13–23.

McCutchen, D. (2006). Cognitive factors in the development of children's writing. In C. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research*. New York, NY: The Guilford Press.

McNamara, D. S., & Magliano, J. P. (2009). Self-explanation and metacognition: The dynamics of reading. In D. J. Hacker (Ed.), *Handbook of metacognition in education*. New York, NY: Taylor and Francis.

Murray, J. (2009). *Non-discursive rhetoric: Image and affect in multimodal composition*. Albany, NY: SUNY Press.

National Academy of Education. (2009). *Standards, assessments, and accountability*. Retrieved from http://www.naeducation.org/Standards_Assessments_Accountability_White_Paper.pdf

Nesi, H., & Gardner, S. (2006). Variation in disciplinary culture: University tutors' views on assessed writing tasks. In R. Kiely, P. Rea-Dickins, H. Woodfield, & G. Clibbon (Eds.), *Language, culture and identity in applied linguistics* (pp. 99–117). London, England: Equinox.

Norris, S. P., & Phillips, L. M. (1994). Interpreting pragmatic meaning when reading popular reports of science. *Journal of Research in Science Teaching, 31*(9), 947–967.

Norris, S. P., & Phillips, L. M. (2002). How literacy in its fundamental sense is central to scientific literacy. *Science Education, 87*, 224–240.

Norris, S. P., Phillips, L. M., & Korpan, C. A. (2003). University students' interpretation of media reports of science and its relationship to background knowledge. *Public Understanding of Science, 12*(2), 123–145.

Page, E. B. (2003). Project essay grade: PEG. In M. D. Shermis & J. C. Burstein (Eds.), *Automated essay scoring: A cross-disciplinary perspective* (pp. 43–54). Mahwah, NJ Lawrence Erlbaum Associates.

Paul, R., & Elder, L. (2005). *A guide for educators to critical thinking competency standards: Standards, principles, performance indicators, and outcomes with a critical thinking rubric*. Tomales, CA: Foundation for Critical Thinking.

Pressley, M. (1990). *Cognitive strategy instruction that really improves children's academic performance*. Cambridge, MA: Brookline Books.

Pressley, M., Harris, K. R., Alexander, P. A., & Winne, P. H. (2006). Cognitive strategies instruction: From basic research to classroom instruction. In P. A. Alexander & P. H. Winne (Eds.), *Handbook of educational psychology* (pp. 265–286). Mahwah, NJ: Lawrence Erlbaum Associates.

Purcell-Gates, V., Duke, N. K., & Martineau, J. A. (2007). Learning to read and write genre-specific text: Roles of authentic experience and explicit teaching. *Reading Research Quarterly, 42*(1), 8–45.

Reder, S. (1994). Practice-engagement theory: A sociocultural approach to literacy across languages and cultures. In B. M. Ferdman, R.-M. Weber, & A. Ramirez (Eds.), *Literacy across languages and cultures* (pp. 33–73). New York, NY: SUNY Press.

Resnick, L. B. (1991). Literacy in school and out. In S. R. Graubard (Ed.), *Literacy: An overview by 14 experts*. New York, NY: The Noonday Press.

Rosenfeld, M., Courtney, R., & Fowles, M. E. (2004). *Identifying the writing tasks important for academic success at the undergraduate and graduate levels* (GRE Board Research Rep. No. 0-04R). Princeton, NJ: ETS.

Rouet, J. F., Favart, M., Britt, M. A., & Perfetti, C. A. (1997). Studying and using multiple documents in history: Effects of domain expertise. *Cognition and Instruction, 15*(1), 85–106.

Russell, D. R. (1997). Rethinking genre in school and society: An activity theory analysis. *Written Communication, 14*, 504–554.

Shanahan, T. (2006). Relations among oral language, reading, and writing development. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 171–183). New York, NY: The Guilford Press.

Shermis, M. D., Burstein, J., & Bliss, L. (2004, April). *The impact of automated essay scoring on high stakes writing assessments*. Paper presented at the National Council on Measurement in Education, San Diego, CA.

Souvignier, E., & Mokhlesgerami, J. (2005). Using self-regulation as a framework for implementing strategy instruction to foster reading comprehension. *Learning & Instruction, 16*(1), 57–71.

Street, B. V. (2003). What's 'new' in new literacy studies? Critical approaches to literacy in theory and practice. *Issues in Comparative Education, 5*(2), 1–14.

Swales, J. M. (1990). *Genre analysis: English in academic and research settings*. Cambridge, England: Cambridge University Press.

Tower, C. (2003). Genre development and elementary students' informational writing: A review of the literature. *Reading Research and Instruction, 42*(4), 14–39.

van Gelder, T. (2005). Teaching critical thinking: Some lessons from cognitive science. *College Teaching, 45*(1), 1–6.

van Gelder, T., Bissett, M., & Gumming, G. (2004). Cultivating expertise in informal reasoning. *Canadian Journal of Experimental Psychology, 58*(2), 142–152.

Venezky, R. L. (1991). The development of literacy in the industrialized nations of the West. In R. Barr, M. L. Kamil, P. B. Mosenthal, & P. D. Pearson (Eds.), *Handbook of reading research* (Vol. II, pp. 46–67). New York, NY: Longman.

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

Vipond, D., & Hunt, R. A. (1987). Aesthetic reading: Some strategies for research. *English Quarterly, 20*(3), 178–183.

Vipond, D., Hunt, R. A., Jewitt, J., & Reither, J. (1990). Making sense of reading. In R. Beach & S. Hynds (Eds.), *Developing discourse practices in adolescence and adulthood* (pp. 110–135). Norwood, NJ: Ablex.

Voss, J. F., Greene, T. R., Post, T. A., & Penner, B. C. (1983). Problem-solving skill in the social sciences. In G. H. Bower (Ed.), *The psychology of learning and motivation* (pp. 165–213). New York, NY: Academic Press.

Voss, J. F., & Wiley, J. (2006). Expertise in history. In K. A. Ericsson (Ed.), *The Cambridge handbook of expertise and expert performance* (pp. 569–584). Cambridge, England: Cambridge University Press.

Vygotsky, L. S. (1978). *Mind and society: The development of higher mental processes*. Cambridge, MA: Harvard University Press.

Weigle, S. C., Bachman, L. F., & Alderson, J. C. (2002). *Assessing writing*. Cambridge, England: Cambridge University Press.

White, E. M. (1985). *Teaching and assessing writing*. San Francisco, CA: Jossey-Bass.

White, E. M. (2004). The changing face of writing assessment. *Composition Studies, 32*(1), 109–116.

White, E. M. (2005). The scoring of writing portfolios: Phase 2. *College Composition and Communication, 56*(4), 581–600.

Wineburg, S. S. (1991a). Historical problem solving: A study of the cognitive processes used in the evaluation of documentary and pictorial evidence. *Journal of Educational Psychology, 83*(1), 73–87.

Wineburg, S. S. (1991b). On the reading of historical texts: Notes on the breach between school and academy. *American Educational Research Journal., 28*(3), 495–519.

Wineburg, S. S. (1994). The cognitive representation of historical texts. In G. Leinhardt, I. Beck, & C. Stainton (Eds.), *Teaching and learning in history* (pp. 85–135). Hillsdale, NJ: Lawrence Erlbaum Associates.

Wineburg, S. S. (1998). Reading Abraham Lincoln: An expert/expert study in the interpretation of historical texts. *Cognitive Science, 22*, 319–346.

Yancey, K. (1999). Looking back as we look forward: Historicizing writing assessment. *College Composition and Communication, 50*(3), 483–502.

Yi, L. Y. (2007). Exploring the use of focused freewriting in developing academic writing. *Journal of University Teaching and Learning Practice, 4*(1), 41–53.

Zeits, C. M. (1994). Expert-novice differences in memory, abstraction, and reasoning in the domain of literature. *Cognition and Instruction, 12*(4), 277–312.

Notes

[1] This is a convenience sample, not balanced for representativeness.

[2] Within the writing community, there is both support and opposition to the use of automated essay scoring. A typical objection is that found in the Conference on College Composition and Communication (CCCC) Position Statement on Teaching, Learning and Assessing Writing in Digital Environments (retrieved Nov. 9, 2009, from http://www.ncte.org/cccc/resources/positions/digitalenvironments), which makes the very important point that current automated essay scoring systems do not measure rhetorical and conceptual quality and, if used alone, eliminate the human audience that is intrinsic to writing as a mode of communication. See also Charney (1984).

Appendix

Reflective Strategies, Genres, and Writing Development

The tables that follow summarize current thinking within the CBAL writing assessment project about the kinds of reflective conceptual strategies that students need to master to achieve high levels of reading, writing, and critical thinking skill (Table A1), how particular genres draw upon these strategies (Table A2), and some rough initial estimates about the grade levels at which particular genres might reasonably be introduced (Table A3). These form the basis for planned efforts to continue to build a range of writing assessments to cover primary and elementary grade writing assessment.

The lists of genres, strategies, and estimates of grade levels presented here draw rather heavily upon research into genres, especially genres used in academic contexts such as college and graduate school, and genre pedagogy (Bazerman, 2004; Donovan & Smolkin, 2006; Gardner, 2008; Goldman & Bisanz, 2002; Hyland, 2003; Martin & Rose, 2006; Purcell-Gates, Duke, & Martineau, 2007; Rosenfeld, Courtney, & Fowles, 2004; Swales, 1990; Tower, 2003). This list uses some of the terminology for genres that comes from this literature but adapted both genre labels and descriptions in the light of the research cited above on the cognition of writing and its relation to strategies and critical thinking.

It is important to recognize that these tables are intended as rough summaries. Table A1 provides a rough summary idea of the kinds of critical thinking that are also important to support reading comprehension and effective writing. Table A2 provides a rough summary of the kinds of scaffolding tasks that might be appropriate to support students learning to write in particular genres. Table A3 is designed to help focus future development work, but will be no substitute for the actual articulation of a sequence of assessments at different grade levels, for such a sequence, when completed, will be far more self-explanatory than the contents of this appendix. The problem for future work will be to translate the vision presented in this paper into a concrete series of assessment models articulated over multiple grades.

Table A1

Strategies for Reading, Writing, and Rethinking

Exploration	Explication	Modeling	Judgment
Contingency modeling—systematically modeling contingencies that could affect a plan by systematically varying starting conditions, outcomes, and interventions	Means/end planning—Explicitly setting or recognizing goals, sub-goals, obstacles and methods to overcome them; involves metacognitive and self-regulation strategies that support chunking tasks into pieces of manageable size	Heuristics—considering a range of cases to extract a common principle that can then be used to define strategies for solving new cases; involves synthesis by analogy across cases	Standard-setting (ethos)—appealing to ethical, moral and efficacy standards to determine whether a course of action is appropriate; requires the ability to apply standards to specific cases, plus the active application of principles of moral reasoning and decision-making to work out and refine standards and define consistent and appropriate ways to apply them
Self-explanation—reviewing and rethinking facts or events to gain new insights particularly with regard to reasons, causes and why one feels as one does; substrategies include rereading, notetaking, think-aloud, and freewriting	Social simulation—modeling the perspective, motivations, goals, actions, and reactions of different participants so as to understand the dynamics that govern an event or event sequence	Reconciliation—considering alternate accounts of the same events and finding ways to integrate them while reconciling differences among them based on differences in perspective, reliability, and immediacy of evidence	Appeal-building (pathos) —creating motivation for people to accept particular explanations, characterizations, or courses of action by appealing to their purposes, emotions, and values
Guiding questions—generating content predictions and high-level questions to elaborate representation of content; stimulated by skimming, pre-reading, and brainstorming activities	Close questioning—self-reflection (such as devising a series of specific questions) to identify causal/factual gaps, inconsistencies, and vagueness in what one knows, and use them to devise a clearer formulation or to supply bridging inferences	Close reading—formulating readings of texts based upon close attention to phrasing, implication, allusion, subtext, and other elements that reflect transactions among author, audience, immediate context, and literary tradition; involves integration of multiple clues to support an interpretation	
Concept mapping—systematically exploring what one knows about a particular domain by explicitly mapping out major entities, facts, and relationships; substrategies include knowledge-based inferencing and use of graphical representations (concept maps), plus consultation of external references	Defining—using context and background knowledge to define terms and conceptual categories; substrategies include analogy, comparison/contrast, identification of necessary/typical attributes		Argument-building (logos)—constructing chains of reasoning that support conclusions on the basis of evidence; substrategies include active use of logical reasoning to elaborate one's own knowledge and critical application of reasoning to identify questionable or uncertain information
	Outlining—organizing information in terms of relatedness and relative importance, often graphically; involves visualization, paraphrasing, and selection of key ideas	Hypothesis-testing—formulating a hypothesis to cover observations, making predictions, and conducting experiments to confirm the resulting predictions	

Table A2

Genres Strongly Exercising Particular Conceptual Strategies

Strategy	Genres strongly exercising that strategy
Means-end planning	Procedure—directions how to perform an action Problem statement—broad descriptions of a task to be accomplished
Contingency modeling	Method—directive text that explains reasons as well as procedures Proposal—text proposing a specific plan detailing how goals will be accomplished Causal account—explanation of phenomena in terms of causes & consequences
Heuristics	Case study—specific case presented as illustration of principles Manual—multiple procedures synthesized into systematic account
Self-explanation	Reader response—free reaction to reading Note taking—self-explanation as an aid to memory Anecdote—description for expressive purposes Description—concrete presentation of things one knows Summary—self-explanation of core content of reading
Guiding questions	Description, report—systematic presentation answering key questions Annotation—comments on text raising questions and issues Explication—systematic explanation of the information presented in a text, intended to clarify and expand on key information
Outlining	Recount—basic presentation of events in sequence Summary, synopsis—summary of a narrative focusing on key events & their causes Survey—text combining information from multiple sources to create coherent picture
Defining	Gloss—annotations defining key ideas or terms Comparison/contrast—ideas defined by identifying shared and unique attributes
Simulation/ Roleplay	Narration—presentation of a story with full attention to literary elements Commentary—explanation elaborating on story elements and their significance Interpretive review—explanation of story justifying interpretations Interpretive account—systematic analysis of reasons and motivations
Close reading	Explication, interpretive account, historical account—combination of information from multiple sources to describe historical events and their causes Literary analysis—coherent interpretation drawing on multiple literary texts
Reconciliation	Historical account – synthesis giving sequential and causal account of events based on analysis of sources taking reliability and perspective into account Survey – synthesis of information on a topic based on integration of source materials Discussion—information objectively presented on an issue without taking sides
Hypothesis-testing	Theoretical account—model presented and fitted it to range of facts/observations Experimental report—data presented and organized to evaluate how well it fits a model
Appeal-building	Promotion—persuasion focused on action and emotional appeal Recommendation—evaluation of choices; persuasion focused on alternatives
Standard-setting	Apology—defense of rightness of actions Exemplum—story implicitly presenting actions as model to emulate or avoid
Argument-building	Discussion, essay—advance specific thesis and logically defend it with evidence Critique—evaluation of the arguments advanced in a text Rebuttal—text examining arguments of others and presenting reasons to reject them

Table A3

Approximate Grade Ranges at Which Particular Genres in Table A2 Might Be of Interest for Assessment Research

Grade levels	Genre categories
3–5	Anecdote, reader response
3–5	Recount, procedure
4–6	Description, report
4–6	Comparison/contrast, illustrative account
5–7	Synopsis, narration
5–7	Summary, account
6–8	Gloss, annotation, note-taking
6–8	Apology, exemplum
7–9	Explication, commentary, interpretive review
7–9	Problem statement, survey
7–9	Promotion, recommendation
8–10	Method, proposal, experimental report
8–10	Discussion, essay
9–11	Rebuttal, critique
9–11	Case study, manual
10–12	Theoretical account, interpretive account
10–12	Literary analysis, historical account