

CHAPTER 5.

RETHINKING K-12 WRITING ASSESSMENT TO SUPPORT BEST INSTRUCTIONAL PRACTICES

Paul Deane, John Sabatini, and Mary Fowles

Educational Testing Service

The work described in this chapter arises in a specific current United States context: one in which there is tension between best practices in writing instruction and standard approaches to writing assessment, particularly in the context of standardized tests, such as state accountability examinations. On the one hand, the instructional literature emphasizes the importance of teaching deliberate, well-developed writing processes; indicates the value of peer review; and strongly supports an instructional approach in which explicit strategies are taught in meaningful contexts where content matters (Perin, 2009). On the other hand, the requirements of standardized testing often favor creation of relatively decontextualized writing tasks in which students produce essays under timed conditions with little access to external sources of information—a state of affairs that may have deleterious effects on writing instruction, since the predominance of high-stakes assessments often forces instructors to focus on test preparation (Hillocks, 2002, 2008).

Not surprisingly, current reform efforts such as the Race to the Top program partially conceive of assessments as interventions, and various scholars have argued that such reforms should be guided by modern theories of learning and cognition (Pellegrino, 2009). This point has been advanced in particular by Bennett and Gitomer (2009), who propose a research program that they term CBAL (Cognitively-Based Assessment of, for and as Learning). The goal of this research program is to conceptualize and try out the components of an integrated system in which summative assessments, formative assessments, and teacher professional support combine to encourage and enhance effective teaching and learning.

This chapter presents some initial results from the CBAL program of research as it applies to writing. In particular, we have developed a framework that draws upon an extensive review of the literature on writing and related literacy

skills; developed methods for designing summative and formative assessments to measure writing skill while modeling best practices in writing pedagogy; and begun analyzing results of preliminary (but in many cases, large-scale) pilots.

Our work is still in its early stages (though see Deane, 2011; Deane, Fowles, Baldwin, & Persky, 2011; Deane, Quinlan & Kostin, 2011; Deane, Quinlan, Odendahl, Welsh & Bivens-Tatum, 2008), but one central theme has begun to emerge: that writing must be conceptualized within an integrated system of socially-embedded literacy skills.

Effective assessment design requires us to construct an interpretive argument that connects construct theory to task demands through a chain of evidence elicited by those tasks (Kane, 2006; Mislevy & Haertel, 2006). Reading science has documented a developmental trajectory in which certain skills, such as decoding and verbal comprehension, start relatively independent, and gradually become integrated and intertwined as expertise develops (Hoover & Gough, 1990; Vellutino, Tunmer, & Jaccard, 2007). Over time, some cognitive processes become increasingly fluent and automatized, while others become increasingly strategic and more finely sensitive to details of the situation in which communication or understanding must take place (Scarborough, 2001). It thus becomes necessary to consider skills both as stand-alone capabilities and as capacities invoked as part of a larger, more complex skill-set. Similar, parallel considerations apply to the description of the development of writing expertise. Developing writing expertise involves both skill development and their attachment as strategic resources within an activity system. The problem is that we must measure both the fluency and accuracy of skills as stand-alone tasks and the effectiveness with which readers/writers/thinkers can employ those skills flexibly to accomplish specific literacy goals; thus, we must conceptualize reading and writing not merely as individual skills, but also as interdependent and mutually supporting tools for social interaction.

The literacy framework developed for the CBAL project is based upon this kind of developmental trajectory and predicts that parallel expressive, interpretive, and deliberative skills invoke common, shared 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. One of the advantages and contributions of a combined reading/writing (English Language Arts) model is that it helps us to focus on the presence of common, shared cognitive resources deployed in literacy activity systems, whether the channel/modality itself is primarily reading, writing, or thinking. But in actual educational practice, reading and writing are typically treated separately, particularly in middle and upper grades, though both are taught by

English Language Arts (ELA) teachers. While classroom practice often integrates reading and writing, there are only scattered attempts to integrate theoretical models of reading and writing development, and even less effort to build an assessment system sensitive to an integrated literacy model. Ironically, reading and writing are frequently combined in high stakes assessments; on reading tests, students are asked to write to show they understand what they have read; on some writing tests, students are not even asked to produce writing, but only to edit, revise, or identify errors in sentences or simulated written compositions.

Treating reading and writing separately seems a missed opportunity with potentially negative learning consequences. If reading literacy is privileged in school at the expense of writing skill development, students may be ill-prepared for secondary and post-secondary academic learning which puts increasingly higher demands on the ability to express one's thinking (about what one reads) in written forms. Building more complex, integrated literacy assessments aligns well with best instructional practice and may assist learners in developing the full range of literacy skills.

In general, the community of reading researchers tends to acknowledge that writing instruction supports reading development, but relatively few researchers cross the precipice and see them as jointly determined. Often, writing instruction has been the historical runt of the reading, writing, and 'rithmetic litter. When state accountability scores drop, explicit reading and math interventions often squeeze out time for writing instruction. Yet reading scores continue stagnating nationally, which means that millions of children, adolescents, and adults have inadequate reading, writing, and likely, thinking skills. An integrated approach that emphasizes learning to write fluently and thoughtfully may also provide the most effective and efficient pathway to thoughtful reading, if only because cognitive reading processes are mostly invisible, whereas the processes of written composition can be made visible, transparent, and an object of metalinguistic reflection (Olson, 1991; Olson & Hildyard, 1985). The model described here is the starting point for mapping out the interdependencies we want to foster, even as the assessment challenge of separating them remains the target of most standards statements and external testing programs. We must first reform the social construct in order to assess the cognitive construct more productively.

MAJOR THEMES

Our perspective treats reading, writing, and critical thinking as integrated activity systems in the Vygotskian sense (Vygotsky & Luria, 1994) and along

lines discussed in Bazerman (2004). We envisage writing (and, in particular, specific genres of writing) as forming part of an integrated set of tools for social communication. Thus, as Deane (2011) outlines in greater detail, writing skill is inherently intertextual, involving an interplay of skills that might, in isolation, be considered reading or critical thinking (Bloome, 1993). This point can be supported in part by considering the many shared elements that play roles not only in theories of writing but also in theories of reading comprehension, such as verbal comprehension (Hoover & Gough, 1990) and text macrostructure (Graesser, Singer, & Trabasso, 1994). It can be reinforced by observing the key role that reading, deliberation, and reflection skills play in classical models of writing (Alamargot & Chanquoy, 2001; Bereiter & Scardamalia, 1987; Hayes, 1996; Hayes & Flower, 1980). But the importance of intertextuality emerges most strongly when we consider particular genres of writing and analyze the specific configurations of skills that are required in particular genres, along lines exemplified by Coirier, Andriessen and Chanquoy (1999).

In addition, we have found that it is useful to think of writing skill as involving the acquisition of specific skill bundles (only some of which are writing-specific) and their progressive elaboration and generalization. Beyond general fluency and accuracy of written production, there is specific evidence that progress in writing in particular genres is tied to the developmental sequence observed for specific skills.¹ For instance, progress in writing narratives seems to depend critically upon acquiring the ability first to represent event sequences causally in terms of character motivations and goals; and second, in acquiring the ability to represent narratives metacognitively as interpretive acts enacted by the author (McKeough, 2007; Nicolopoulou & Bamberg, 1997; Nicolopoulou, Blum-Kulka, & Snow, 2002). Similarly, the development of skill in argumentative writing partially reflects the underlying development of argumentation skills (Felton & Kuhn, 2001; Kuhn, 1999; Kuhn & Udell, 2003). When we analyze particular genres in this fashion, intertextual dependencies also emerge from the ecology of the activity system. For example, argumentative writing critically depends upon summary skills and not just the ability to create arguments, since participation in argument nearly always entails a response to prior and opposing points of view. Similarly, literary analysis critically depends upon the ability to find and explain evidence for interpretations in a text, and more generally upon the ability to evaluate and respond to such interpretations. Many of these dependencies are recognized in educational standards, at least implicitly. For instance, in the Common Core State Standards that have been adopted by 44 of the 50 US states (<http://www.corestandards.org/the-standards/english-language-arts-standards>), many of the language arts standards specifically address such skills as building a mental model of the events in a narrative, creating arguments, or finding evidence in a source text.

The considerations sketched thus far lead in specific directions: (i) toward assessment design strategies that borrow many features of performance assessments, (ii) by assessment designs that incorporate a (relatively) meaningful context and arrange task sequences so that their application outside the assessment context is transparent to students and teachers.

DESIGNING WRITING ASSESSMENTS TO SUPPORT INSTRUCTION

DESIGN CONSIDERATIONS

Any kind of formal assessment creates tradeoffs. The more we seek to standardize tests, to make tests equivalent and generalizable, the harder it is to capture interdependencies among tasks, to communicate why one task supports another, or to communicate the social context that motivates particular skills. On the other hand, pure performance tasks may create measurement and scoring difficulties. Within the overall CBAL research framework, we are researching both high-stakes assessments (where the pressures for standardization are greatest), and formative, classroom assessments (where performance tasks are often favored, yet must still be reliable enough to support instructional decisions based upon student performance).

In this chapter we primarily discuss our designs for high-stakes tests, with an emphasis on features intended to make high-stakes testing more supportive of instruction. We emphasize, in particular, features that make the high-stakes assessments more transparent—in the sense that each test exemplifies appropriate reading and writing practices and provides instructionally actionable results. Some of these considerations are not specific to writing but are particularly problematic for writing because of the time required to collect a single written response of any length. For instance, reliable estimates of ability require multiple measurements; and in the case of writing, that means a valid writing assessment will collect multiple writing samples on multiple occasions. Solving this problem is fundamental to the CBAL approach, which is focused on exploring the consequences of distributing assessments throughout the year (which makes it easier to collect multiple samples, but is likely to be feasible only if the high-stakes assessments are valuable educational experiences in their own right.)

Our goal is to develop a series of assessments that might be given at intervals, sampling both reading and writing skills, over the course of the school year. One logical way to do this is to focus each individual test on a different genre (but to sample systematically from all the reading, writing, and thinking skills

necessary for success at each genre.) Having taken this first step, it becomes possible to introduce many of the features of a performance task into a summative design without sacrificing features necessary to produce a reliable instrument under high-stakes conditions.

STRUCTURE OF INDIVIDUAL ASSESSMENTS.

Certain design decisions are well-motivated if we conceive of individual writing assessments as occasions to practice the skills needed for success in particular genres, and plan from the beginning to provide multiple assessments during the course of the school year. In particular, the CBAL writing assessments have a common structure, involving:

- A unifying scenario
- Built-in scaffolding
- Texts and other sources designed to provide students with rich materials to write about
- Lead-in tasks designed to engage students with the subject and measure important related skills
- A culminating extended writing task

The Scenario

Rather than presenting a single, undifferentiated writing task, each test contains a series of related tasks that unfold within an appropriate social context. The scenario is intended to provide a clear representation of the intended genre and social mode being assessed, to communicate how the writing task fits into a larger social activity system, and to make each task meaningful within a realistic context. By their nature, such scenarios are simulations, and may not capture the ultimate social context perfectly; but to the extent that they transparently represent socially meaningful situations within which students may later be required to write (either inside or outside of school), the scenario helps to make explicit connections between items that (i) communicate why each item has been included on the test; (ii) help teachers connect the testing situation to best instructional practices; and (iii) support the goal of making the assessment experience an opportunity for learning.

Scaffolding

Building scaffolding elements into the test helps the assessment model best practices in instruction. Such elements include:

- Lead-In Tasks, which may involve reading or critical thinking activities and consist of selected-response as well as sentence- or paragraph-length writing tasks. Lead-in tasks are intended to satisfy several goals at once, to prepare students to write, to measure skills not easily measured in an extended writing task, and to exercise prerequisite skills.
- Task supports such as rubrics that provide explicit information about how student work will be judged, tips and checklists that indicate what kinds of strategies will be successful, and appropriate reference materials and tools to support reading comprehension and thinking. These materials are included to minimize irrelevant variation in student preparation that could obscure targeted skills.

Supporting Texts

Rather than asking students to write about generic subjects, we provide supporting texts intended to inform students about a topic and stimulate their thinking before they undertake the final, extended writing task. The goal is to require students to engage in the kinds of intertextual practices that underlie each written genre.

The Extended Culminating Writing Task

In each test, we vary purpose and audience, and hence examine different social, conceptual and discourse skills, while requiring writers to demonstrate the ability to coordinate these skills to produce an extended written text.

This general design is instantiated differently depending on what genre is selected for the culminating extended writing task. Each genre has a well-defined social purpose, which defines (in turn) a specific subset of focal skills. For the classic argumentative essay, for example, focal skills include argument-building and summarization. Given this choice, the problem is to create a sequence of lead-in tasks that exercise the right foci and thus scaffold and measure critical prerequisite skills. For a different, paradigmatic writing task, such as literary analysis, focal skills include the ability to identify specific support for an interpretation in a text, and to marshal that evidence to support and justify one's own interpretations. These are different kinds of intertextuality, supporting very different literacy practices. The final writing task is meaningful only to the extent that writers are able to engage in the entire array of reading and writing practices associated with each genre.

AN EXAMPLE: TWO MIDDLE SCHOOL WRITING/READING TESTS

Tables 1 and 2 show the structure of two designs that might be part of a single year's sequence of reading/writing tests: one focuses on the classic argumentative essay; the other, on literary analysis. The argumentation design contains a series of lead-in tasks designed, among other things, to measure whether students have mastered the skills of summarizing a source text and building an argument, while simultaneously familiarizing students with the topic about which they will write in the final, extended writing task. The literary analysis design contains a series of lead-in tasks intended to measure whether students have the ability to find textual evidence that supports an interpretation, can assess the plausibility of global interpretations, or can participate in an interpretive discussion.

An important feature of this design is that the lead-in tasks straddle key points in critical developmental sequences. Thus, the test contains a task focused on classifying arguments as pro or con—a relatively simple task that should be straightforward even at relatively low levels of argument skill. It also contains a rather more difficult task—identifying whether evidence strengthens or weakens an argument—and a highly challenging task, one that appears to develop relatively late, namely the ability to critique or rebut someone else's argument.

Note that a key effect of this design is that it includes what are, from one point of view, reading or critical thinking tasks in a writing test, and thus enables us to gather information about how literacy skills vary or covary when applied within a shared scenario. In addition, since the tests are administered by computer, we are able to collect process data (e.g., keystroke logs) and can use this information to supplement the information we can obtain by scoring the written products. The CBAL writing test designs thus provide a natural laboratory for exploring how writing skills interact with, depend upon, or even facilitate reading and critical thinking skills.

Also note that the culminating task is not (by itself) particularly innovative. One could be viewed as a standard persuasive essay writing prompt; the other, as a fairly standard interpretive essay of the kind emphasized in literature classes. This is no accident; the genres are well-known, and exemplary tasks have been chosen as exemplars because of the importance of the activity system (i.e., the social practices) within which they play key roles. Our contribution is to examine how these complex, final performances relate to other activities, often far simpler, that encapsulate key abilities that a skilled reader/writer is able to deploy in preparation for successful performance on the culminating activity.

In other words, within the perspective we have developed, we characterize all of these skills as literacy skills, and treat writing as a sociocognitive construct.

We are currently developing a variety of materials—some of them designed for use in the classroom as formative assessments, and some designed as teacher professional support—as part of the CBAL language arts initiative. In collaboration with classroom teachers at several pilot sites, we have begun exploring a wide range of questions about writing and its relation to reading and thinking skills. As noted above, the literature contains considerable evidence that reading and writing share a large base of common skills. But there is relatively little evidence about how these skills interact in the course of complex literacy tasks. The designs we have been developing are intended in part to support research intended to construct a shared reading/writing literacy model.

Table 1. Design: Argumentation

(Lead-in tasks help prepare students to write an essay on a controversial issue. Task supports include summary guidelines, essay rubrics, and planning tools, embedded as tabs accessible from the same screen as each item.)		
Item Description	Timing (min.)	Task Description
Lead-in Section (Task 1, Part 1) (Five selected-response items)	15	Apply the points in a summarization rubric to someone else's summary of an article about the issue.
Lead-in Section (Task 1, Part 2) (Two short constructed-response items)		Read and summarize two articles about the issue. (One with a simple macrostructure, another with a more complex one.)
Lead-in Section (Task 2, Part 1) (Selected Responses, 10 binary-choice items)	15	Determine whether statements addressing the issue are presenting arguments pro or con.
Lead-in Section (Task 2, Part 2) (Six multiple-choice items)		Determine whether specific pieces of evidence will weaken or strengthen particular arguments.
Lead-in Section (Task 3) (One short constructed-response item)	15	Critique someone else's argument about the issue.
Culminating Task (One long constructed-response-item)	45	Write an argumentative essay taking a position on the issue.

Table 2. Design: Literary analysis

Note: Lead-in tasks help students prepare to write an essay on how an author develops ideas over several passages in a literary work. Task supports include essay rubrics and planning tools, embedded as tabs accessible from the same screen as each item.		
Item Description	Timing (min.)	Task Description
Lead-in Section (Task 1) (Five selected-response items)	20	Choose evidence to support an inference about a literary text.
Lead-in Section (Task 2) (One short constructed-response item)	15	Contribute to an interpretive discussion about a literary text.
Lead-in Section (Task 3, Part 1) (Six selected-response items)	15	Decide on the best justification for a global interpretation of a literary text.
Lead-in Section (Task 3, Part 2) (One short constructed-response item)		Explain briefly what has been learned by reading and interpreting passages from a literary text (including interpretation of figurative language embedded in the text).
Culminating Task (One long constructed-response item)	45	Write a literary analysis describing the effects achieved by a combination of passages in a literary text and identifying evidence from the text that helps illustrate how these effects were achieved.

MEASURING RELATIONSHIPS AMONG READING, WRITING AND CRITICAL THINKING SKILLS IN A COGNITIVELY BASED ASSESSMENT

The assessments based on this framework have been designed systematically to probe reading, writing, and thinking interrelationships. Several such assessments have been field-tested, including the two described in preceding parts of this chapter. As part of the field testing, we collected various sources of evidence: not only the student responses, but also keystroke logs capturing timing data for the culminating, written response, and we subjected student responses to analysis using natural language processing techniques. The result is a rich dataset that enables us to address a variety of issues relating reading and writing skills. In this section of the chapter, we briefly explore some of the research questions that can be addressed as a result.

FIELDTEST DESIGN

In 2009 multi-state field tests, 2,606 eighth grade students were administered two test forms selected from four forms total, in a counterbalanced design (that is, randomly selected students took each ordered combination of tests, which allows us to verify that the order or specific choice of tests did not change our results). Two of these forms are discussed in this chapter.² A total of 1,054 students completed the first form analyzed here (the argumentation design), and 1,109 completed the second form (the literary analysis design), with 293 students completing both forms. The sample was about half female, with a range of ability levels, socioeconomic status (SES), and race/ethnicity. SES, race/ethnicity and English language proficiency information was collected by survey, and noncompletes ranged as high as 45% on some questions; but of the students that responded, about half were low-SES (on free and reduced lunch), less than five percent reported having English Language Learner (ELL) status, and a majority (62%) were white, with substantial African-American (22%) and Hispanic (12%) subpopulations. Each assessment was completed in two 45-minute sessions, and focused on a specific written genre. The lead-in tasks occupied the first 45 minutes. The essay was written in the second 45 minute session, which took place either immediately after the first session, or with a few days' gap in between. Test reliability was high (literary interpretation, $\alpha=.81$, 13 items; persuasive essays, $\alpha=.76$, 25 items³).

RELATIONSHIP BETWEEN READING AND WRITING SCORES

If we conceptualize the lead-in tasks as basically reading/thinking tasks, and the culminating tasks as writing tasks, the correlations were moderate to high when all reading tasks were summed to give a total score, as shown in Table 3.⁴ This level of correlation led us to ask whether these assessment task designs capture evidence of shared thoughtful (deliberative) cognitive processing deployed in reading (interpretive) and writing (expressive) tasks, and how they are inter-related. Specifically, we wondered:

1. Do the scores for each reading/thinking task set contribute unique variance to the prediction of holistic composition scores?
2. Do holistic written composition scores contribute unique variance to the prediction of scores on the reading/thinking tasks over and above that contributed by other reading/thinking task set scores?⁵

Table 3. Pearson correlations between human essay scores and total lead-in tasks within and across prompts ($p < .001$)

	Argumentative Lead-in Task Score	Literary Interpretation Lead-In Task Score
Ban Ads Essay Score	.684	.607
Mango Street Essay Score	.584	.705

To address these questions, we ran a series of regression models predicting Essay scores (Task 4) from Task Sets (1-3). In each case, the results were highly significant ($R^2 = .474$ & $R^2 = .464$ for Literary and Persuasive respectively). Each Task Set was a significant predictor and added unique variance when added stepwise to the model. A second series of regression models were run predicting each Task Set, with other Task Sets entered first and Essay scores last. Again, in most cases, essay score predicted additional, unique variance (see tables 4, 5, 6 and 7).⁶

Table 4. Predicting the argument essay score from lead-in tasks (Adj. $R^2 = .464$)

DV	B	Standard Error	Beta	t	Significance	Correlation
(Constant)	-.046	.113		-.407	.684	
Task_1	.163	.022	.193	7.504	.000	.440
Task_2	.184	.021	.234	8.679	.000	.506
Task_3	.443	.028	.428	15.789	.000	.611

Table 5. Predicting the interpretative essay score from lead-in tasks (Adj. $R^2 = .474$)

DV	B	Standard Error	Beta	t	Significance	Correlation
(Constant)	0.357	0.061		5.901	0	
Task_1	0.082	0.009	0.245	8.667	0	.545
Task_2	0.444	0.029	0.399	15.168	0	.605
Task_3	0.131	0.019	0.192	6.888	0	.514

Table 6. Predicting the lead-in tasks from argumentative essay score

Model	Adjusted R Square	R Square Change	Significance of R Square Change
Model 1: Task 1 predicted from Tasks 2 & 3	.186	.043	.000
Model 2: Task 1 predicted from Tasks 2,3 and Essay	.229		.000
Model 3: Task 2 predicted from Tasks 1 and 3	.260	.051	.000
Model 4: Task 2 predicted from Tasks 1, Task 3 and the Essay	.311		.000
Model 5: Task 3 predicted from Tasks 1 and 2	.268	.146	.000
Model 6: Task 3 predicted from Task 1, Task2 and the Essay	.414		.000

Table 7. Predicting the lead-in tasks from the interpretive essay score (Adj. R^2 = .366)

Model	Adjusted R Square	R Square Change	Significance of R Square Change
Model 1: Task 1 predicted from Tasks 2 and 3	.379	.041	.000
Model 2: Task 1 predicted from Tasks 2, 3, and Essay	.420		.000
Model 3: Task 2 predicted from Tasks 1 and 3	0.281	0.128	.000
Model 4: Task 2 predicted from Tasks 1, 3, and Essay	0.409		.000
Model 5: Task 3 predicted from Tasks 1 and 2	.362	.027	.000
Model 6: Task 3 predicted from Tasks 1, 2, and Essay	.389		.000

As we can see in Tables 4 and 5, each reading task contributes separate variance to predicting the writing score; and, as Tables 6 and 7 indicate, the writing score contributes additional unique variance above and beyond that contributed by the other lead-in tasks. The pattern of performance is consistent with (though of course not sufficient to demonstrate) the kind of interpretation we would suggest—in which reading and writing draw upon a common base of

shared skills, and typically are most efficiently acquired and exercised as an integrated skill-set.

These examples illustrate one kind of research strategy enabled by the CBAL assessment framework. This strategy provides a research-based justification for an integrated approach to ELA literacy instructional and assessment that views reading, writing, and thinking as mutually reinforcing skills that draw upon shared mental representations. This study and its results comprise promising first steps. The test forms demonstrated feasible implementation and scoring, acceptable psychometrics, and patterns of results in the directions predicted by the framework and design.

CONCLUSIONS AND FUTURE DIRECTIONS: FROM WRITING TO READING AND THINKING

This chapter represents collaboration between researchers who previously focused separately on writing and reading. In the parlance of visual art, we have taken a “one-point” perspective so far, focusing on the development of writing skills and assessments. We have not neglected reading (interpretive) and critical thinking (deliberative) skills, but they have been viewed through the writing lens. Once we commit to the idea that writing skill must be assessed within a larger context, then we may want to view this landscape from two- or three-point perspective. That is, once we recognize that the act of writing may incorporate a whole series of literacy acts that do not directly involve text production, it becomes necessary to give a much more detailed accounting of the relationship among skills that puts equal focus on reading and thinking as activities in their own right.

In fact, the kind of integrated model we have proposed leads naturally to a position in which a three-point perspective is viewed as the norm toward which educational practice should strive. Shared, mutually supportive cognitive representations do not necessarily emerge spontaneously in the untrained, developing reader/writer/thinker. The pedagogical literature suggests (Langer, 2001) that these kinds of skills are promoted and developed by classroom learning and instruction that take advantage of and foster their integrated construction and use in social literacy practice.

The relationship of reading to critical thinking, though often contentious, is well established in the literature of reading comprehension and assessment, including its more recent incarnation, *reading for understanding* (Kintsch, 1998; Pearson & Hamm, 2005). Nearly every reading comprehension assessment blueprint in the past several decades has some variation of a cross of text types

(typically narrative vs. expository vs. persuasive) against Bloomian-derived critical thinking skill types (typically inference, analysis, synthesis, evaluation, explanation, application) (e.g., NAGB, 2005). The ongoing challenge for reading theorists has been to find a way to distinguish some “purified” construct of advanced reading from an equally “pure” construct focused on verbal reasoning/critical thinking/problem solving, recognizing that the latter could be assessed using non-language stimuli (e.g., matrix rotations), or logic problems that rely minimally on verbal understanding.

But we can ask ourselves, is not a scenario-based, scaffolded writing test as we have described in this chapter also a test of reading proficiency? Is it not also a test of critical-thinking skill? Put differently, do we not have considerable *de facto* evidence of reading and critical thinking proficiency when a writer produces a well-constructed essay or composition that cites evidence derived from foundational texts and articulates a well-thought-out position, claim, argument, interpretation, description, or explanation? Such a performance arguably provides evidence that an individual has the complete literacy package. There are other ways of assessing advanced reading comprehension and thinking skills that do not require a student to compose a written product (e.g., giving an oration, producing a multi-media or video, performing an experiment or other actions, selecting correct answers to questions on an exam), but perhaps permit individuals to express their understanding in a specified well-known genre—a sanctioned, conventionalized, and therefore accepted social literacy communication format—is also one of the cleanest and fairest ways to gather evidence of reading and thinking skill.

This view, which forces us to speak of tasks in a compound way, variously as reading-for-writing, writing-for-reading-comprehension, text-production-to-stimulate-reasoning, or reasoning-in-support-of-writing, creates significant measurement issues because it is incompatible with simple factorial models of skills and ability. The entire direction of literacy development is toward greater integration and mutual dependency among skills, so that (for instance) an expert writer, by employing a knowledge-transforming composition strategy, is far more dependent upon skilled reading (both for knowledge acquisition and self-evaluation) and upon verbal reasoning/critical thinking skills, than is a novice writer who relies almost exclusively upon knowledge-telling. Tracking the development of writing expertise thus requires a highly nuanced account, since expert writers are distinguished from novice writers not by the possession of any single skill, but by the ability to coordinate many skills strategically to achieve writing goals.

Conversely, building a complex mental representation or model of a text or the integration of several text (and non-text) sources, and connecting and

integrating those sources by updating one's existing knowledge of the domain, often demands iterations of writing (notes, outlines, explanations) and concomitant deliberation and reflection. We can flex this Rubik's Cube in exponential permutations to form myriad patterns, but ultimately we always have three-dimensional consequences.

NOTES

1. Outside the English Language Arts, such progressions are often called "Learning Progressions" (Duncan & Hmelo-Silver, 2009). We avoid this term here primarily to avoid confusion, since it is not clear in the current stage of research whether the developmental sequences observed with general literacy skills follow the same kinds of principles that may govern the learning of mathematical or scientific concepts. In our work for the CBAL program, we have ended up hypothesizing specific "Skills Foci" that correspond to disciplinary and academic genres and well-established literacy practices, and then proposing "hypothesized developmental sequences" that might underlie student learning and the kinds of curricular goals expressed in the standards. The assessments presented in this chapter depend in part on such an analysis being performed, since (for instance) we seek to include items that measure different levels of performance (and possibly different points in a developmental sequences) for targeted skills.
2. The other two focused on (i) arguing for a choice among alternatives and (ii) writing pieces of an informational pamphlet. These involved very little reading from extended texts and are therefore excluded from the present analysis.
3. Alpha (α) is a standard statistical measure that indicates the reliability or internal consistency of a test; high alpha is consistent with the hypothesis that all the items are measuring performance on a common underlying construct.
4. This level of agreement between reading and writing scores is about at the level seen between nationally normed, standardized tests of reading comprehension and writing.
5. For the purpose of this analysis, we do not distinguish between reading and thinking items, since the items designed to probe such skills as argumentation were pitched specifically to measure performance on argument tasks that combined reading and thinking.
6. A regression analysis creates a predicted score by assigning a weight to each of the predicting variables and adding the weighted variables together with a constant to produce a predicted score. The weights are adjusted to make the predicted score match the actual scores as closely as possible. R Square is a measure of the quality of the model. When R Square is 1 the dependent variable is fully predicted by the predictors; when it is 0, the dependent variable is not dependent on the predictors. A mid-range score

like .474 or .464 corresponds to a model that predicts about half of the variance—it works reasonably well, but with a significant amount of noise. In a regression analysis, R is the positive square root of R Square, and indicates the level of correlation between the values predicted by the model and the observed values of the dependent variable. Significance levels near 0 indicate that the results would be very unlikely if the null hypothesis were true. In Tables 4 and 5, the Beta value is important, since it is a standardized weight—it indicates the relative importance of each of the variables used in the regression in terms of a standard unit of measure. The higher the Beta, the more effect that variable has on the final score. In Tables 6 and 7, the R Square change is the most important figure, since it shows us how much of the prediction provided by the model can be produced by the other two lead-in tasks and how much is added by including the writing score.

REFERENCES

- Alamargot, D., & Chanquoy, L. (2001). Planning process. In G. Rijlaarsdam, D. Alamargot & L. Chanquoy (Eds.), *Through the models of writing* (Vol. 9, pp. 33-64). Dordrecht, Netherlands: Kluwer.
- 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.
- 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*. Berlin: Springer.
- Bereiter, C., & Scardamalia, M. (1987). *The psychology of written composition*. Hillsdale, NJ: Lawrence Erlbaum.
- Bloome, D. (1993). The social construction of intertextuality in classroom reading and writing lessons. *Reading Research Quarterly*, 28(4), 305-333.
- Coirier, P., Andriessen, J. E. B., & Chanquoy, L. (1999). From planning to translating: The specificity of argumentative writing. In P. Coirier & J. Andriessen (Eds.), *Foundations of argumentative text processing* (pp. 1-28). Amsterdam: Amsterdam University Press.
- Deane, P. (2011). *Writing assessment and cognition*. (ETS Research Report RR-11-14). Princeton, NJ: Educational Testing Service.
- Deane, P., Fowles, M., Baldwin, D., & Persky, H. (2011). *The CBAL summative writing assessment: A draft eighth-grade design*. (ETS Research Memorandum RM-11-01). Princeton, NJ: Educational Testing Service.
- Deane, P., Quinlan, T., & Kostin, I. (2011). *Automated scoring within a developmental, cognitive model of writing proficiency*. (ETS Research Report RR-11-16). Princeton, NJ: Educational Testing Service.

- Deane, P., Quinlan, T., Odendahl, N., Welsh, C., & Bivens-Tatum, J. (2008). *Cognitive models of writing: Writing proficiency as a complex integrated skill. CBAL literature review-writing* (ETS Research Report RR-08-55). Princeton, NJ: Educational Testing Service.
- Duncan, R. G., & Hmelo-Silver, C. E. (2009). Learning progressions: Aligning curriculum, instruction, and assessment. *Journal of Research in Science Teaching*, 46(6), 606-609.
- Felton, M., & Kuhn, D. (2001). The development of argumentative discourse skill. *Discourse Processes*, 32(2/3), 135-153.
- Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, 101(3), 371-395.
- 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.
- 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.
- Hillocks, G., Jr. (2002). *The testing trap*. New York: Teachers College Press.
- Hillocks, G., Jr. (2008). Writing in secondary schools. In C. Bazerman (Ed.), *Handbook of research on writing: History, society, school, individual, text* (pp. 311-330). Mahwah, NJ: Lawrence Earlbaum.
- Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing: An Interdisciplinary Journal*, 2, 127-160.
- Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17-64). Westport, CT: Praeger.
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge, UK: Cambridge University Press.
- Kuhn, D. (1999). A developmental model of critical thinking. *Educational Researcher*, 28(2), 16-46.
- Kuhn, D., & Udell, W. (2003). The development of argument skills. *Child Development*, 74(5), 1245-1260.
- 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.
- McKeough, A. (2007). Best narrative writing practices when teaching from a developmental framework. In S. Graham, C. MacArthur, & J. Fitzgerald (Eds.), *Best practices in writing instruction* (pp. 50-73). New York: Guilford.
- Mislevy, R. J., & Haertel, G. (2006). Implications for evidence-centered design for educational assessment. *Educational Measurement: Issues and Practice*, 25, 6-20.

- National Assessment Governing Board. (2005). *Specifications for the 2009 NAEP reading assessment*. Washington, DC: National Assessment Governing Board, American Institutes for Research.
- Nicolopoulou, A., & Bamberg, M. G. W. (1997). Children and narratives: Toward an interpretive and sociocultural approach. In M. Bamberg (Ed.), *Narrative development: Six approaches* (pp. 179-215). Mahwah, NJ: Lawrence Erlbaum.
- Nicolopoulou, A., Blum-Kulka, S., & Snow, C. E. (2002). Peer-group culture and narrative development. In S. Blum-Kulka & C. E. Snow (Eds.), *Talking to adults* (pp. 117-152). Mahwah, NJ: Lawrence Erlbaum.
- Olson, D. R. (1991). Literacy as a metalinguistic activity. In D. R. Olson & N. Torrance (Eds.), *Literacy and orality* (pp. 251-270). Cambridge, U.K.: Cambridge University Press.
- Olson, D. R., & Hildyard, A. (1985). *Literacy, language, and learning: The nature and consequences of reading and writing*. Cambridge, U.K.: Cambridge University Press.
- Pearson, P. D., & Hamm, D. N. (2005). The assessment of reading comprehension: A review of practices- past, present, and future. In S. G. Paris & S. A. Stahl (Eds.), *Children's reading comprehension and assessment* (pp. 17-30). Mahwah, NJ: Lawrence Earlbaum.
- Pellegrino, J. W. (2009, December). The design of an assessment system for the Race to the Top: A learning sciences perspective on issues of growth and measurement. Paper presented at the Exploratory Seminar: Measurement Challenges Within the Race to the Top Agenda. Abstract retrieved from <http://www.k12center.net/rsc/pdf/PellegrinoPresenterSession1.pdf>.
- Perin, D. (2009). Best practices in teaching writing to adolescents. In S. Graham, C. MacArthur, & J. Fitzgerald (Eds.), *Best Practices in writing instruction* (pp. 242-264). New York: Guilford.
- Scarborough, H. S. (2001). Connecting early language and literacy to later reading (dis)abilities: Evidence, theory, and practice. In S. Neuman & D. Dickinson (Eds.), *Handbook for research in early literacy* (Vol 1., pp. 97-110). New York: Guilford.
- Vellutino, F., Tunmer, W. E., & Jaccard, J. (2007). Components of reading ability: Multivariate evidence for a convergent skills model of reading development. *Scientific Studies of Reading*, 11(1), 3-32.
- Vygotsky, L., & Luria, A. R. (1994). Tool and symbol in child development. In R. van der Veer & J. Valsiner (Eds.), *The Vygotsky reader* (pp. 99-174). Oxford, UK: Blackwell.