

# Symposium

---

## Assessment

### **Closed Systems and Standardized Writing Tests**

*Chris M. Anson*

Imagine this odd but not entirely implausible scenario. Sam, an overzealous dad bent on establishing himself in his new neighborhood, trains his toddler daughter to act smart. Using a combination of fatherly coaching and operant conditioning, he gets little Alyssa to identify some framed prints of famous artwork displayed around his house—a Renoir here, a Paul Klee there. To the amusement of his neighbors as they sip margaritas and nosh on tortilla chips and canned salsa, every now and then Alyssa points to a painting and exclaims, “Jack-ssson! Poll-ICK!” or “Dats a REM-bant!”

In educational terms, Sam has tacitly established a set of learning outcomes for Alyssa, motivated by a desire to draw attention to his daughter, entertain his guests, and show off his family’s educational status: Alyssa will correctly identify and re-identify the artists of a dozen cheap prints of famous paintings. With enough promised trips to Dairy Queen and serious rehearsal, Alyssa ends up matching painter to painting nearly 100 percent of the time, the small margin of error coming from playful resistance, crankiness, or total distraction. For our purposes, Sam’s educational method, based on its results, is a huge success.

CCC 60:1 / SEPTEMBER 2008

Alyssa's performance, such as it is, represents a use of oral response to demonstrate the acquisition of knowledge—in simple terms, it's the measurable outcome of a certain kind of learning. No one, of course, believes that Alyssa really knows about art and could describe the difference between abstract expressionism and French impressionism, or analyze the relationship between pointillism and computer pixilation, or explain the effect of light and shadow in *The Night Watch*. It's just a cute game, and Alyssa seems all the cuter for playing it.

When we turn to the ways students are prepared to take national and state standardized tests in writing, however, too few stakeholders understand that training young people to produce a type of discourse found nowhere in the natural world in order to display command of simplistic, routinized textual habits and structures is, like Sam's project, a game, but with more important consequences. Instead, it is widely assumed that what students can do in such assessments accurately measures abilities they can draw on to succeed in other tasks, such as a lab report in a heat transfer experiment, a new-historical analysis based on conflicting primary documents, or a marketing proposal for managers in a retail firm. This belief in the generalizability of simplistic textual forms across contexts is wrongheaded. It deceives teachers, parents, educational administrators, and the general public into endorsing an unprincipled method of assessing students' writing abilities that drives an equally unprincipled pedagogy. Further, it provides a convenient way for politicians and others whose livelihoods depend on the improvement of "results"—on "gains" and the closing of "gaps"—to regress the cycle of teaching and assessing writing to the point at which high scores, like Alyssa's, are more easily attainable, harvesting political capital for themselves but bankrupting students' education in the process.

### ***Open and Closed Discursive Systems and the Problem of Transfer***

In the sense in which activity theorists and genre theorists describe it, writing takes place in an open system: as constantly evolving, contextually mediated, and contextually determined practices, influenced by social and institutional histories, conventions, and expectations (see Bazerman and Russell; Russell). As we know from descriptive research of students moving across the landscape of college courses (Curtis and Herrington; Haswell; Herrington and Curtis; McCarthy; Sternglass; Walvoord), or moving from successful college careers into unfamiliar professional settings as writers (Anson and Forsberg; Beaufort), proficiency in writing is not a matter of simply mapping a discrete set of

learned skills onto new tasks in unfamiliar contexts; it requires the kind of rhetorical, discursive, and textual flexibility and sensitivity that we hope our programs and courses provide.

In contrast, a closed system is one in which the activities admit little variation, are habituated over long periods of time, and are learned through repeated practice. If a person's total writing experience were limited to producing a 100-word, audience-less synopsis of their day, every day, using a chronological narrative structure and simply filling in the details, they would be working in a relatively closed system. Of course, an infinite number of events can instantiate the form, and the form itself admits at least some variation in its lexicon, the structure of its sentences, and so on. But if we assume that the writer's "ability" has been contained mainly or only to this textual world, we can easily imagine how difficult it would be to perform well outside it. Furthermore, without being able to "step back from their own decisions as writers and develop greater *metacognitive awareness* of the discourse practices they are entering, of how those practices differ, of how they as writer switch among them, and the strategies they use" (Flower 27–28), the writer will inappropriately replicate the habituated form. While it might be possible to explain to these narrative writers how "rhetorical actions are themselves situated in the larger circle of social and cultural assumptions" (28), the lack of experience writing in those larger circles would doom them to adaptive failure.

Attempts to get computers to work naturally with natural language tell a similar story. Applications of artificial intelligence to natural texts have had to operate within closed discursive systems because the relationship between world knowledge and discourse, in all its infinite genres and structures and socially determined features, is far too complex for computers to manipulate. For example, experts have programmed computers to produce brief versions of longer news reports (Radev and McKeown; Schank and Abelson) or to judge the effectiveness of someone else's summary of a longer text (see Anson). The fact that the computer systems can't "understand" what they are summarizing is obvious; the fact that they can't produce equally accurate and convincing summary versions of romance novels, wall-writings on Facebook, or contraindications on the tiny-print informational pages stuffed into the boxes of over-the-counter medications demonstrates their inability to transfer their automatized summary ability to new forms and contexts of discourse. Like programming computers to summarize news stories, we teach writing in a closed system when we reduce its rhetorical, linguistic, and performative complexity by creating a fixed genre with artificially stable features and then teach-

ing to it as if the goal were to repeat the writing behavior “automatically,” without regard to differing contexts and purposes.

To continue their literacy development in college, students need more than experience in a closed discursive system. As many writing scholars have argued, we must help students to understand that writing is situated in social practice and therefore context-dependent (see Lave and Wenger). Through diverse experiences, they begin to recognize that they are participating in conversations within a number of intellectual and pragmatic domains. “Since there are many such conversations that are important in our social and cultural world,” writes Applebee, “writing development may in turn be a matter of developing a voice in a wider array of conversations” (106). Such conversations take place in a variety of forms and media defined both by broad generic features, such as the IMRAD format of a report in a soil science class, and more idiosyncratic, localized expectations within an activity system, such as those constructed in a learning-based, invented dialogue between two classical thinkers in Philosophy 101. Although *continued* writing development involves acquiring, through exposure and experience, the ways of thinking and communicating that define these different communities, students are obviously better prepared to do so when they have had similar experiences earlier in their education. In such experiences, students learn the relationship between their social context and the various linguistic and rhetorical decisions they must make as they write and revise in order to be effective. In a single context, they may be repeating specific discursive patterns—for example, if they prepare a monthly Internet newsletter for a student organization—and therefore appear to be writing in a more closed system. But it is working across and between those contexts that gives them the experience of being in an open system, one that does not *predict* the way that future texts must be created in new and unfamiliar settings but provides the strategic knowledge to be discursively flexible—to know how to “read” a context in order to write in it (Anson and Forsberg).

The fact that writing successfully depends on local knowledge and experience does not, of course, mitigate the need for general abilities. As Mike Carter has pointed out in a compelling analysis of the tensions between cognitive theories of writing that posit universal, transferable skills and socially based theories in which writing is contextually situated, determined, and acquired, the idea of “expertise” involves applying global strategies to specific contexts where they are not yet fully experienced; but it is the nature of those global strategies that differentiates them from what students learn when they prepare for most standardized writing tests:

Competence . . . describes writers who have developed the writing skills that enable them to perform capably in a variety of writing domains. They have a firm possession of the relatively general writing strategies that give them a great deal of flexibility, for instance strategies of discourse analysis, revision, generating ideas, getting started, overcoming writer's block, determining and writing for audiences, etc. All of these are strategies that enable the writer to achieve some success in writing in a domain without extensive experience in that domain. (282)

Scholarship on the transfer of knowledge and skills across contexts is especially relevant to the advantages of working in open systems. Perkins and Salomon, for example, describe two important dimensions in transfer: one is the distance (or difference) between two contexts and the other is the nature of the two tasks. In a near-transfer situation, where one context is very similar to the other, transfer may not pose a major challenge. If someone has learned to use the Internet Explorer browser and then switches to Netscape, the distance between these contexts is not far, and the tasks are similar in scope and difficulty. Her knowledge of the Explorer functions, behaviors, and settings will transfer across the two browsers. But if she has learned to present a scripted 30-minute tour guide in a special exhibit of a museum and is then asked to create the script for a tour of a resort that various tour guides can all use, the distance between these contexts is greater, and the task, though somewhat related, may challenge her abilities. If she wants to be well prepared to do this kind of work, she will expect her teachers to vary her tasks and contexts and help her to develop strategies to be successful in all of them. More extensive preparation would involve some meta-level strategies for figuring out what questions to ask about the new task, such as what visitors will want to know or see, how to organize and arrange the information, or how to transform one kind of experience (answering questions during a live tour) to another (anticipating questions that might be asked during the resort tour). In writing, as Amy Devitt has pointed out in her analysis of genre transfer, "if the only genres the writer feels familiar with differ considerably from the needed genre, the antecedent genres will not likely serve the writer well. . . . Writers with fuller genre repertoires may have more baggage to carry, but they can move among different locations—with their different genres—more easily" (222).

Once a skill is habituated (like shifting gears in a car), it can be used in reflexive or "low-road" transfer (like shifting gears in a truck instead of a car). High-road or *mindful* transfer requires effortful abstraction and a search for connections between contexts (such as studying historical events and then applying them to current events in order to make connections and extend them

into critical questions about, for example, courses of action). Working across contexts requires a combination of reflexive and mindful transfer, or *adaptive expertise*—the ability to match new situations to previous experiences combined with the ability to abstract general problem-solving skills from previous experiences to apply in new situations (see Beaufort 180; Ingram). Experts' knowledge is organized by strategies that direct their thinking in multiple situations rather than by separate lists of facts that are relevant to individual situations (Perkins and Salomon 36).

In their review of research on transfer, Donahue and Wardle show that transfer across contexts is hampered when “material is taught in only one setting or context rather than in multiple contexts (Bjork and Richardson-Klavehn)”; and when “individuals learn procedures without learning underlying concepts, or when learning is overly contextualized, that is, when the classroom does not connect with everyday practices (Bransford) or details of the learned material are elaborated in one context and become attached to that context (Eich).” In contrast, transfer is facilitated when learners can explicitly abstract principles from a situation (Gick and Holyoak) and when they engage in self-reflection and mindfulness (Belmont, Butterfield, and Ferretti; Bransford; Langer; Bereiter). According to a review of twelve longitudinal studies of writing in college, “students’ pre-existing conceptions of writing from other contexts can prevent transfer; students who transfer writing ability successfully begin ‘seeing texts as accomplishing social actions’ and develop a ‘complex of activities’ rather than a set of generalizable skills; and students whose teachers help them deconstruct the genres of their field transfer writing knowledge/ability more effectively” (Rogers, quoted in Donahue and Wardle).

Good writing instruction, then, assumes that it is important for learners to experience a range of writing tasks, contexts, and purposes, and that it is better for them to gain adaptive expertise than a narrowly defined set of skills relevant to a specific, artificial genre (such as a test essay). Such instruction provides students with more time for sharing their writing and extending it into new domains, a wider range of curricular experiences, and a broader range of genres (Frater), and it “creates opportunities for students to be in different kinds of writing situations, where the relationships and agendas are varied” (Writing Study Group). As Carter puts it, these experiences help students “to learn some writing strategies that provide for a less restrictive approach to a writing task, a recognition that writing changes from domain to domain and the ability to write with some effectiveness in different domains” (282). Meet-

ing the need for both general, global skills and the local knowledge required to write effectively in a specific setting or community requires an instructional approach that Brown, et al. call “cognitive apprenticeship”: working in “authentic learning situations . . . that show how students learn general principles from working in specific domains” (qtd. in Carter, 283).

### ***Large-Scale Writing Assessment: Closed Systems***

For creators and administrators of large-scale standardized writing tests, assessing performance in an open discursive system presents serious challenges. Providing opportunities for students to write in different genres, to show their reasoning as they make certain rhetorical or stylistic decisions in a specific context, or to demonstrate their ability to incorporate actual (not excerpted or artificial) texts into their own original arguments or syntheses requires time to think, plan, and revise; space in which to display writing ability across a range of tasks and genres; and readers willing to take more than a few minutes to reach judgments on the (varied) results. Testing agencies are not willing to afford such opportunities because of cost and the loss of the controlled environment required to monitor and certify the work of thousands or tens of thousands of test-takers. Such rich, authentic assessment is *possible*, but for the machine-like industry of mass testing, it is highly *undesirable*.

Consider as emblematic of large-scale writing assessments the newly implemented SAT writing test (one of the highest-stakes writing exams in the country and perhaps the most carefully orchestrated because of the resources the College Board has at its disposal to create, administer, and monitor it). This test exists within a relatively closed discursive system because (1) it is rhetorically a-contextual: it has no purpose other than to test, and certainly none determined by the writer; (2) it has no functionally appropriate audience beyond completely anonymous scorers (and perhaps eventually machines); (3) the conditions under which it is composed (a timed, 25-minute period in a sterile, panopticon-like environment lacking any writing resources) are mirrored almost nowhere in the world of meaningful written discourse; (4) its assessment protocol necessarily narrows its structure, language, and other possible variations in responses to a prescribed form (generalized to the infamous five-paragraph theme); and (5) it provides no pragmatically or pedagogically useful feedback other than a score. As Perelman has put it, the text demanded of the SAT is a “completely artificial and unnatural piece of writing” (Jaschik).

Such a limited test of an ability as complex and multifaceted as writing is bad enough. It is not unlike assessing an ability as complex and multifaceted as teaching through a five-minute videotaped lecture on a “sprung” topic, with no time to prepare, delivered to a camera, and scored at lightning speed by anonymous, holistically trained raters who look for generalized features such as speaking stance and eye contact. Like the SAT writing test, such an assessment ignores most of the characteristics of “good teaching,” such as strategic planning; the artful orchestration of class sessions based on learning goals; insightfulness about specific students and their needs as learners; the translation of thoughtful and consciously considered instructional ideologies into pedagogical routines; innovation and creativity; spontaneity; ability to lead meaningful discussions that help students apply and scaffold their knowledge; evidence that students are developing and learning along multiple trajectories; and countless other characteristics best displayed in multiple ways, including reflective portfolio entries. In contrast, the one-shot video performance can’t possibly capture “good teaching” or accurately predict novice teachers’ future success in the classroom.

But the test itself is not the only or most serious problem. Far more important is the effect that such high-stakes standardized tests have on curriculum and instruction across the country. Whereas such tests are said by politicians and testmakers to spur greater rigor in the schools and improve students’ learning, ironically they appear to have the opposite effect. In a study of the relationship between high-stakes mandated tests and student achievement in seventeen states, Amrein and Berliner found *no significant effect* of standardized testing on student performance:

At the present time, there is no compelling evidence from a set of states with high-stakes testing policies that those policies result in transfer to the broader domains of knowledge and skill for which high-stakes test scores must be indicators. Because of this, the high-stakes tests being used today do not, as a general rule, appear valid as indicators of genuine learning, of the types of learning that approach the American ideal of what an educated person knows and can do. (N.p.)

Paradoxically, these findings appear to be causally related to the influence of the assessments on instruction. Hillocks’s analysis of the effects of state writing assessments in five states led him to the following conclusion:

When states establish writing assessments and determine certain categories of writing to test, they privilege the selected categories in the eyes of the schools

and teachers. When school scores are reported to a central authority and compared to the scores of other schools, the procedure of testing and reporting demands attention to what is taught. Teachers teach what is on the test and ignore what is not. (204)

Hillocks's empirical findings are daily confirmed and elaborated anecdotally across the United States. In one school district, administrators and teachers were aware of what research overwhelmingly says about the deleterious effects of direct, sustained formal grammar instruction on writing ability yet were faced with a standardized test "seeming to say that you have to go back to some old methods" (Cloud). In place of richer forms of instruction, they predicted a movement "toward teaching [students] to scribble hasty compositions in 25 or 30 minutes." Similarly, Michael Winerip recounts the experience of Becky Karnes, a high school teacher who recently completed a graduate-level writing course at Grand Valley State University that "opened her eyes" about how to teach writing in an open system where students write all sorts of different pieces, some evolving from others, some personal, some based on research. In her view, "writing formulas don't work; formulas don't let [students] think."

Karnes returned to her school in the fall eager to put all the ideas from her course into practice, but soon found that impossible.

There are three essays on the state test and we start prepping right at the start of the year. We have to teach to the state test [Michigan Educational Assessment Program, or MEAP]. . . . MEAP is not what writing is about, but it's what testing is about. And we know if we teach them the five-paragraph essay formula, they'll pass that test. There's a lot of pressure to do well on the MEAP. It makes the district seem good, helps real estate values.

Winerip points out that as a result of these pressures,

the five-paragraph essay has become the law of the land: introductory paragraph; three supporting paragraphs, each with its own topic sentence as well as three supporting ideas; and summary paragraph. Students lose points for writing a one-sentence paragraph. Many English teachers have developed a standard five-paragraph form with blanks to fill in.

Topic sentence: \_\_\_\_\_.

Literary example: \_\_\_\_\_.

Historical example: \_\_\_\_\_.

Current event: \_\_\_\_\_.

Concluding sentence: \_\_\_\_\_.

Nancy Patterson, who teaches the college course that Becky Karnes took, sees these formulaic writing experiences “spreading like kudzu, down to middle school, down to the fourth grade, down to the third grade. If you give kids the formula to write an essay, you’re taking away the very thinking a writer engages in . . . . [Students] are less apt to develop a writer’s thinking skills” (qtd. in Winerip).

These essential thinking skills are further subverted by attempts to outsmart or, as *LA Times* reporter Karin Klein puts it, “game” the test. The more narrowly circumscribed and reductive a task is, the more easily a learner can be coached, like Alyssa, to perform it. In an exposé, Klein describes how she persuaded the College Board and Pearson Educational Measurement to let her train to score essays for the new SAT. Among many observations from her experience, Klein found that the higher-scoring essays used multiple examples even if those added nothing to the argument; had at least one nonpersonal anecdote (especially from current events); and were more than one page (a finding statistically supported by Perelman). After her experience scoring, Klein says that if she had to prepare her own children to take the test, she would advise that they “prepackage some thinking. Get familiar with a couple of Greek myths or literary classics that would work for multiple themes. Prepare a few highly burnished words that can be applied to almost any situation. A prepared sentence or two wouldn’t hurt.”

### ***Toward Constructive Alignment***

Principled, authentic assessment is anchored to the learning goals and pedagogies of particular educational contexts. Although such goals can be successfully proposed on the national level in the form of generalized outcomes (not imposed tests), such as the *WPA Outcomes Statement for First-Year Composition* (WPA), the most meaningful assessments provide formative information to those closest to the learners whose abilities are being assessed, in the context of their own curriculum and educational outcomes. In this respect, we can consider the effects of high-stakes, large-scale, standardized assessment in terms of what John Biggs calls “constructive alignment,” which refers to the relationship between the intended goals and objectives of a course or curriculum and what is actually tested (Biggs, *Teaching*). Misalignment happens when, for example, we want students to learn to think critically about historical events and understand that history is inscribed and socially constructed, but we then test those students for the recall of factual historical

information. Eventually, students will simply memorize the facts (what Biggs calls “dealing with” the test), and the original goals and objectives are lost. On a national scale, such misalignment caused by reductivist testing programs weakens the entire educational system and produces students ill-prepared to do higher-level college work and enter an increasingly complex and demanding workforce.

If we agree that adaptive expertise is an important goal of preparation in writing, there is a serious lack of constructive alignment between this goal and tests that assess adherence to a formulaic essay structure. Lavelle and Guarino, for example, have shown that students write at varying positions along a scale from “deep” to “surface” thinking, based on prior experience and constructs of their writing situations. In the educational implications of their study, they suggest that “active, comprehensive revision is the defining element of deep writing”:

Our results support the contention that deep writing may be further differentiated from surface writing via consideration of the level of attentional focus (Biggs, [“Approaches”]). Both the elaborative and reflective-revision scales suggest a global focus involving attention to theme, voice, and audience as opposed to microconcerns such as grammar, rewording, and rules suggested in the low-self-efficacy, spontaneous-impulsive and procedural scales. (302)

To foster deep writing and more effective learning, they recommend that teachers “encourage global initiatives such as attention to genre, voice, audience and intentionality,” all aspects of writing conspicuously diminished in the a-contextual “dummy runs” (Britton et al.) represented by timed writing tests such as the SAT.

In the context of such educational goals, we must ask why anyone would want to impose high-stakes assessments that deny students varied, purposeful, and learner-centered experience as writers, especially through the effects such assessments have on the curricula and pedagogical practices that precede them. Scholars and critics have offered many reasons, of course, including money, accountability, comparability, and pressure from the public. In addition, we need to expose political motivations of the kind that have troubled early reading instruction, where federal guidelines and funding support a specific kind of pedagogy that is said to be most effective based on an inaccurate representation of empirical research. Further, this instruction is tied to a particular ideology of education that focuses on the acquisition of bottom-up skills,

the reinforcement of teacher-dominant, masculinist pedagogies, and the imposition of discipline and rigor (see NCTE Board of Directors)—all of which, as Jonathan Kozol has described in *The Shame of the Nation*, affect disproportionately the poorest schools that struggle the hardest in the face of relentless accountability.

In the urge to retain control over our pedagogies at all levels, teachers and scholars of writing need to be more public about the insufficiencies of the models being imposed on schooling through standardized assessment. We need to reveal to multiple audiences the regression of education for the purpose of monetary or political gain, or deceptive, “results-oriented” agendas. We also need far more research exploring the effects of high-stakes, large-scale assessment on students’ adaptive expertise—research such as Amrein and Berliner’s study that showed that “although states may demonstrate increases in scores on their own high-stakes tests, transfer of learning is not a typical outcome of their high-stakes testing policy” (n.p.).

Finally, we need much more attention to alternatives—to localized, contextually sensitive assessments that measure something more than the minimal skills required for reflexive transfer and encourage pedagogies as rich, varied, engaging, purposeful, and interactive as is the world of written discourse. Many successful examples of such localized programs can serve as models, but they are too often silenced by the political power and reach of the large-scale testing industry and the politicians and policy-makers at various levels who blindly adopt the industry’s products. For its part, even in the absence of regressive testing, instruction in writing must focus more strongly on helping students to develop those abilities that will serve them well in other contexts. In case studies of former students, for example, Elizabeth Wardle has found that although first-year composition students believe the abilities acquired there can transfer to other courses and beyond, they are often sharply critical of instruction that does not explicitly help them to see potential connections between contexts. Her study, combined with the results of other research on transfer, suggests to Wardle that “*meta-awareness about writing, language, and rhetorical strategies* in FYC may be the most important ability our courses can cultivate” (82). As more and more teachers in high school narrow students’ practice to a single, artificial, but nationally institutionalized test, it is not only important that we advocate for the cultivation of such meta-awareness and the diverse practices that nourish it, but that we also let the over-tilled fields of testing lie fallow and plow new ground based on more authentic forms of assessment.

Meanwhile, back at Sam's house, Alyssa's mom Angela has intervened. Alyssa is now doing her own toddleresque artwork, some for her own enjoyment, some as postcards for her grandparents and other relatives, some for the walls of her room, some using simple digital paint tools on the family computer. Angela has also been archiving Alyssa's work in several portfolios. Here and there, Alyssa has tried to create crayon or watercolor imitations of some of the framed prints in the house, but she's also doing her own representations of her long-tailed stuffed capuchin, her friend Jen, and the family beagle. Neighbors and relatives occasionally notice Alyssa's work and encourage her to create more. But the externalizing and display of Alyssa's developing knowledge and abilities are no longer what matters. Now Alyssa's connection to art is no longer about Sam, or Angela, or the neighbors. It is no longer about the imposition of a "test" of Alyssa's rote knowledge for others to measure their own sons and daughters against, or for Sam to accrue intellectual capital in his neighborhood.

Now it is about Alyssa.

### Works Cited

- Amrein, Audrey L., and David C. Berliner. "High-stakes Testing, Uncertainty, and Student Learning." *Education Policy Analysis Archives* 10.18 (2002). 10 March 2007 <<http://epaa.asu.edu/epaa/v10n18/>>.
- Anson, Chris M. "Can't Touch This: Reflections on the Servitude of Computers as Readers." *Machine Scoring of Student Essays: Truth and Consequences*. Ed. Patricia Freitag Ericsson and Richard Haswell. Logan: Utah State UP, 2006. 38–56.
- Anson, Chris M., and L. Lee Forsberg. "Moving beyond the Academic Community: Transitional Stages in Professional Writing." *Written Communication* 7 (1990): 200–31.
- Applebee, Arthur. "Alternative Models of Writing Development." *Perspectives on Writing Research, Theory, and Practice*. Ed. Roselmina Indrisano and James R. Squire. Newark, DE: International Reading Association, 2000. 90–110.
- Bazerman, Charles, and David Russell (Eds.). *Writing Selves/Writing Societies: Research from Activity Perspectives*. Fort Collins, CO: WAC Clearinghouse and Mind, Culture, and Activity, 2002.
- Beaufort, Anne. *Writing in the Real World: Making the Transition from School to Work*. New York: Teacher's College P, 1999.
- Belmont, J. M., Earl C. Butterfield, and R. P. Ferretti. "To Secure Transfer of Training Instruct Self-Management Skills." *How and How Much Can Intelligence Be Increased?* Ed. Douglas Detterman and Robert Sternberg. Norwood, NJ: Ablex, 1982. 147–54.
- Bereiter, Carl. "Situated Cognition and How to Overcome It." *Situated Cognition: Social, Semiotic, and Psychological Perspectives*. Ed. David Kirshner and

- James A. Whitson. Hillsdale, NJ: Erlbaum, 1997. 281–300.
- Biggs, John. *Teaching for Quality Learning at University*. 2nd ed. Philadelphia: Society for Research into Higher Education and Open University Press, 2003.
- Biggs, John. "Approaches to Learning and Essay Writing." *Learning Strategies and Learning Styles*. Ed. R. R. Schmenk. New York: Plenum, 1988. 185–228.
- Bjork, Robert A., and Alen Richardson-Klavehn. "On the Puzzling Relationship between Environment Context and Human Memory." *Current Issues in Cognitive Processes*. Ed. Chizoko Izawa. Hillsdale, NJ: Erlbaum, 1989. 313–34.
- Bransford, John. "Learning and Transfer." *How People Learn: Mind, Brain, Experience, and School*. Ed. John Bransford et al. Washington, DC: National Academy Press, 2000. 51–78.
- Britton, James N., Tony Burgess, Nancy Martin, Alex McLeod, and Harold Rosen. *The Development of Writing Abilities (11–18)*. London: Macmillan, 1975.
- Brown, John Seely, Allan Collins, and Paul Duguid. "Situated Cognition and the Culture of Learning." *Educational Researcher* 18 (Jan/Feb. 1989): 32–42.
- Carter, Michael. "The Idea of Expertise: An Exploration of Cognitive and Social Dimensions of Writing." *College Composition and Communication* 41.3 (1990): 265–286.
- Cloud, John. "Inside the New SAT." *Time Magazine*, Monday, 27 Oct. 2003. <<http://www.time.com/time/magazine/article/0,9171,1005973-2,00.html>>.
- Curtis, Marcia, and Anne Herrington. "Writing Development in the College Years: By Whose Definition?" *College Composition and Communication* 55.1 (2003): 69–90.
- Devitt, Amy. "Transferability and Genres." *The Locations of Composition*. Ed. Christopher J. Keller and Christian R. Weisser. Albany: SUNY Press, 2007. 215–227.
- Donahue, Christiane, and Elizabeth Wardle. "The Unexplored Assumption: Knowledge Transfer and Teaching Composition." *TeachingComp*. New York: McGraw-Hill Partners in Teaching Module, 2007. <<http://www.mhhe.com/socscience/english/tc/>>.
- Eich, Eric. "Context, Memory, and Integrated Item/Context Imagery." *Journal of Experimental Psychology* 11 (1985): 764–70.
- Flower, Linda. *The Construction of Negotiated Meaning: A Social Cognitive Theory of Writing*. Carbondale: Southern Illinois UP, 1994.
- Frater, Graham. *Bridges for Literacy*. London: Basic Skills Agency, 2002.
- Gick, Mary, and Keith J. Holyoak. "Schema Induction and Analogical Transfer." *Cognitive Psychology* 15 (1983): 1–38.
- Haswell, Richard H. *Gaining Ground in College Writing: Tales of Development and Interpretation*. Dallas: Southern Methodist UP, 1991.
- Herrington, Anne, and Marcia Curtis. *Persons in Progress: Four Stories of Writing and Personal Development in College*. Urbana, IL: National Council of Teachers of English, 2000.
- Hillocks, George, Jr. *The Testing Trap: How State Writing Assessments Control Learning*. New York: Teachers College P, 2002.
- Ingram, Laura. "An Analysis of the Problem of Writing Skill Transference and Its Implications for Professional Writing Instructors." M.A. Thesis, Dept. of English, North Carolina State University, June 2007.

- Jaschik, Scott. "Fooling the College Board." *Inside Higher Ed*, 26 March 2007. <<http://insidehighered.com/news/2007/03/26/writing>>.
- Klein, Karin. "How I Gamed the SAT." *Los Angeles Times*, 3 April 2005. <<http://www.latimes.com/news/opinion/sunday/commentary/la-op-sat3apr03,1,3742834.story>>.
- Kozol, Jonathan. *The Shame of the Nation: The Restoration of Apartheid Schooling in America*. New York: Crown, 2005.
- Langer, Ellen J. *Mindfulness*. Reading, MA: Addison-Wesley, 1989.
- Lave, Jeanne, and Etienne Wenger. *Situated Learning: Legitimate Peripheral Participation*. Cambridge: Cambridge UP, 1991.
- Lavelle, Ellen, and Anthony J. Guarino. "A Multidimensional Approach to Understanding College Writing Processes." *Educational Psychology* 23.3 (2003): 295-305.
- McCarthy, Lucille Parkinson. "Stranger in Strange Lands: A College Student Writing Across the Curriculum." *Research in the Teaching of English* 21.3 (1987): 233-65.
- NCTE Board of Directors and Executive Committee. "On the Reading First Initiative." NCTE Position Statement. Urbana, IL: National Council of Teachers of English, 2002. <<http://www.ncte.org/about/over/positions/category/read/107475.htm>>.
- Perelman, Les. "Teacher-Based, Technology Assisted Writing Assessments: Test to the Teaching." Paper presented at the Conference on College Composition and Communication, New York, 23 March 2007.
- Perkins, David N., and Gavriel Salomon. "Transfer of Learning." *International Encyclopedia of Education*. 2nd ed. Oxford: Pergamon, 1992. <<http://learnweb.harvard.edu/alps/thinking/docs/traencyn.htm>>.
- Radev, Dragomir R., and Kathleen R. McKeown. "Generating Natural Language Summaries from Multiple Online Sources." *Computational Linguistics* 24.3 (1998): 469-500.
- Rogers, Paul. "Methodological and Developmental Lessons from Longitudinal Studies of Writing Across the College Span." Working draft 2006. University of California at Santa Barbara Graduate School of Education.
- Russell, David. "Activity Theory and Its Implications for Writing Instruction." *Reconceiving Writing, Rethinking Writing Instruction*. Ed. Joseph Petraglia. Hillsdale, NJ: Erlbaum, 1995. 51-78.
- Schank, Roger C., and Robert P. Abelson. *Scripts, Plans, Goals, and Understanding*. Hillsdale, NJ: Erlbaum, 1977.
- Sternglass, Marilyn S. *Time to Know Them: A Longitudinal Study of Writing and Learning at the College Level*. Mahwah, NJ: Erlbaum, 1997.
- Walvoord, Barbara. *Thinking and Writing in College: A Naturalistic Study of Students in Four Disciplines*. Urbana, IL: National Council of Teachers of English, 1990.
- Wardle, Elizabeth. "Understanding 'Transfer' from FYC: Preliminary Results of a Longitudinal Study." *WPA: Writing Program Administration* 31.1-2 (2007): 65-85.
- Winerip, Michael. "Study Great Ideas, but Teach to the Test." *New York Times*, 13 July 2005. <<http://www.nytimes.com/2005/07/13/education/13education.html?ex=1278907200&en=3d7a79e4b3b9ea89&ei=5088&partner=rssnyt&emc=rss>>.

*WPA Outcomes Statement for First-Year Composition.* Council of Writing Program Administrators. Available at [www.wpacouncil.org](http://www.wpacouncil.org).

Writing Study Group of the NCTE Executive Committee. "NCTE Beliefs about the Teaching of Writing." Urbana, IL: National Council of Teachers of English, 2004. <<http://www.ncte.org/prog/writing/research/118876.htm>>

**Chris Anson** is University Distinguished Professor of English and Director of the Campus Writing and Speaking Program at North Carolina State University, where he helps faculty in nine colleges to use writing and speaking in the service of students' learning and improved communication. He has written and spoken widely on composition, literacy, and writing across the curriculum. His website can be found at <<http://www.home.earthlink.net/~theansons/Portcover.html>>.

## Information Illiteracy and Mass Market Writing Assessments

*Les Perelman*

Long before it was launched, the College Board hailed the new SAT writing essay as a major step in improving writing instruction in America (McGrath 24). Their literature continues to report that the new test is placing a new emphasis on writing skills in American High Schools (Mattern, Camara, and Kobrin). It and its sister College Board tests such as the AP Language and Composition Test are placing a new emphasis on writing, but, unfortunately, the emphasis is on composition techniques that, instead of fostering good writing and critical thinking, encourage students to embrace habits that produce mechanistic prose lacking any intellectual substance.

Using training samples and other sample essays all published by the College Board, I developed a cynical but effective formula for coaching three high school seniors retaking the SAT. I told them to follow the rigid structure of the five-paragraph essay, fill up both pages of the test booklet, include lots of detail even if it is made up or inaccurate, use lots of big words, especially substituting "plethora of" and "myriad number of" for "many," and to insert a famous quotation near the conclusion of the essay even if it is irrelevant to the rest of the essay.

My formula appears to work. All three students who followed it improved their raw scores on the essay section by at least 2 points out of the 12 possible.

Here is part of one of the student essays that was scored as a 5, out of 6 possible points, by each of the two College Board readers<sup>1</sup>:

A major reason why cooperation is a preference to competition is because competition induces civil struggle at a time of crisis while cooperation reduces tension. In the 1930's, American businesses were locked in a fierce economic competition with Russian merchants for fear that their communist philosophies would dominate American markets. As a result, American competition drove the country into an economic depression and the only way to pull them out of it was through civil cooperation. American president Franklin Delenor Roosevelt advocated for civil unity despite the communist threat of success by quoting "the only thing we need to fear is itself," which desdained competition as an alternative to cooperation for success. In the end, the American economy pulled out of the depression and succeeded communism.

The College Board's Scoring Guide states that an essay receiving a score of 5 "effectively develops a point of view on the issue and demonstrates strong critical thinking, generally using appropriate examples, reasons, and other evidence to support its position; is well organized and focused, demonstrating coherence and progression of ideas; and exhibits facility in the use of language, using appropriate vocabulary" ("How the Essay Is Scored").

The student who wrote this essay, however, knew it was badly written and that it did not exhibit mastery of anything. He told me that his current English teacher would have graded such an essay as a D. He was amazed that the College Board could be so easily fooled. He knew that his explanation of the Great Depression was wrong, but he could not remember the correct facts quickly so he just made some up. One could argue that with my small sample, the scores on the essays were just mistakes and do not represent the way most essays are graded. The College Board, however, does not admit that their scoring process can produce any errors. The College Board's "Essay Score Verification Policy Guidelines" state that their regular scoring process "already includes a built-in validation of the reader's scores," and, consequently, essays scores will not be rescored nor can they be appealed" (Hand Scoring).

I do not have space in this article to enumerate and discuss all of the bad habits reinforced by these tests and evident in the above passage, such as the consistent use of inappropriate vocabulary, meaningless sentences, and non sequiturs. See the NCTE Task Force on SAT and ACT Writing Tests, *The Impact of the SAT and ACT Timed Writing Tests*, for a more comprehensive discussion of the deficiencies in the new SAT Writing Test. Instead, I focus primarily on one deficiency, which I define as "information illiteracy." Infor-

mation illiteracy not only makes it more difficult for individuals to find information, but it makes it more difficult for them to differentiate between truth and falsehood. Indeed, information illiteracy often retards the desire to do so, reducing all assertions to the equal status of someone's opinion.

I derive the negative term "information illiteracy" from the positive form "information literacy," a term developed by librarians over the past twenty-five years as "the abilities to recognize when information is needed and to locate, evaluate, effectively use, and communicate information in its various formats" (State University of New York Council of Library Directors). In 2000, the Association of College and Research Libraries (a division of the American Library Association) published *Information Literacy Competency Standards for Higher Education*, which defines "an information literate individual" as someone who can:

- Determine the extent of information needed
  - Access the needed information effectively and efficiently
  - Evaluate information and its sources critically
  - Use information effectively to accomplish a specific purpose
  - Understand the economic, legal, and social issues surrounding the use of information, and access and use information ethically and legally.
- (2-3)

Information literacy is an inherent component of almost all writing and, as such, constitutes part of general literacy.

Information literacy permeates all phases of the writing process. Discovering what information is needed, how to obtain it, and then evaluating it properly are, in most writing situations, part of Invention (or *Prewriting*). First a writer needs to assess which facts constitute information. As James Kinneavy in *A Theory of Discourse* (93) notes, "To be informative a statement must enable us to relate the factual basis to some explicit or implicit system about which information is desired. In most instances, data are needed to help inform and refine a thesis, as well later to be used as evidence in supporting it." The writer of the SAT essay at the beginning of this article did none of these things.

In addition to using information effectively when writing informative and persuasive documents, writers need to learn the complex rules that govern

the appropriate use of information. In most venues, information is most effective as paraphrase, when it has been internalized by the writer and reformulated in his or her own words. Writers, however, also need to determine when direct quotations are more appropriate and how to identify information that, in a particular context, needs to be acknowledged and cited. When revising, a writer may discover that information is incomplete, inconclusive, or contradictory. Then, either the information has to be reinterpreted or additional information needs to be obtained. Finally, editing involves all the mechanics of proper citation and formatting as well as checking on the accuracy and ethics of how the information is reported.

### ***Data Smog***

Curiously, as our ability to access information online is increasing exponentially, our ability to use information effectively is decreasing dramatically. In pharmacology, a similar phenomenon is termed a “paradoxical effect”: a large dose of a drug produces the opposite effect of a smaller clinical dose. The paradoxical effect of too much information is almost the same as having no information at all. Daniel Shank defined the consequence of too much information as “data smog.” There is so much information coming at us, from so many directions and so quickly, that it is difficult if not impossible to differentiate good information from incomplete, biased, misleading, or just incorrect information. Moreover, the speed of the information coming at us produces the contradictory feelings that we need to make decisions quickly while being anxious because we know that there is still more information out there that we have not processed. Information literacy provides us with the skills and strategies necessary to cope with data smog.

Large-scale national tests are now incorporating what they define as document-based writing exercises, yet the specific formulations appear to test how comfortably students can function within data smog rather than how to emerge from it through information literacy. The SAT Writing Essay is the most egregious example. Graders of the essay are specifically instructed to reward the quantity of detailed and specific information but to ignore the quality of information. Blatantly false data is just as good as true information. Consequently, the student author of the passage above did not waste time trying to remember what were the real causes of the Great Depression. Instead, he brings in some notions from the Cold War to frame a detailed narrative that makes a strange sort of sense but has no connection with the real history. But within

the data smog-filled universe of the SAT, in which the rapid inscription of information is much more important than its accuracy, such a strategy is rewarded.

### ***Document-Based Essay Questions***

The College Board and the Educational Testing Service pay lip service to the notion of information literacy, and during the past several years they have incorporated Document-Based Essay Questions (DBQs) into the Advanced Placement Composition and History Tests. Completing one of these DBQs, however, has as much to do with displaying real information literacy as completing a child's eight-piece jigsaw puzzle has to do with painting a portrait.

The sample DBQ for the English Language and Composition Test given in the May 2007 and May 2008 Course Description (College Board, *English*), for instance, presents the reader with six sources, five texts and one chart, all of which are concerned with the topic of the relationship between American television and U.S. presidential elections.

The prompt reads:

This question requires you to synthesize a variety of sources into a coherent, well written essay. *Refer to the sources to support your position; avoid mere paraphrase or summary. Your argument should be central; the sources should support this argument.*

After introducing the topic of the influence of television in presidential elections since 1960, the prompt gives the following assignment:

*Assignment:* Read the following sources (including any introductory information) carefully. Then, in an essay that synthesizes at least three of the sources for support, take a position that defends, challenges, or qualifies the claim that television has had a positive impact on presidential elections.

The first source is a three-paragraph excerpt from an encyclopedia; the second source is a one-paragraph excerpt from what is listed as an online article but is actually an online version of an article from the Museum of Broadcast Communication's *Encyclopedia of Television*. Moreover, the excerpt contains a typo: the date on the website is "April 1993," not "April 1992," the date listed in the College Board *Guide*. This typo is not trivial. Because the incorrect date is before Clinton's victory, a student could argue that Clinton's association with MTV helped his election. Moreover, the date is part of the subheading for this section of a larger article, "April 20, 1993—Bill Clinton's

MTV Appearance”. Although the excerpt mentions that President Clinton was interacting with a “member of the MTV Generation,” there is no indication that this event took place on MTV. Moreover, there is no ellipsis indicating missing text. The third source is curious. It is listed as an article by Louis Menand, but it is actually a review by Menand of two new books. The excerpt consists of three paragraphs, the first two largely paraphrasing and quoting from the section on the Kennedy-Nixon debates in Theodore H. White’s *The Making of the President 1960*. The third paragraph consists of two sentences quoting “one commentator.” In the original review, these two sentences follow a detailed three-paragraph summary of a 1961 book by Daniel Boorstin, the “one commentator.” The last paragraph also omits Menand’s following sentences, which place Boorstin’s comments in historical perspective: “In 1961, this observation seemed alarming or alarmist. Today, no wisdom is more conventional.”

The last three sources consist of a chart of the Nielsen Ratings for the 1960 and the 1976 through 1996 presidential elections, and then short excerpts from a book on television and politics and from the memoirs of newscaster Ted Koppel. While the chart containing the Nielsen data lists the number of people watching presidential debates, it does not list the overall U.S. population or represent the audience watching presidential debates as a percentage of that number. Because the American population increased dramatically during this thirty-six-year period, percentages rather than raw numbers would provide a much better basis of comparison. In addition, although there were usually several presidential debates in each election, the table presents data for only one. Finally, the table displays Nielsen Rating numbers without any explanation of what these numbers represent.

These passages and data are to real sources of information what “sound bites” are to complete arguments. Rather than teaching students how to work with information, incomplete, misleading, and incomprehensible *word bites* such as those included in this exercise teach students to be comfortably lost in the data smog.

The DBQ essay prompt also encourages students to use information improperly. That the prompt calls for a minimum of three sources is probably not arbitrary. Three sources fit nicely into the three body paragraphs of a five-paragraph essay, the College Board’s favorite textual form. Even more disconcerting is that the prompt tells the students to avoid “mere paraphrase” but does not say anything about using too many quotations. Indeed, several of the sources have been edited down to be largely an author quoting other sources.

Yet we teach students to seek, when possible, the original source of a quotation rather than just citing a secondary source quoting the original. Students writing this essay in forty minutes are going to quote their sources quoting the original rather than *merely* paraphrasing the essential information. The assignment invites students to mindlessly patch together groups of quotations around a topic sentence to form each of the three paragraphs. Moreover, there is nothing in the prompt nor in most of the short introductory sentences for each of the passages that encourages students to evaluate the quality of the information. The scoring guide also contains nothing indicating that a critical assessment of the sources should be rewarded. Top-scoring essays, according to the guide, “effectively support their position by effectively synthesizing” and citing at least three of their sources. A footnote at the bottom of the pages elaborates: “For the purposes of scoring, synthesis refers to combining the sources to form a cohesive, supported argument and accurately citing all sources.” Thus this assignment reinforces in students the false notions that all information is equal, and that it is better to excessively quote sources rather than to internalize the concepts and restate them in an original form.

The College Board’s description of the AP Language and Composition course notes that one of the course goals is to help students “move beyond such programmatic responses as the five-paragraph essay” (12). Similarly, the course emphasizes researched argument papers that “use citations for substance rather than show, for dialogue rather than diatribe” (14). Unfortunately, the AP examination subverts these goals, and because the score on the exam is much more important to students for college admission, placement, and possible credit than the course grade, the demands of the test sometimes become the primary basis of the curriculum. Moreover, College Board publications sometimes make claims for their tests and products that are, at best, extremely dubious. A College Board report exploring the structure of the new SAT Writing Section to high school and college instructional practices and philosophy, for example, claims that having students write a 25-minute essay in response to a two- or three-sentence quotation is “using writing and reading as tools for critical thinking” (Milewski et al. 24).

The prompts in the AP English Language and Composition Test are representative of the similar attempts by testing organizations such as the College Board and the Educational Testing Service to incorporate the use of sources in essay assessments. The Document Based Essay Questions on the AP History Tests have been around longer, and they are more complex. The example

in the 2007–8 Guide for the U. S. Course (College Board, *AP History*), a question on the success of the New Deal in solving the problems of the Great Depression, includes nine sources, one of them a political cartoon and one a government advertisement for the new Social Security program. Still they are extremely formulaic. A student at MIT told me that her AP teacher regularly graded the tests and taught her students how to receive a high score on the DBQ. All of the sources except two will support one side of the argument. First, develop a thesis that will incorporate all the sources except for the two dissenting ones. Second, modify the thesis to acknowledge the two dissenting sources. Then write an essay that incorporates all the sources by including a paragraph toward the end that cites the counter-arguments of the two opposing sources. My student told me that every one in her class received scores of 4 or 5 on the test.

Students do need to learn to address opposing arguments and contradictory evidence. But in the real world they sometimes might not encounter opposing arguments or, if they do, they will rarely encounter exactly two opposing pieces of evidence. Also, sometimes students appropriately ignore a specific source if it is irrelevant to their argument. Moreover, while sources of information for any topic are always finite, there are almost always potentially many more than six to eight sources. The limited and predefined corpus of two- to three-paragraph word bites that comprise these assignments fits the formulaic five-paragraph essay; they have, however, little resemblance to real world information.

The SAT and AP tests are not the only high-stakes tests that attempt to access a student's ability to judge and use information. The Collegiate Learning Assessment (CLA) is an institutional assessment instrument jointly developed by Council for Aid to Education (CAE), Rand Corporation, and the Educational Testing Service (ETS). The instrument is taken by freshmen and seniors to measure the "value added" by a university's curriculum (Council for Aid to Education, CLA). Student scores are aggregated because the instrument is designed to measure institutional rather than individual student performance. The CLA claims to measure critical thinking, analytic reasoning, and problem solving through three written tasks, a *performance task*, a *make-an-argument prompt*, and an *analyze an argument prompt*. The CLA claims that these three measures assess, among other cognitive abilities, the ability to evaluate evidence; the ability to "synthesize data and information; and the ability to consider alternate interpretations and viewpoints," or as it is framed in the

CLA document, “How well does the student consider other options and acknowledge that their [sic] answer is not the only perspective. . . ?” (Council for Aid to Education, “Critical Thinking”).

The *make-an-argument prompt* closely resembles the argumentative prompts on the SAT Writing Essay. The sample prompt asks students to construct an argument responding to an assertion that it is reasonable for public figures to lose some of their privacy. It is again asking the student to construct an impromptu argument on a topic that he or she may never have thought about. The *analyze an argument prompt* is virtually identical to a standard writing prompt on ETS’s Graduate Record Examination, as well as on the Graduate Management Admissions Test, which used to be designed and administered by ETS but is now the product of Pearson Education.<sup>2</sup> The *performance task* is the writing exercise that resembles the DBQs of the Advanced Placement Test. In the sample prompt in the CLA brochure and on their website, the student is asked to assume the role of an assistant to the president of a high-tech company. A sales manager has recommended that the company buy a particular plane to transport sales personnel to and from customers, but before the plane was purchased there was an accident involving that particular model. The student is provided with the following documents: 1) a newspaper report of the accident; 2) an FAA report on “in-flight breakups in single engine planes;” 3) emails from the president to you and from the sales manager to the president; 4) charts displaying the performance of this particular line of planes; and 5) an article from a magazine for amateur pilots comparing this plane to others in its class.

Because the sample prompt just names the documents but does not display them, it is difficult to assess how much information is given to students. Still, given that students are expected to read the entire prompt and write the essay in ninety minutes, it is difficult to believe that these passages are substantial. In addition, given that in the scenario, the writer of the recommendation might be flying in the small plane, he or she, in a real situation, would probably want to check other sources of information before writing the recommendation. Again, students are given a six- or eight-piece puzzle, and we are told that mastering it will tell us how well they can navigate the vast sea of information that surrounds us. These tests do not encourage students to learn how to obtain, assess, and use information appropriately. These tests teach them to formulaically manipulate prepackaged information bites.

In a 1996 article, Jeremy J. Shapiro and Shelley K. Hughes in “Information Literacy as a Liberal Art” argue persuasively that information literacy is now

an essential liberal art. Using Enlightenment, medieval, and classical notions of what constitutes a “liberal art,” they demonstrate that in the twenty-first century, information literacy is a necessary skill for a free (the *libre* in liberal) person to participate in civic life. None of the assessments I have discussed encourages this liberal art of information literacy and in several key aspects, they impede it.

Even when large testing organizations try to address the issue of information literacy directly, they end up focusing on largely irrelevant elements that are easy to assess. In 2001, the Educational Testing Service “convened an international panel to study the growing importance of existing and emerging Information and Communication Technologies (ICT) and their relationship to literacy” (*Digital Transformation* iii). Curiously, although the ostensible mission of the group was to assess the effect of technologies on information literacy and literacy itself, the group contained no one with expertise in either area. Indeed, although the panel had representation from Australia, Brazil, Canada, and France, in addition to members from the United States and was chaired by a professor of communications, the panel had neither a specialist in writing and literacy nor an academic librarian among its members. Of course, the result of the panel’s report, *Digital Transformation: A Framework for ICT Literacy*, has resulted in a new and aggressively marketed test, the *iSkills* Assessment. As might be expected, the focus of this test is on mechanically using technology rather than on the complex components of literacy. Some of the tasks on the test include filling out an organization page, selecting (through multiple choice) elements of a Web page, selecting (again through multiple choice) lines for a PowerPoint presentation, downloading a video file, and managing electronic files on a computer. Some sections of the advanced test have students use a very limited bibliographic search engine and categorize bibliographic items as “useful,” “not authoritative,” “problematically biased,” or “not current” based only on titles and two-line descriptions. Like all the previously mentioned tests, the net effect of the *iSkills* Assessment is to trivialize the notion of information literacy while assuring various stakeholders that this important skill is being addressed.

Why do these tests have these effects? It is ridiculous to think that all of these sponsoring groups and assessment are engaged in a giant cabal to make American students stupid. No, the tests have the form they do because all of the stakeholders in the development and administration of mass-market writing tests place the most value on economic efficiency and scoring reliability. Compared to multiple-choice tests, essay tests are extremely expensive be-

cause each test carries a substantial marginal cost. Although the initial capital expenditure in developing multiple-choice test items may be high, once they are developed, the cost of administration per student is minuscule. In contrast, in most holistic readings, each student essay is read by a minimum of two readers. With mass market tests, such as the ones I have been discussing, that are given to hundreds of thousands of students per year, the cost of grading essays constitutes a substantial part of the test's overall cost. Test providers, then, whether they are for-profit companies or large non-profit entities such as the Educational Testing Service, seek essay prompts that will minimize the time needed by a reader to score the essay while simultaneously ensuring inter-reader reliability that can be used as a self-referential indicator of validity. Evaluating different paraphrases of a source, for example, is much more difficult and time consuming, and has more potential for disagreement than the mechanical counting of quotations. Indeed, in general, quantitative features, such as the frequency of infrequently used words like *plethora*, can be graded much more quickly and reliably by human readers than qualitative features, such as creativity in accurately rephrasing an idea. (The same bias holds for essays scored by machine, but that is the topic of another article.) The design of all these Document-Based Essay Questions is largely informed by the desire to have readers make reliable decisions on each student's score very quickly.

Alternate models are often rejected because they are not "practical," that is, the costs to score them reliably are considered prohibitive. Yet, it is time to question such assumptions. Although it is probably impossible to construct a large-scale assessment that covers all aspects of information literacy, except the use of portfolios, our own experience here at MIT and those of our partner institutions in using an online essay assessment has demonstrated that the Internet makes it possible to develop assessments that involve several important aspects of the process of evaluating, selecting, and using information.

At MIT, one of the two essays in our Online Freshman Essay Evaluation consists of an analytic summary essay of four to eight complete articles on a single topic. Students have to first decide which information is important and then synthesize the information from the articles into coherent summary paragraphs. The prompt instructs them to paraphrase their sources rather than quote them. Although this exercise is not a perfect exercise in information literacy, it reinforces some key elements. Students have to read lengthy texts and identify the important information in each one. They have to then identify the key topics and construct the "conversation" among the various texts.

Although we do not ask them to evaluate the quality of information, they have to use the information to construct an essay that resembles an academic literature review. Our assignment grew out of a slightly different but similar essay prompt developed by Irvin Peckham at Louisiana State University that also employs the online essay evaluation service iMOAT.

These online assessments seem to be perfect vehicles for low stakes assessments such as the CLA, although they may not be appropriate in their current form for high-stakes tests like the SAT or AP. In any case, these successful assessments at MIT and LSU might serve as prototypes for new and various kinds of tests. This development, however, should not be undertaken by the large non-profit and private testing organizations. They will always value reliability and economic efficiency over all other factors. Instead, these assessments should be developed by stakeholders of information literacy, writing teachers and librarians. If at all possible, writing teachers should include both K-12 and college instructors, and the National Writing Project, with its nearly 200 local sites, provides not only a viable model but also possible participants. The publications of the American Library Association and the Association of College and Research Librarians indicate that they would be eager to collaborate in an endeavor to design assessments that would reinforce the acquisition of information literacy and real critical thinking, instruments that would reinforce learning goals rather than subvert them and help our students become smarter rather than dumbing them down with word bites.

### **Acknowledgment**

*I want to thank Jay K. Lucker, Ann Wolpert, Director of MIT Libraries, and Steven Gass, associate director for public services, MIT Libraries, for educating me about information literacy and for reviewing this manuscript. I also want to thank my colleague Dr. Suzanne Lane for her excellent suggestions. As always, she has been an excellent writing teacher.*

### **Notes**

1. The prompt for the essay was:

Think carefully about the issue presented in the following excerpt and the assignment below.

While some people promote competition as the only way to achieve success, others emphasize the power of cooperation. Intense rivalry at work or play or engaging in competition involving ideas or skills may indeed drive people either to avoid

failure or to achieve important victories. In a complex world, however, cooperation is much more likely to produce significant, lasting accomplishments.

**Assignment:**

Do people achieve more success by cooperation than by competition? Plan and write an essay in which you develop your point of view on this issue. Support your position with reasoning and examples taken from your reading, studies, experience, or observations.

2. The sample *analyze an argument* prompt for the CLA has also appeared as a sample prompt on the Graduate Record Examination. For the text and an example of a high-scoring response see <<http://www.ets.org/Media/Tests/GRE/pdf/awintro.pdf>>.

**Works Cited**

- Association of College and Research Libraries. *Information Literacy Competency Standards for Higher Education*. Chicago: American Library Association, 2000.
- College Board. "AP English: English Language and Composition; English Literature and Composition: Course Description." May 2007, May 2008. 7 July 2007 <[http://www.collegeboard.com/student/testing/ap/sub\\_englang.html?englang](http://www.collegeboard.com/student/testing/ap/sub_englang.html?englang)>.
- . "AP United States History: Course Description." May 2007, May 2008. 9 July 2007 <[http://www.collegeboard.com/student/testing/ap/sub\\_ushist.html?ushist](http://www.collegeboard.com/student/testing/ap/sub_ushist.html?ushist)>.
- . "Hand Scoring/Requests for Score Verification: Multiple-Choice Hand Score Verification." 8 February 2008 <<http://www.collegeboard.com/student/testing/sat/scores/sending/handscore.html>>.
- . "How the Essay Is Scored: Scoring Guide." 16 March 2008 <[http://www.collegeboard.com/student/testing/sat/about/sat/essay\\_scoring.html](http://www.collegeboard.com/student/testing/sat/about/sat/essay_scoring.html)>.
- Council for Aid to Education. *Collegiate Learning Assessment*. New York: Council for Aid to Education, 2007.
- . "Collegiate Learning Assessment (CLA): Critical Thinking, Analytic Reasoning, Problem Solving, and Writing Skills—Definitions and Scoring Criteria." 9 July 2007 <[http://www.cae.org/content/pdf/CLA\\_Scoring%20Criteria.pdf](http://www.cae.org/content/pdf/CLA_Scoring%20Criteria.pdf)>.
- Educational Testing Service, International ICT Literacy Panel. *iSkills: Digital Transformation: A Framework for ICT Literacy*. Princeton, NJ: Educational Testing Service. <[http://www.ets.org/Media/Tests/Information\\_and\\_Communication\\_Technology\\_Literacy/ictreport.pdf](http://www.ets.org/Media/Tests/Information_and_Communication_Technology_Literacy/ictreport.pdf)>.
- Kineavy, James L. *A Theory of Discourse*. Englewood Cliffs, NJ: Prentice-Hall, 1971.
- Mattern, Krista, Wayne Camara, and Jennifer L. Kobrin. *SAT Writing: An Overview of Research and Psychometrics to Date*. Report RN-32. New York: College Board, 2007.

- McGrath, Charles. "The New SAT: Writing to the Test." *New York Times* 7 Nov. 2004: sec. 4A, 24.
- Milewski, Glenn B., et al. *A Survey to Evaluate the Alignment of the New SAT Writing and Critical Reading Sections to Curricula and Instructional Practices*. Research Report 2005-1. New York: College Board, 2005.
- NCTE Task Force on SAT and ACT Writing Tests. *The Impact of the SAT and ACT Timed Writing Tests: Report from the NCTE Task Force on SAT and ACT Writing Tests*. Urbana, IL: National Council of Teachers of English, 2005. <[http://www.ncte.org/library/files/About\\_NCTE/Press\\_Center/SAT/SAT-ACT-tf-report.pdf](http://www.ncte.org/library/files/About_NCTE/Press_Center/SAT/SAT-ACT-tf-report.pdf)>.
- Peckham, Irvin. "Turning Placement into Practice." *Writing Program Administration* 29 (2006): 65–84.
- Shank, Daniel. *Data Smog: Surviving the Information Glut*. New York: HarperCollins, 1997.
- Shapiro, Jeremy J., and Shelley K. Hughes. "Information Literacy as a Liberal Art: Enlightenment Proposals for a New Curriculum." *Educom Review* 31.2 (1996). 5 July 2007 <<http://www.educause.edu/pub/er/review/reviewarticles/31231.html>>.
- State University of New York Council of Library Directors. "Information Literacy Initiative: Final Report." 30 Sept. 1997. 5 July 2008 <<http://www.sunyconnect.suny.edu/ili/final.htm>>.

**Les Perelman** is Director of Writing Across the Curriculum in the Program in Writing and Humanistic Studies at the Massachusetts Institute of Technology, where he has also served as an associate dean in the Office of the Dean of Undergraduate Education. He is currently working on several articles on the machine scoring of writing and a book-length study of mass market writing assessments.

## Genre, Testing, and the Constructed Realities of Student Achievement

*Mya Poe*

The passage of the No Child Left Behind (NCLB) legislation in 2002 marked a milestone in American educational testing. For the first time in the United States, the federal government mandated an educational testing policy that dictated how federal educational testing would be conducted, who would be tested, when they would be tested, and how resources would be distributed based upon test results. Although officially NCLB was a reauthorization of the Elementary and Secondary Education Act of 1965, the NCLB legislation also added a number of components to the original legislation, including mandates about teacher quality, school choice, and educational testing. It is these components of NCLB, especially the legal mandates regarding testing, that are es-

pecially controversial. Briefly, NCLB requires testing of reading and math at grades 3–8 (science was added this past year). Schools must demonstrate “adequate yearly progress” on state standardized achievement tests. Schools receiving Title I funding may lose that funding if they do not demonstrate adequate yearly progress. Given these new requirements on public education, the effects of NCLB have been notable. NCLB *has* raised student test scores albeit at the expense of narrowing the school curriculum, unfairly burdening poor school systems, and destabilizing public education (Meier and Wood; Jennings and Rentner).

This year the Bush administration seeks to reauthorize NCLB and expand the legislation to include high school students (Spellings). In this commentary, I look more closely at NCLB, drawing upon Bakhtinian genre theory and specifically the concept of genre ecologies in my analysis. Broadly, Bakhtinian genre theory shows how official genres make certain behaviors or viewpoints appear normal. More specifically, a dominant genre such as NCLB legislation mediates social practice and can spawn entire ecologies of supporting genres that further mediate those social practices. Through this ecology of genres a “legend” of a testing center evolves (Giltrow). The successful reauthorization of NCLB suggests that this center has sufficiently held to expand federal oversight of educational testing to include secondary education, but it is a center that may not hold.

### ***Creating the Center***

NCLB mediates testing activity on three levels: (1) NCLB establishes norms for acceptable behavior, naturalizing federal control over testing, linking federal funds to test scores, and instantiating the theory of “yearly adequate progress”; (2) NCLB establishes legal and economic consequences for failure to comply with the law; and (3) NCLB creates institutional “rules of the game,” meaning that it sets forth constraints on state and local government actions—for example, it dictates whom states will test, when they will be tested, what will be tested, what counts as a suitable test, and who will do the testing.

In establishing norms for acceptable behavior, NCLB has reconfigured the relationship of the federal government to state government in terms of educational testing. While previously states were required to report student performance on assessments, the federal government “left the responsibility of managing public education to the individual states” and did not sanction schools for students’ test performance (PBS Online). Today, federal control over testing is normative practice. Or as Margaret Spellings writes in *Building on*

*Results: Blueprint for Strengthening the No Child Left Behind Act*, “In five short years, the *No Child Left Behind Act* has evolved from an idea to a law to a way of life” (3). Given this new “way of life,” NCLB sets a precedent for future activity. It becomes “natural” for the federal government to exercise control over all educational testing. The clearest demonstration of this growth phenomenon is found in Secretary Spellings’ stated goal of expanding NCLB testing through high school. If the federal government exercises control over elementary educational testing, it seems only natural that it should also exercise control over secondary school testing. In doing so, it also becomes normative to link funding to test scores at the secondary school level and demand “adequate yearly progress” for high school students.

Beyond establishing norms for acceptable behavior, NCLB includes penalties for failure to comply with the law. In the case of NCLB, the Department of Education may withhold funds from states until the secretary of education approves the state’s assessment plan. The Department of Education may also sanction schools if students’ test scores do not show “adequate yearly progress.” Although *all* schools are expected to show “adequate yearly progress,” *only* Title I schools are subject to federal sanctions for not showing progress on state tests. This policy has been particularly potent at mediating the activity of many public schools. Title I funding is given to districts that serve socioeconomically disadvantaged students. In many states, virtually all school districts receive Title I funding. In a final measure to mediate the activity of state educational testing, NCLB disperses the burden of policing local enactment of the law on state departments of education:

Each State accountability system shall . . . include sanctions and rewards, such as bonuses and recognition, the State will hold local educational agencies and public elementary schools and secondary schools accountable for student achievement and for ensuring that they make adequate yearly progress in accordance with the State’s definition under subparagraphs (B) and (C) (1446).

In the end, NCLB diffuses responsibility for maintaining testing activity across multiple levels of government. However, it reserves funding decisions solely to the federal government and leaves the determination of “acceptable behavior” far removed from local school districts, students, and teachers.

In setting forth the rules of the game, NCLB not only assigns roles to various actors, but it also determines how the assessment game will be played—in this case, what “tools” constitute acceptable forms of testing. First, NCLB assigns stated educational agencies as the nexus of state activity. The state de-

partment of education works in “consultation” with other “lesser” players to design individual state assessments:

For any State desiring to receive a grant under [Title I], the State educational agency shall submit to the Secretary a plan, developed by the State educational agency, in consultation with local educational agencies, teachers, principals, pupil services personnel, administrators . . . other staff, and parents, that satisfies the requirements of this section. (No Child Left Behind 1444)

Second, NCLB specifies what tools will be used in the assessment game. Although NCLB does not specify a particular test that states must use, it does stipulate that assessments must be based on “scientifically based research,” a term that is found more than 100 times in the legislation and defined as follows (Hess and Petrilli). NCLB delineates at length what constitutes “scientifically based research”:

(37) SCIENTIFICALLY BASED RESEARCH.—The term ‘scientifically based research’—“(A) means research that involves the application of rigorous, systematic, and objective procedures to obtain reliable and valid knowledge relevant to education activities and programs; and “(B) includes research that—“(i) employs systematic, empirical methods that draw on observation or experiment; “(ii) *involves rigorous data analyses that are adequate to test the stated hypotheses and justify the general conclusions drawn*; “(iii) relies on measurements or observational methods that provide reliable and valid data across evaluators and observers, across multiple measurements and observations, and across studies by the same or different investigators; “(iv) *is evaluated using experimental or quasi-experimental designs* in which individuals, entities, programs, or activities are assigned to different conditions and *with appropriate controls* to evaluate the effects of the condition of interest, with a preference for random-assignment experiments, or other designs to the extent that those designs contain within-condition or across-condition controls; “(v) ensures that *experimental studies are presented in sufficient detail and clarity to allow for replication* or, at a minimum, offer the opportunity to build systematically on their findings; and “(vi) has been accepted by a peer-reviewed journal or approved by a panel of independent experts through a comparably rigorous, objective, and scientific review (1551, emphasis added)

By limiting the official assessment tools to include those that are hypothesis-driven, that are based on experimental or quasi-experimental research, that include controls, and that are repeatable, NCLB sanctions tools that “stand outside the daily life of the classroom but are intended to hold classrooms, teachers, and schools accountable for results” (Bazerman 429). “Scientifically

based research” becomes a rule of the game, eliminating other forms of research such as qualitative studies and ultimately limiting what we know about assessment. As Dan Liston, Jennie Whitcomb, and Hilda Borko write in “NCLB and Scientifically-Based Research: Opportunities Lost and Found”:

The federal government’s legislation of a particular method has a dampening effect on research in education writ large, and by extension, teacher education. Not only does the law limit the possibilities for future federally funded research, but also and equally important, it narrows the set of existing research available to inform policy decisions that guide k–12 and teacher education practices (4).

### ***Ecologies of Control***

Clay Spinuzzi writes of genre ecologies: “A given genre mediates an activity, but it does not do so alone; it works in conjunction with the entire ecology of genres available” (“Describing” 6). Indeed, NCLB has spawned an entire ecology of supporting genres. Through this ecology of supporting genres, NCLB also stabilizes particular viewpoints or ideologies in “official” genred forms. In Bakhtinian terms, “official” genres act as *centripetal* forces, making certain activities appear normal or “natural.” Legal genres are especially efficient and effective means to maintain a particular worldview because they circulate in multiple activity systems with authority (e.g., state and local governments, unions, and court systems). This proliferation of genres via “center,” “official” genres provides a basis upon which subsequent policies and reports can build, thereby creating a unified sense of a testing “center.” Each of those genres comment on each other’s “officialness,” validating and naturalizing certain behaviors, the viewpoints of certain members of a community, and certain assessment tools (Giltrow).

One example of how genre ecologies mediate activity by pointing to a center is evidenced in Massachusetts. Here the NCLB legislation was taken up by the state department of education to advance the state’s use of the Massachusetts Comprehensive Assessment System (MCAS). MCAS then became the “official” test genre by which the Massachusetts Department of Elementary and Secondary Education could “align [the] state standards with college and work ready standards developed by Achieve and college-ready standards from ACT and the SAT” (Massachusetts Department of Elementary and Secondary Education). In other words, NCLB gave the Massachusetts department of education the authority to use its own test as a means to promote the state’s interest in college admissions and placement testing (Pollard). Although the various

genres spawned by NCLB in Massachusetts all have different individual purposes, they all work as a collective achievement in performing the “official” center of NCLB.

The ecologies of genres stemming from NCLB not only reinforce the authority of NCLB, but they also ignore or deny the social conflicts that led to the creation of NCLB. Assessment is often a reflection of broader social tensions in a culture, ranging from immigration, national security, segregation and social stratification, affirmative action, and privatization of education (Gonzalez; Trimbur; Kozol; Kohn). Assessment policies enact these social tensions and reproduce certain social arrangements (Inoue). By noticing who is singled out for surveillance in NCLB, we can find traces of those social tensions. Specifically, we should look for the ways that NCLB requires that test data be disaggregated. Currently NCLB requires that test data be disaggregated by the following categories:

race, ethnicity, gender, disability status, migrant status, English proficiency, and status as economically disadvantaged (1457).

Although NCLB names other groups of students in its Statement of Purpose—“neglected or delinquent children, and young children in need of reading assistance”—these other groups are erased when state test data are reported (1440). NCLB also does not require states to report intergroup variations or make connections across categories such as race, socioeconomic status, and geographic location. The collapsing of intergroup differentiation allows for the monolithic construction of singular test-taking identities (e.g., “Latino test scores”), which, in turn, allows for the reproduction of educational stereotypes along race, gender, and socioeconomic lines. Although NCLB promises equality, fairness, and high expectations for all students, it can be argued that this incongruity between the stated egalitarian objective of NCLB and the way that test results are required to be reported in state report cards means that “some gaps count more than others” in managing social tensions (Karp). There is a tacit understanding that it is African American, Latino, and Native American students who need “higher expectations” if they are ever to “catch up” to white and Asian students. Certainly, reporting racial disparities in educational performance can reveal inequalities in education (as well as health care, housing, and so on), and race-based reporting remains important evidence in civil rights lawsuits. However, when these are the *only* ways of reporting test data, they reinforce public fears about the failings of African American, Latino, and Na-

tive American students. Such are the “cascading discursive consequences” of genre ecologies where social, economic, and public resource inequality is thereby reduced from a complex social phenomenon with multiple causes and multiple potential resolutions to a contained entity called “the achievement gap,” which can be righted by the careful scrutiny of “student achievement”(Russell, personal).

### ***The Center Doesn't Have to Hold***

Any analysis of NCLB, however, is always partial unless we realize that ecologies of testing include both forces working to unify as well as forces working to disrupt any perceived unity. In Bakhtinian terms, *centripetal* forces work inwardly to standardize a system, while *centrifugal* forces work outwardly toward diversity and heterogeneity. In the case of NCLB, there are multiple possible “systematic destabilizations” to the law’s stabilizing tendencies (Spinuzzi *Tracing* 49).

Legal challenges chip away at portions of NCLB, undermining the authorial “super-structure” of the genre. For example, the National Education Association along with districts in three states has charged that NCLB is an unfunded mandate. The NCLB legislation, however, specifically prohibits states to “spend any funds or incur any costs not paid for under this Act” (1983). The National Education Association lawsuit has been successfully held up in the Sixth Circuit Court of Appeals and represents the kind of legal attack that can undermine NCLB. By chipping away at one piece of the NCLB legislation, the plaintiffs have successfully shaken the stability of NCLB’s centering powers. Subsequent legal challenges can continue to destabilize NCLB, adding “counter-genres” to the NCLB genre ecology. These “counter-genres” resist the inward pull of NCLB.

Other destabilizing genres include counter-policy statements, counter-assessments, and analysis of NCLB. Counter-policy statements such as NCTE’s *The Impact of the SAT and ACT Timed Writing Tests* and the CCC Position Statement on Writing Assessment act as centrifugal forces, setting forth alternative authorial visions of testing. Unofficial documents like fairtest.org materials also act as centrifugal forces in that they denormalize the activity of testing. Counter-assessments can show that “scientifically based” research is not a one-size-fits-all measure. And finally, structural analysis of NCLB focuses attention on the material outcomes of NCLB and its supporting ecology, revealing that NCLB debates are not rooted simply in ideological differences.

Likewise, looking at other kinds of test data and examining those data through different means allows us different narratives of academic achievement. For example, in *Black-White Test Score Gap*, Christopher Jencks and Meredith Phillips cite NAEP studies that show differences in black-white achievement actually narrowed through the 1970s and 1980s. In another NAEP analysis published by Columbia University, it was found that “after accounting for the fact that private schools serve more advantaged populations, public schools perform remarkably well, often outscoring private and charter schools” (Lubienski and Lubienski 2). Geneva Smitherman’s research on NEAP writing scores shows how measurement tool and reader training can change test results. Smitherman found that from 1969 to 1979 Black English syntax “correlated significantly and negatively with rater score for both primary trait and holistic scores” on essays written by African American junior high students (“Blacker” 16). Black English syntax continued to affect *holistic* scores for the 1984 and 1988 writing assessments. On the other hand, a comparison of *primary trait scores* from 1969–79 and 1984–88 yielded another finding. While the NAEP data from 1969–79 showed a negative correlation between Black English syntax and primary trait score, the NAEP data from 1984 and 1988/89 did not show a high correlation between Black English syntax and primary trait scores. Smitherman concluded that this change was the result of better rater training:

Ideally, if raters are assessing task accomplishment for specific features of discourse production, then features of grammar and syntax would count minimally or not at all in whether a writer accomplished the rhetorical demands of the task. This ideal was not achieved in either 1969 or 1979, but 1984 and 1988/89 NAEP raters accomplished this goal. (“Black English” 56)

Such analysis of NAEP data is rich evidence that assessment data can yield alternative narratives of student success. Such analysis of state NCLB data would likely yield such competing evidence of academic achievement.

Yet another way to deconstruct NCLB is to expose the disparity between its purported goals of testing policies and its actual practices and outcomes. While there has been a rise in test scores at some schools, NCLB has not improved test scores at many schools, despite federal threats to cut school funding. As Ann Owens and Gail L. Sunderman of the Harvard Civil Rights Project report in their policy brief “School Accountability under NCLB: Aid or Obstacle for Measuring Racial Equity,” many repeatedly poor-performing schools are minority-serving institutions. Thus, ironically, NCLB, which purports to

improve student learning, actually “concentrates sanctions in schools serving disadvantaged and minority students” (Owens and Sunderman 1).

A final way to deconstruct NCLB is to address its cascading discursive consequences. The cascading discursive consequences of social stereotypes about racial identity and educational achievement ultimately affect students’ attitudes toward tests and their performances on tests. Claude Steele’s research on “stereotype threat” has shown how cultural stereotypes about race and educational performance ultimately influence students’ test-taking abilities. In Steele’s widely cited article “A Threat in the Air,” he describes a series of experimental studies indicating African American students performed more poorly on tests in which they believed a negative stereotype existed about their intellectual ability and linguistic proficiency. Rather than internalizing failure or rejecting success, African American students in Steele’s studies have internalized cultural expectations that their performance will be secondary to that of white students—expectations that are reinforced through teacher expectations and tracking (Ferguson; Ogbu). Given Steele’s research it seems that one of the best possibilities for righting stereotypes in educational testing is by changing the stereotyped conditions in which learning occurs. For example, Geoffrey Cohen, Claude M. Steele, and Lee D. Ross showed that invoking standards in responding to student writing can actually improve student performance (see also Cohen et al., “Reducing”). Cohen, Steele, and Ross summarized their findings: “when the feedback [on student writing] was accompanied both by an invocation of high standards and by an assurance of the student’s capacity to reach those standards, Black students responded as positively as White students and both groups reported enhanced identification with relevant skills and careers” (1302).

Given the consequences of NCLB on the U.S. educational system and the current administration’s desire to expand NCLB to include students at all educational levels, the renewal of NCLB this year is significant. The United States now has a precedent for federal control over public education and, specifically, over educational testing. The potential impact of this change in federal power on college learning and particularly on composition cannot be understated. My hope is that by understanding the ecology by which NCLB controls American education, we can better understand how NCLB allows certain perspectives to dominate. More specifically, we can understand the role of various genres in mediating that control as well as how competing genres can destabilize that control. Genres can be used remarkably effectively to create a legend

of a testing center; however, that legend of the center does not have to hold. The center can be ruptured.

### Acknowledgments

*I would like to thank David Russell, John Cogan, and Lisa Dush for their advice on earlier drafts of this commentary.*

### Works Cited

- Bakhtin, Mikhail. *The Dialogic Imagination*. Trans. Caryl Emerson and Michael Holquist. Austin: U of Texas P, 1981.
- Bazerman, Charles. "What Is Not Institutionally Visible Does Not Count: The Problem of Making Activity Assessable, Accountable, and Plannable." *Writing Selves/Writing Societies: Research from Activity Perspectives*. Ed. Charles Bazerman and David Russell. Fort Collins, CO: WAC Clearinghouse, 01 Feb. 2003. 6 June 2006 <[http://wac.colostate.edu/books/selves\\_societies/bazerman/bazerman.pdf](http://wac.colostate.edu/books/selves_societies/bazerman/bazerman.pdf)>.
- Cohen, Geoffrey L., Julio Garcia, Nancy Apfel, and Allison Master. "Reducing the Racial Achievement Gap: A Social-Psychological Intervention." *Science* 313, no. 5791 (01 Sept. 2006): 1307–10.
- Cohen, Geoffrey L., Claude M. Steele, and Lee D. Ross. "The Mentor's Dilemma: Providing Critical Feedback across the Racial Divide." *Personality and Social Psychology Bulletin* 25.10 (1999): 1302–18.
- Ferguson, Ronald. "Teachers' Perceptions and Expectations and the Black-White Test Score Gap." *The Black-White Test Score Gap*. Ed. Christopher Jencks and Meredith Phillips. Washington, DC: Brookings Institution, 1998. 273–317.
- Giltrow, Janet. "Meta-Genre." *The Rhetoric and Ideology of Genre: Strategies for Stability and Change*. Ed. Richard Coe, Loreli Lingard, and Tatiana Teslenko. Cresskill, NJ: Hampton P, 2002. 187–205.
- Gonzalez, Roseann Duenas, ed. *Language Ideologies: Critical Perspectives on the Official English Movement*. Vol. 2: *History, Theory, and Policy*. Urbana, IL: National Council of Teachers of English, 2001.
- Hess, Frederick, and Michael Petrilli. *No Child Left Behind: Primer*. New York: Peter Lang, 2006.
- Inoue, Asao B. "The Technology of Writing Assessment: Structural Racism and a Case for Racial Validity." *Handbook of Research on Assessment Technologies, Methods, and Applications in Higher Education*. Hershey, PA, in press.
- Jencks, Christopher, and Meredith Phillips. "The Black-White Test Score Gap: An Introduction." *The Black-White Test Score Gap*. Ed. Christopher Jencks and Meredith Phillips. Washington, DC: Brookings Institution, 1998. 1–54.
- Jennings, Jack, and Diane Stark Rentner. "Ten Big Effects of the No Child Left Behind Act on Public Schools." *Phi Delta Kappan* 88.2 (01 Nov. 2006). Center on Education Policy. 03 August 2007 <[http://www.pdkintl.org/kappan/k\\_v88/k0610jen.htm](http://www.pdkintl.org/kappan/k_v88/k0610jen.htm)>.
- Karp, Stan. "Some Gaps Count More Than Others." *Rethinking Schools Online*, Winter 2003. 23 July 2007 <[http://www.rethinkingschools.org/special\\_reports/bushplan/gaps182.shtml](http://www.rethinkingschools.org/special_reports/bushplan/gaps182.shtml)>.

- Kohn, Alfie. "NCLB and the Effort to Privatize Public Education." *Many Children Left Behind: How the No Child Left Behind Act Is Damaging Our Children and Our Schools*. Ed. Deborah Meier and George Wood. Boston: Beacon P, 2004. 79–92.
- Kozol, Jonathan. *Shame of the Nation: Restoration of Apartheid Schooling in America*. New York: Three Rivers P/ Crown, 2006.
- Liston, Dan, Jennie Whitcomb, and Hilda Borko. "NCLB and Scientifically-Based Research: Opportunities Lost and Found." *Journal of Teacher Education*. 58.2 (March/April 2007): 99–107.
- Lubienski, Christopher, and Sarah Theule Lubienski. "Charter, Private, Public Schools and Academic Achievement: New Evidence from NAEP Mathematics Data." Jan. 2006. National Center for the Study of Privatization in Education, Columbia University. 15 July 2007 <[http://www.ncspe.org/publications\\_files/OP111.pdf](http://www.ncspe.org/publications_files/OP111.pdf)>.
- Massachusetts Department of Elementary and Secondary Education. "Secondary School Improvement." American Diploma Project. 01 July 2007 <<http://www.doe.mass.edu/hsreform/adp.html>>.
- Meier, Deborah, and George Wood. *Many Children Left Behind: How the No Child Left Behind Act Is Damaging Our Children and Our Schools*. Beacon Press, 2004.
- No Child Left Behind Act*. Public Law 107-110. 8 Jan. 2002. Stat. 115.1425.
- Ogbu, John U. *Black American Students in an Affluent Suburb: A Study of Academic Disengagement*. Mahwah, NJ: Lawrence Erlbaum, 2003.
- Owens, Ann, and Gail L. Sunderman. "School Accountability under NCLB: Aid or Obstacle for Measuring Racial Equity?" 10 Oct. 2006. Policy Brief: The Civil Rights Project at Harvard University. 30 Aug. 2007 <[http://www.civilrightsproject.ucla.edu/research/esea/NCLB\\_Policy\\_Brief\\_Final.pdf](http://www.civilrightsproject.ucla.edu/research/esea/NCLB_Policy_Brief_Final.pdf)>.
- PBS Online. "The Federal Role in Education before No Child Left Behind." 21 August 2005. *The Online NewsHour*. 2 Feb. 2008 <[http://www.pbs.org/newshour/indepth\\_coverage/education/no\\_child/before.html](http://www.pbs.org/newshour/indepth_coverage/education/no_child/before.html)>.
- Pollard, Rebecca. "Excerpts from Capstone Report to Rappaport Institute." 2006. Rappaport Institute of Greater Boston, Kennedy School of Government, Harvard University. 3 July 2007 <<http://www.ksg.harvard.edu/rappaport/service/fellows/alumni/reports/pollard.htm>>.
- Russell, David R. Personal communication. 3 Sept. 2007.
- Smitherman, Geneva. "Black English, Diverging or Converging? The View from the National Assessment of Educational Progress." *Language and Education* 6.1 (1992): 47–61.
- . "'The Blacker the Berry, the Sweeter the Juice': African American Student Writers and the National Assessment of Educational Progress." Paper presented at the Annual Meeting of the National Council of Teachers of English. Pittsburgh, PA. November 1993. ED 366 944.
- Spellings, Margaret. "Building on Results: Blueprint for Strengthening the No Child Left Behind Act." Jan. 2007. US Department of Education. 15 Feb. 2008 <<http://www.ed.gov/policy/elsec/leg/nclb/buildingonresults.pdf>>.
- Spinuzzi, Clay. "Describing Assemblages: Genre Sets, Systems, Repertoires, and

Ecologies." Computer Writing and Research Lab, Department of Rhetoric and Writing, U of Texas, Austin, 2004. 7 July 2008 <<http://www.cwrl.utexas.edu/node/158>>.

———. *Tracing Genres through Organizations: A Sociocultural Approach to Information Design*. Cambridge, MA: MIT P, 2003.

Steele, Claude. "A Threat in the Air: How Stereotypes Shape Intellectual Identity and Performance." *American Psychologist* 52.6 (1997): 613–29.

Trimbur, John. "Literacy and the Discourse of Crisis." *The Politics of Writing Instruction: Postsecondary*. Ed. Richard Bullock, John Trimbur, and Charles Schuster. Portsmouth, NH: Heinemann, 1991. 277–95.

**Mya Poe** is Director of Technical Communication at the Massachusetts Institute of Technology, where she teaches undergraduate and graduate courses in rhetoric and scientific writing. Her research interests include the role of racial identity in writing assessment, scientific writing, and writing across the curriculum. Currently, she is conducting research for *Writing and Speaking as a Scientist and Engineer: Perspectives from MIT*, which will be published by MIT Press.

## The Call of Research: A Longitudinal View of Writing Development

*Nancy Sommers*

Consider the situation of first-year students in a college composition course. If the students are fortunate, they have had some writing instruction in high school and have learned something about the structure of a "tell them what you are going to tell them, tell them, and tell them what you told them" five-paragraph theme. And if the students are even more fortunate, they have had some form of instruction, most probably in a history course, in writing research papers about encyclopedic topics such as the Salem Witch Trials or the American Civil War. But college writing, as one of my students called it, is "something more and something deeper than high school writing." In college, students are asked to write *more* than a plot summary, *more* than a cut-and-paste presentation of secondary sources. Rather, they are instructed in the language of analysis, argument, and counter-argument and are urged, in our responses to their drafts, to "analyze more" and to "go deeper into their sources." As academic writers, they are asked to engage with the ideas of others, while most likely encountering and trying to comprehend these ideas, and their authors, for the first time.

It is healthy to remind ourselves how daunting and complex the conventions of academic writing look to first-year college students, even to most undergraduates, as they practice what seems, at first, a set of secret handshakes and esoteric codes, requiring arcane passwords and goofy stances, practiced few other places in life beyond their windowless academic rooms. For if writing were only as simple as arranging twenty-six letters into poses and moves, our students would not need years of instruction and practice in rehearsing other people's arguments before being able to articulate their own. And it is salutary to remind ourselves about these complexities of college writing when we begin discussions about writing development. For if students were only required to interpret and produce *one* set of handshakes or *one* kind of secret code, rather than the wide range they are asked to master as they move across academic disciplines, we might be able to develop a standardized test to describe students' growth and development as writers. But problems begin to multiply, especially in our "no child, every teacher left behind" culture of assessment, when we are asked to provide data to one simple question: Do students leave college as stronger writers than when they entered?

This is a stunning and seductive question, really—do seniors graduate as stronger, i.e., better writers than when they entered as freshmen?—a question with huge implications, both for writing programs and undergraduate education. Let's ponder for a moment how we might go about answering such a question. If we think of writing development in the way child development is often described, with sequential, observable, and concrete milestones, we might define similar developmental stages to measure growth in writing. If this were possible—imagine here a series of monographs: *Your Student Writer as a Freshman*; *Your Student Writer as a Sophomore*—we could create milestones to describe the developmental pathways of student writers to assess their growth or lack of progress. Such milestones, in fact, are not unlike the WPA Outcomes Statement that identifies "the common knowledge, skills, and attitudes sought by first-year composition programs." These bold twenty-two outcomes, though, which are described in the Statement as skills that *should* characterize undergraduate writers at the end of their first year—skills such as "students should be able to respond to the needs of different audiences and different rhetorical situations"—sound more like the common dream of writing program administrators, an idealized view of writing development, rather than developmental pathways based upon decades of composition research.

It seems plausible, though, that students make steady progress in their writing, but this conception of steady progress does not match the research

findings from longitudinal studies completed at six institutions—City College of New York, University of Dayton, Pepperdine University, University of Massachusetts/Amherst, Harvard, and Stanford University.<sup>1</sup> These studies converge to offer two paradoxes of writing development: (1) writing development is not always visible on the page—students may be able to articulate standards of good writing before being able to put them into practice; (2) writing development involves steps both forward and backward, gains and losses, and requires some amount of “bad” writing while new skills are practiced. These steps backward, which often defy our best attempts to describe progress, are often indicators that students are struggling to learn something new. It is not uncommon to see students regress in one area as they practice another. From a longitudinal perspective, writing development is neither linear or sequential, nor entirely predictable.

The big picture of undergraduate writing is hard to grasp because there are so many puzzling pieces that do not easily fit together into a seamless whole. It isn't the case, for instance, that strong high school writers necessarily make a smooth transition to college writing, or that weaker writers necessarily continue to lag behind stronger writers. It isn't the case that students who prosper in the carefully choreographed composition classroom, with sequenced writing assignments and opportunities to revise, necessarily prosper elsewhere in the college without such guided instruction. And it isn't the case that a student who writes a well-organized paper for a political science course will be able to do the same in art history, with a different set of conventions and expectations.

The big picture of undergraduate writing is also difficult to grasp because, like child development, which is similarly difficult to measure and infinitely varied (but, which, at least, has a presumed end point of adult competencies), we do not have an end point for college writing. The freshmen and sophomores in our study, for the most part, were still struggling with the public dimension of academic writing, unable to meet the WPA outcome of responding to the needs of different audiences. So, without knowing the developmental progression that would lead to such an outcome, is it realistic or unrealistic to expect that as seniors these students will be able to produce prose that addresses the needs of different readers? And is the end point for college writing the ability to meet such an outcome in every piece of writing, in every discipline, or just in one?

The challenge, then, of describing writing development when it is not always happening on the page, when it consists of steps backward and for-

ward, starts and stops, and has no recognizable end point, is a challenge that consumed our research team as we dug, deeper and deeper, into the 600-pound avalanche of student writing we collected while following 400 students from the Harvard class of 2001. Although the Harvard Study is based on one cohort of students at one particular institution, the findings from the study resonate with findings of longitudinal studies at other institutions. As all these studies have concluded, to reduce the story of an undergraduate education to a single question—do students graduate as stronger or better writers than when they entered?—is to miss the complexity of a college education. In fact, to reduce an education to any one-dimensional question—do students graduate, for instance, as better geologists or sociologists, moral or quantitative reasoners, community or global citizens than when they entered—is, again, to miss the richness, meaning, and process of a college education, no matter how legitimate these questions might seem to the stakeholders involved. And to reduce four years of college to a series of outcomes or disciplinary rubrics is to find ourselves, in the current culture of assessment, to be asked to measure something that we do not know how to interpret. We might be able to count the grammatical and stylistic errors students make when they arrange their alphabets, but we have not determined how to measure the unpredictable and uneven path of writing development.

The Harvard Study was exploratory in nature, designed to observe undergraduate writing through the eyes of students and to listen to the stories students wanted to tell us in interviews, not the stories we set out to hear or could have heard without going through the efforts of conducting a lengthy study of such a large group. But the methodological challenge of an exploratory study is to figure out what can be learned from a single story as well as from 400 stories, and to determine from such stories why some undergraduates progress as writers while others lag.<sup>2</sup> To study writing development, then, is to study the composing process writ large—across drafts and disciplines, inside and outside of classrooms—and to acknowledge that learning to write well happens over time (though it may not occur at all), and that even with the best pedagogy, some students make very few gains as college writers.

From a longitudinal perspective, the first-year writing course is hardly a make or break academic experience; its prognostic power is provisional and tentative at best. Although the course advertises itself as a foundation for academic writing, students who go through their undergraduate years *only* on the lessons of first-year writing, without additional instruction, especially instruction in the methods of their chosen discipline, find that such a foundation is

inadequate for the more and deeper research assignments they are asked to complete in their junior and senior years. Why some students prosper within their disciplines, while others stall, is one of the perplexing questions we explored, and a question that is intimately connected to issues of audience. Students who prosper as writers, we observed, cultivate a desire to enter disciplinary debates and to find their place in an academic exchange, with something to gain and much to give.

Let me take you inside the Harvard Study and offer one story, that of Luisa, whose college writing experiences illustrate many of the study's themes.

Luisa's first-year writing would look familiar to veteran composition teachers—opinions stated as facts, sources floating without proper introduction, abrupt shifts between points, and meandering sentences. In the margins of her papers, her instructors scramble to get their bearings by writing comments such as “explain what you mean here,” or “I'm not sure what you mean.” She is surprised by all the question marks her instructors place in the margins of her papers because, as Luisa tells us, what she writes makes perfect sense to her. Dubbing herself the “queen of the B minus,” she struggles through her freshman year to find a place for herself, socially and academically, especially aware, as the first child in a large Mexican-American family to attend college, that her fellow students have different life experiences. At the end of her first year, Luisa would have met few of the WPA Outcomes, especially the ability to write for different audiences. But the outcome she met—one not identified on the WPA list—is one I would call *understanding a larger role for writing than just completing a single assignment*. Although Luisa's instructors are frustrated by her unsupported opinions or her confusion of opinion and idea, Luisa sees the larger role of writing as an opportunity for advocacy, a chance to write about topics that matter to her—welfare mothers, voters' rights, and her own Mexican-American identity. Writing gives Luisa a provisional sense of belonging in an alien academic culture, one that seems otherwise unwelcoming. And writing gives her a belief in herself—a belief that following her own interests is important to her success as a student. The story of Luisa's freshman year is not one of dramatic changes on paper; it is the story of changes within the writer herself.

By junior year, the “queen of the B minus” begins to move from writing idiosyncratic prose that makes perfect sense to her, though not always to others, to writing public prose that engages readers, albeit inconsistently. Luisa's progress, like the writing development of all the students we followed, is not

linear, sequential, or predictable. Her prose remains uneven throughout her undergraduate years: passages of clarity and genuine argument are followed, in the same paper, by passages of vague, soggy sentences. As a political science major, she is asked to synthesize complex theories, and her strongest writing comes in essays in which she has learned a subject sufficiently well to interpret these theories rather than to summarize them. As it turns out, though, the “B minus” grades Luisa receives in her freshman and sophomore years are auspicious. We observed that students who receive “B minuses” often receive the longest and fullest responses to their papers, either because the messiness of their prose compels lengthy responses or because instructors feel a need to use their comments to justify the grade. When Luisa is asked as a junior what writing advice she would give to freshmen, she tells us: “It is good to think about what kind of comments you receive from course to course, especially when they are the same kind of comments.” And in all the lengthy comments she receives, the message is the same: if Luisa is to engage others in the arguments that matter to her, she needs to explain her logic by employing the methods and conventions of the disciplinary communities she enters.

But how do undergraduates like Luisa learn to explain their logic? How do they learn to use disciplinary methods to engage others in the field who might not agree with them, but who also care about similar ideas? In her junior year, Luisa accounts for her growing comfort as a writer by telling us “practice, repetition, and instruction make the difference.” What surprised us was how much practice, repetition, and instruction it takes for students to progress as writers. Luisa wrote, on average, twenty papers each year; and she took three methods courses in political science—a writing-intensive apprenticeship: “here’s *what* we think about in political science, and here’s *how* we think.” In this apprenticeship, she learned that if she wants to engage in the disciplinary debates of political science and contribute to them—that is, if she wants political scientists to listen to her ideas and speak back to theirs—she needs to learn their language and use their methods and approaches, not just her own idiosyncratic frame of reference. To reach an audience of political scientists, she needs to write like one—testing theories and hypotheses, synthesizing large bodies of information, and anticipating and refuting counterarguments.

Luisa’s undergraduate writing experience taught us about the role of expertise in writing development, both expertise in subject and method. If she succeeds as a college writer, she succeeds not by writing *out* of expertise but by writing *into* expertise. The origin of her expertise does not come in a series of breakthrough moments but in her repeated struggles, her many approxi-

mations of political science methodology, borrowing the materials and methods of a course before making them her own. Her strongest paper during her junior year is one in which she enters a lively debate on school vouchers, asking why the state of Florida passed a school voucher law if public opinion in the state was divided, and using the case study of Florida to challenge a general theory about how school voucher programs are enacted into law. This is her second paper about school vouchers, and she is familiar enough with the topic to position herself in a debate rather than to merely summarize the debate for herself.

For Luisa, expertise and comfort with political science subjects and methods mean that she has the knowledge to define problems worthy of investigation, identify disagreements among sources, and orchestrate debates. One of our students defined expertise in content as “gaining great purchase with the material,” while another explained the value of such knowledge: “writing papers is just a lot easier when you have a background in the subject.” To Luisa, expertise with the subject of school vouchers meant she could make a statement. As she told us, “This is an argument I don’t think has been made before.” She knows enough about school voucher programs to see a gap in the debate, and she has a shared language and shared methods to reach out and engage readers who have not heard this argument before.

The critical role expertise and knowledge play in preparing students to address an audience helps to account for many patterns we observed over four years. Students do not simply struggle with audience at the beginning of their college careers: they struggle with it whenever they find themselves as novices, unfamiliar with the ideas and methods of a particular discipline or subject matter. Expertise—or more accurately, the lack of it—also helps explain why students who never pursue one subject or one discipline in depth have more difficulty engaging readers. Take, for example, a history major we followed who globe-trotted across the curriculum rather than settling into a particular historical time period. With the exception of two writing assignments, she always worked with unfamiliar sources, always learning about different cultures and shifting into new historical landscapes. While other students were digging deeper into their disciplinary materials, our globe-trotter skittered along on the surface of her materials, always breaking fresh ground. Her broad-based program simply did not give her the necessary mastery and expertise in her subject matter to move from private report to public writing.

In a similar way, a biochemistry major, who never wrote on the same topic twice and who received writing instruction only in his first-year composition

course, seemed surprised when he encountered new disciplinary expectations. His comments made us appreciate just how difficult it is as novices to engage an audience in conversation: “All the humanities seem to blend together, kind of vague in my eyes. Papers in these areas make me nervous because I feel as if I’m throwing darts when I write papers.” Without sustained instruction in disciplinary methods, and without expertise in content, such students learn to write and connect with readers in a more haphazard way, more a process of luck than instruction. We observed that when students do not know that there is a method in *one* discipline, they are less likely to look for disciplinary conventions elsewhere. As writers, they stay longer in the stages of imitation and mimicry, and the process of writing papers too often remains magical and mysterious. But when students have a chance to build expertise by returning to the same topics and when they are given instruction and practice in disciplinary methods, they are more likely to engage with writers who have wrestled with similar issues and thus find their place in an academic exchange.

Learning to write to an audience is not a mathematical formula or algorithm that once learned may simply be reapplied in every writing situation. Nor is it an outcome that is easily defined or measured. Every time students read new texts, enter new debates, or practice new sets of disciplinary conventions, they must, in effect, learn how to address a new audience. It is convenient to think of audience in human terms as an assembly of listeners and readers, but it is most often the case that students are unable to know just *who* it is they should be imagining. From analyzing pounds of student writing, we observed that *if* students learn to write *to* and *for* an audience, it does not happen because they have a theoretical grasp of the concept of audience or from an awareness of their readers’ identities. Rather, if students learn to write *to* and *for* an audience, it happens because of a confluence of desires: an institution desires its students to receive explicit, sustained, and incremental instruction in thinking and writing; and students desire to see their writing not as an isolated exercise but as part of an ongoing academic exchange. What we like to call “audience awareness” develops indirectly, from the depth of expertise and investment that imply an audience—from listening to sources, both dusty volumes and living people—and from finding not only something to say but also a reason to say it.<sup>3</sup>

For Luisa, the path to expertise stretched beyond the physical boundaries of college and back to her hometown of Fresno, where she returned, for her senior paper, to research patterns of political participation and voting behaviors among Asian Americans, specifically the Hmong community. Here’s

how she explained her choice of topic: "I'm always thinking where I come from and the ways in which what I learn might have a positive effect on local politics; I went away to college and now I'm bringing something back to my town." Placing Luisa's first-year writing next to her seventy-five-page senior paper, we easily recognize her increased ability to address an academic audience. From the very beginning of her senior paper, Luisa reaches out to her readers with a genuinely provocative puzzle: "Why don't Hmong voters vote, and particularly, why don't they vote in strong enough numbers to elect someone from their particular ethnic group?" Throughout her project, she employs a wide range of political science methods, explains complex concepts, and engages readers, especially people unfamiliar with Fresno politics, to wonder why, given the cohesion of the Hmong community, there are no Hmong on the Fresno City Council. She described the experience of writing her senior paper as "trying to grab a big slippery and uncontrollable fish." At times she thought she had a hold of her fish, but then "it got away."

Luisa's image of "trying to grab a big slippery and uncontrollable fish" describes the challenge of getting a grasp on something as unpredictable and unstable as writing development. At the end of the second year of our study, for instance, we thought we had a hold on one aspect of writing development; that is, the importance of close reading or close textual analysis. The strongest writers, at this point, seemed to be those who were able to pull the "more" and "deeper" from their sources into their analysis of texts. Conversely, students who stayed on the surface of every text they wrote about, reducing the meaning of a text to a simple, homogenized commonplace, seemed limited by their inability to pry open their sources. By junior year, though, many of these "close readers" had stalled and were stuck as writers because their close readings remained isolated exercises, disconnected to any argument or claim that might locate them in larger disciplinary conversations. One such stalled writer, Avery, employed a generic template for every writing assignment: for a sociology paper on homelessness, her thesis was "homelessness is more complicated than has been otherwise suggested." For a paper in a world religions course, her thesis was "religion is more complicated than it first seems." Avery's dazzling close readings, which seemed like significant development in her first two years of college, never served a larger purpose than to protect her beliefs, avoid controversies, and allow her to see the world as "more complicated than it first seems." By junior year, Avery's close readings begin to look more like prepackaged interpretations than true academic inquiry.

Trying to explain almost any slippery aspect of writing development led

us to a proposition that sounds like this: any single element necessary for development is never sufficient for students to succeed as college writers. For instance, close reading is necessary but not sufficient in itself for making arguable claims. Learning to question sources and pose counterarguments are necessary to establish a place in an academic debate but not sufficient to engage readers if a student lacks expertise in the subject matter or lacks expertise with the disciplinary methods of engagement. And learning to ask questions that matter to a student is necessary but not sufficient, especially if these questions matter *only* to the student and not to a community of readers. Every time we try to grasp hold of our slippery and uncontrollable fish, we end up sounding like Avery: writing development is just more complex than has been otherwise suggested.

If we use Lee Ann Carroll's definition of writing development—"doing increasingly complex writing tasks" over a period of time, "not just getting better at the same task" (22–28)—we would have little trouble agreeing that Luisa has developed as a writer. Unlike Avery, who plateaus with her complexity thesis, writing the same kind of paper again and again, no matter what task is assigned, Luisa attempts an ambitious senior paper, with a complexity unimaginable in her first year. In attempting a "slippery and uncontrollable" senior project, though, one with a scope and breadth that she can only approximate as an apprentice, she reaches beyond what she was formerly capable of accomplishing. Over the six months of researching and writing this paper, Luisa is propelled by her desire to say something significant and to give something back to her hometown, but she is hampered by her lack of expertise in blending quantitative and qualitative research. Not having previously written on this topic, she knows little about the Hmong community and needs to handle unfamiliar theories and data.

But if Luisa fails to grab hold of the entirety of her "slippery fish," it is because of the ambitious scope of her project, not because she has not satisfied a series of outcomes for political science majors, such as the disciplinary-specific ones Michael Carter offers in his essay "Ways of Knowing, Doing, and Writing in the Disciplines." In fact, Luisa meets all of Carter's outcomes for political science: she identifies research questions and theories, creates and tests hypotheses, and applies appropriate methodologies. Yet before we reduce a college education to any system of outcome-based assessment, or before we congratulate ourselves if our students meet these outcomes (or lament because they miss the disciplinary mark), we need to be certain that such outcomes are sufficient and accurate measurements of writing development, based

on solid research. For instance, if a student achieves Carter's outcomes for political science majors but is unable to meet the WPA outcome of "responding to the needs of different audiences and different rhetorical situations," how do we interpret such development?

In Luisa's case, writing plays a much larger role in her education than what such political science outcomes would suggest: she writes papers about education and prisons, which lead her to a summer job at a prison education center, and to a teaching position, after college, with Teach for America. From her muddled thinking as a first-year student to her senior paper, she learns to hold a thought and to become, in her own words, "a person with things to say." It is through writing—that is, through understanding a larger role for writing than just completing a single assignment—that Luisa claims her education, argues with views and opinions different from her own, and learns to reach beyond herself to engage readers. And it is through writing that Luisa learns to shape her personal interests into public ones, moving into the wider world as a thoughtful and educated citizen.

The problem with measuring writing development by any set of outcomes is that "outcomes" reduce education to an endpoint, transferring the focus of instruction from students to written products and leaving both students and teachers behind in the process. Developing an awareness of audience is not an endpoint, nor is learning to write in a discipline an endpoint. We tend to discuss audience and disciplines as nouns, but it is the verbs—*reading*, *questioning*, *evaluating*, *interpreting*, and, of course, *writing*—that are the sum and substance—the process—of becoming educated citizens. The verbs matter because they are the specific ways by which students move from muddled thinking to participating in public conversations, with something to gain and much to give. Learning to write *to* and *for* an audience is not an outcome independent of the larger goals of a liberal arts education: it is central to the process of becoming an educated person "with things to say," a process without an endpoint.

### **Acknowledgments**

*In the process of researching and writing this essay, I have incurred many debts: I would like to thank the research associates of the Harvard Study—Laura Saltz, Kerry Walk, Pat Kain, Emily O'Brien, Suzanne Lane, Soo La Kim, Faye Halpern, and Suzanne Young—for their abundant contributions to the study. I would also like to thank Derek Bok for a summer's worth of sustained conversation about the role of writing*

in higher education. And, as always, many thanks to my generous circle of readers for staying with me across the drafts: Joshua Alper, Pat Bellanca, Jenny Lewis, Les Perelman, Maxine Rodburg, Dawn Skorczewski, and Lad Tobin.

### Notes

1. To date, scholars in our field—Marilyn Sternglass, Stephen Wilhoit, Anne Herrington, Marcia Curtis, Lee Ann Carroll, Jenn Fishman, Andrea Lunsford, and colleagues have demonstrated the value of longitudinal studies to provide a wider perspective than research focused upon one college course or one undergraduate year.

2. The undergraduate writing collected from the Harvard class of 2001 now resides in the Harvard archives, available to current and future scholars. To receive copies of the films about the study, *Shaped by Writing* and *Across the Drafts: Students and Teachers Talk about Feedback*, write to: [expos@fas.harvard.edu](mailto:expos@fas.harvard.edu).

3. In recognizing expertise as a basic criterion for academic exchange, we are harkening back to a long-established rhetorical tradition. Aristotle noted that expertise is essential to being a good rhetorician, and spent much of his *Rhetoric* outlining the subjects and strategies an orator should master. Aristotle maintained that good rhetoricians must be well-trained and well-informed, “acquainted with the elements of the question, either entirely or in part.” “For if you know none of these things,” he argued, “you have nothing from which to draw a conclusion” (291). Aristotle’s observations help us explain why students such as Luisa prosper as public writers: their expertise shows them both where and how to step into a scholarly discussion to give advice to their own Athenians.

### Works Cited

- Aristotle. *The Art of Rhetoric*. Trans. J. H. Freese. Cambridge, MA: Harvard UP, 1975.
- Carter, Michael. “Ways of Knowing, Doing, and Writing in the Disciplines.” *College Composition and Communication* 58.1 (Feb. 2007): 385–418.
- Carroll, Lee Ann. *Rehearsing New Roles: How College Students Develop as Writers*. Carbondale: Southern Illinois UP, 2002.
- Council of Writing Program Administrators. “WPA Outcomes Statement for First-Year Composition.” April 2000. <<http://www.english.ilstu.edu/Hesse/outcomes/html>>.
- Fishman, Jenn, Andrea Lunsford, Beth McGregor, and Mark Otuteye. “Performing Writing, Performing Literacy.” *College Composition and Communication* 57.2 (Dec. 2005): 224–52.
- Herrington, Anne J., and Marcia Curtis. *Persons in Process: Four Stories of Writing and Personal Development in College*. Urbana, IL: National Council of Teachers of English, 2000.
- Sternglass, Marilyn S. *Time to Know Them: A Longitudinal Study of Writing and Learning at the College Level*. Mahwah, NJ: Erlbaum, 1997.

Wilhoit, Stephen. "Across the Curriculum, Across the Years: A Longitudinal Study of Writing in College." Paper presented

at the Conference on College Composition and Communication, Chicago, April 1998.

**Nancy Sommers** teaches at the Harvard Graduate School of Education and researches the writing development of college and high school students. As principal investigator of the Harvard Study of Undergraduate Writing, she tracked the college writing experiences of 400 students from the Harvard class of 2001. For twenty years, Sommers led Harvard's Expository Writing Program, where she directed both the first-year writing program and the Harvard Writing Project and held the Sosland Chair in Expository Writing. She is the author of numerous articles and essays about teaching writing and a co-author of Diana Hacker's *A Writer's Reference*.

### **CCCC Studies in Writing & Rhetoric Call for Proposals**

The aim of the CCCC Studies in Writing & Rhetoric (SWR) series is to influence how writing gets taught at the college level. The methods of studies vary from the critical to historical to linguistic to ethnographic, and their authors draw on work in various fields that inform composition—including rhetoric, communication, education, discourse analysis, psychology, cultural studies, and literature. Complete manuscripts run from 25,000 to 40,000 words, or about 125–200 pages.

We ask authors to send us project proposals that clearly situate their work in the field and show how they aim to redirect our ongoing conversations about writing and its teaching. Proposals should include an overview of the project, a brief annotated table of contents, and a sample chapter. Proposals should not exceed 10,000 words. To submit a proposal or to contact the series editor, please go to <http://uwp.aas.duke.edu/cccc/swr/>.