

## Review Essay

### Reinventing Writing Assessment: How the Conversation Is Shifting

William Condon

Adler-Kassner, Linda, and Peggy O'Neill. *Reframing Writing Assessment to Improve Teaching and Learning*. Logan: Utah State UP, 2010. 207 pages.

Broad, Bob, Linda Adler-Kassner, Barry Alford, Jane Detweiler, Heide Estrem, Susanmarie Harrington, Maureen McBride, Eric Stalions, and Scott Weeden. *Organic Writing Assessment: Dynamic Criteria Mapping in Action*. Logan: Utah State UP, 2009. 167 pages.

Elliott, Norbert. *On a Scale: A Social History of Writing Assessment in America*. New York: Peter Lang, 2005. 408 pages.

Ericsson, Patricia Freitag, and Richard Haswell. *Machine Scoring of Student Essays: Truth and Consequences*. Logan, Utah: Utah State UP, 2006. 268 pages.

Huot, Brian and Peggy O'Neill, eds. *Assessing Writing: A Critical Sourcebook*. Boston: Bedford, 2009. 500 pages.

Huot, Brian. *(Re)Articulating Writing Assessment for Teaching and Learning*. Logan: Utah State UP, 2002. 216 pages.

Lynne, Patricia. *Coming To Terms: A Theory of Writing Assessment*. Logan: Utah State UP, 2004. 193 pages.

Neal, Michael R. *Writing Assessment and the Revolution in Digital Texts and Technologies*. New York: Teachers College P, 2011. 152 pages.

O'Neill, Peggy, Cindy Moore, and Brian Huot. *A Guide to College Writing Assessment*. Logan, UT: Utah State UP, 2009. 218 pages.

Paretti, Marie C. and Katrina M. Powell, eds. *Assessment of Writing*. Tallahassee: Association for Institutional Research, 2009. 218 pages.

Weigle, Sara Cushing. *Assessing Writing*. Cambridge: Cambridge UP, 2002. 268 pages.

Wilson, Maja. *Rethinking Rubrics in Writing Assessment*. Portsmouth: Heinemann, 2006. 111 pages.

Wolcott, Willa, with Sue M. Legg. *An Overview of Writing Assessment: Theory, Research, and Practice*. Urbana: NCTE, 1998. 206 pages.

The plethora of books written by scholars within the field of Rhetoric and Composition about writing assessment over the past ten years is a strong indication that the conversation about writing assessment has reached a kind of tipping point. Beginning in the 1970's, Edward M. White led a movement in the California university system for direct testing of writing—with the help, we must acknowledge, of insiders at the Educational Testing Service (Albert Serling and Paul Diederich most prominently) who had long been proponents of timed essay testing within their organization. White's group of faculty, who got more support from ETS than from their own system administrators (White, "The Opening"), represents perhaps the first meaningful engagement of the classroom context with the testing community since 1874, when Adams Sherman Hill was hired at Harvard to assess the writing of incoming students and to teach them expository writing. Almost from the beginning of writing assessment in the United States, as Norbert Elliot and as Peggy O'Neill, Cindy Moore, and Brian Huot document, teachers' voices were not welcome in the negotiations. In 1895, Wilson Farrand, of Newark Academy, put forth a plan for what would become the College English Examination Board (CEEB) in which secondary schools that were feeding students into the universities would collaborate in testing those students, the payoff being the kind of communication that would result in better faculty and curriculum development (Elliot 22-26). While the CEEB grew out of Farrand's proposal, there would be no collaboration with teachers: "teacher judgments about student preparation were found suspect. A test was assumed to be better at helping university admissions personnel make important, consequential decisions about students than judgments of secondary teachers" (O'Neill, et al. 17). Peter Sacks documents the continuing irony of the College Board's arrogant early position: studies overwhelmingly demonstrate that high school grades are better predictors of success in college than test scores of any kind (271). In fact, even in its formation, the CEEB held direct tests of writing

in contempt, an attitude that the Board would maintain until the 1970's—even, as Elliot reports, in the face of their own experts who consistently advocated for the direct test.

The breakthrough that White led in California was soon followed by further developments of tests by other large agencies as well as by colleges and universities themselves. My connection to writing assessment began in 1987 at the University of Michigan, which from 1978 forward conducted its own timed writing test of incoming students in order to place students at an appropriate point in the first-year writing curriculum (Bailey and Fosheim). Even at that point, however, the next stage of the conversation was coming into view. Again, even from the beginnings at Harvard, few teachers were satisfied with the outcomes of a timed test of writing. Elliot quotes Adams Sherman Hill, writing in 1878:

Those of us who have been *doomed to read manuscript written in an examination room* (emphasis added)—whether at a grammar school, high school, or a college—have found the work of even good scholars disfigured by bad spelling, confusing punctuation, ungrammatical, obscure, ambiguous, or inelegant expressions. (qtd. in Elliot 341)

While Elliot describes the ways this dissatisfaction simmered at the College Board and, later, in its testing arm, the ETS, by the 1970's teachers were beginning to take matters into their own hands. Since that time, indirect tests—examinations that do not engage test-takers in writing—have declined as placement instruments and even, in latter years, as college entrance instruments. Direct tests were a good first step away from multiple choice question tests, but the timed sample was itself of dubious quality. There are simply not many instances in education or in life when a writer is called upon to sit and in a limited time (ranging from twenty-five minutes on the new SAT to an hour or two on most college-based tests) produce an essay on a topic of which the writer has no prior or specialized knowledge. In fact, I can think of only one such circumstance: the timed writing test.

Shouting “Validity!” as their battle cry, teachers began looking for ways to examine students’ writing abilities in a more natural context, using more authentic samples. Thus was born the writing portfolio. Portfolios gained fairly widespread use in classrooms during the 1970's, but in 1986, Peter Elbow and Pat Belanoff reported on their program-wide use of writing portfolios for grading students in first-year writing courses at SUNY Stonybrook, and the second phase of the conversation had begun. In the writing portfolio, teachers had found an instrument with which to fight back against the way both indirect and timed direct tests of writing under-represented the construct *writing* (Hamp-Lyons and Condon). In other

words, since portfolios include writing that students do under normal conditions—which may vary from one student to the next and even for one student from one assignment to the next—portfolios represent more fully *how that writer writes*. For decades, developers of timed writing tests at ETS struggled to achieve acceptable reliability in scoring their samples, a struggle that Paul Diederich finally resolved in developing a holistic scoring system that engaged raters in applying standard, if limited criteria in assessing each sample, in double- (and, if necessary, triple-) readings of each sample. The challenge for proponents of portfolio-based writing assessment, then, was twofold: to advance the cause of validity while meeting the challenge of reliability.

During the decade of the 1990's, those challenges were met. Entry-level portfolios at Miami University and the University of Michigan and the Junior Writing Portfolio at Washington State University led the way by demonstrating that portfolio-based writing assessments were logistically possible, that portfolios could be scored as reliably as timed writings, that placements made on the basis of portfolios were more appropriate, and that portfolios engaged with and supported the curriculum in ways that timed writings cannot (see Hamp-Lyons and Condon; Willard-Traub, et al.; Daiker, et al.; Haswell). Writing teachers and a growing cadre of writing assessment experts within academe (White, Huot, Kathleen Yancey, Elbow, Belanoff, Richard Haswell, William Smith, Michael Williamson, Hamp-Lyons, and several others) had established the prominence of construct validity as a necessity for legitimate writing assessment, as well as the value of the multiple kinds of validity that scholars such as Samuel Messick, Lee J. Cronbach, and Pamela Moss were isolating and describing (consequential, predictive, face, concurrent, and other kinds). These experts from within the academy reasserted the value of traditional validity—that the test actually addresses the construct it sets out to assess—as the means for insisting on including actual writing in any test that purports to yield results that speak to the test-takers' writing ability(ies). In furthering the cause of authentic tests of writing, the field has followed the development of additional kinds of validity, which require that tests not only match the construct, but also offer evidence that their consequences provide educational benefits, that the predictions they make turn out to be accurate once the student is placed into a curriculum, as well as the basic assurances that the assessment is conducted in a manner that is fair to test takers. O'Neill, et al. point out that timed tests have difficulty meeting these descriptions of validity, especially since the American Psychological Association (APA), the American Educational Research Association (AERA), and the National Council on Measurement in Education (NCME) testing standards require

a test to meet a unified standard of validity (27). Michael Neal puts the dilemma of timed essay testing this way:

Many of the current writing assessment technologies are aligned with values of efficiency, uniformity, speed, and mechanization. Some of these technologies are so deeply entrenched in educational contexts that they seem nearly impossible to challenge, and yet there are voices that have and continue to speak into these contexts. [Large-scale testing agencies] have mistaken elevated uniformity and consistency for fairness, resulting in writing assessments that are inconsistent with many of the most fundamental values and best practices associated with the field. These current assessment models are bolstered by their strong connection to large-scale social, public perceptions of technological fixes and reductive views of literacy that tend to reduce language to surface feature, formulaic arrangement, diminished writing processes, and social dynamics of written communication, and often function outside a rhetorical context for writing. These assessments allow us to manage and compare mass populations of students and perhaps reduce the “burden” of classroom assessments, but they fall short of more desirable outcomes. (132)

And the situation is even more grave today, as David Nye points out: the “written component of the new SAT Verbal ... is the antithesis of nearly every current theory of composition and writing assessment and does not take into account the composing technologies students use in authentic environments” (qtd. in Neal 49). Timed direct tests of writing, at least on the commercial side of the enterprise, fall so far short of the demands of unified validity that they do not meet the most basic standards for responsible practice laid out by the APA, AERA, and NCME.

Well-designed portfolio-based writing assessments, because of the broader educational benefits they bestow on students and on teachers, can meet those standards. As a result, portfolio practice and the scholarship around that practice grew exponentially during the 1990's. By 1996, Robert Calfee and Pamela Perfumo reported that sixty percent of secondary English teachers used portfolios within their classrooms, while twenty percent used them across classes, and another ten percent were using portfolios in ways that reached beyond their school buildings. The practice of assessing students' writing—and beyond that, their overall learning—via a writing portfolio had clearly taken hold. Willa Wolcott and Sue M. Legg include a chapter on portfolios in a volume otherwise devoted to timed direct tests—but oriented to K-12 classroom teachers. That 1998 volume, *An Overview of Writing Assessment*, acts as a kind of threshold to the conversation this

article describes. Wolcott and Legg describe the basics of writing assessment so that teachers can understand not only how direct testing happens, but also the limits of these tests. In addition, the book has a “how to” flavor, an implication at least that teachers could be active participants in and even designers of responsible writing assessments (White did much the same thing for college-level teachers in his 1999 work *Assigning, Responding, Evaluating*, which is now in its fourth edition, 2007).

From that jumping-off place, the conversation over the past decade has grown to the point that teachers of writing increasingly recognize the inadequacy of the old paradigms of writing assessment and demand not just connections between the classroom and the assessment enterprise, but a meaningful role in those assessments as well. Over the years since Wolcott and Legg, too many books for one review article have emerged, and while no one article can hope to include them all, I address the remainder of this article to identifying in the last decade of writing assessment trends, currents, and crosscurrents, sampling the range of scholarship coming out of academic research on writing assessment, and describing how this re-entry into the conversation may be changing the dialog, tipping the balance away from the psychometrics-dominant past and toward a view of writing assessment that prizes the traditional emphasis on responsible assessment (reliability, validity, sound practices) while insisting, at last (pace, Wilson Farrand!), that what happens in classrooms matters. In so doing, I note the ubiquity of the prefix “re-” in these books, and not only in their titles (Reframing, (Re)articulating, Rethinking, Revolution). I hope readers of this essay will forgive my following these authors’ lead, as their trope reinforces the basic theme of this essay that current scholars are (ahem) reconstructing writing assessment to provide teachers with a greater voice.

#### REHISTORICIZING WRITING ASSESSMENT

Let us be as clear as possible. In the twentieth century, “assessment largely promoted reductive views of language in favor of ... efficiency, mechanization, and cost effectiveness” (Neal 5). The ever-continuing descent of commercial writing assessment is driven by (1) cost-cutting in the testing industry resulting from a profit motive on the part of assessment companies that forces assessments to become less and less valid (because tests become shorter and shorter, thus severely restricting the construct being tested); and (2) a budget-cutting desire on the part of colleges and universities to offload the cost of placement onto students. Thus, we find COMPASS, E-rater, Criterion, and others that use tests so limited that they can be scored—not read, mind you, but scored—by computers. We see the new SAT, a cyni-

cal response to the California system's threat to discontinue using the SAT for admissions unless it included a direct test of writing. The resulting test is weighted so that two-thirds of a student's "writing" score is based on an indirect test of vocabulary and grammar, and one-third is based on a twenty-five-minute (!) written sample that, as Les Perelman has demonstrated, ties quality of writing so directly to essay length that the samples can be scored from across the room (Anson, et al.). And because many institutions adopt these tests for their own reasons of economy—these are cheap assessments that also allow the college to pass the costs on to the students—writing teachers are put into a difficult position. Writing assessment as a technology informs "many of the habits and notions we have developed in contemporary educational settings. The most detrimental effect of such assumptions is that educators have become (often unintentionally or against our better judgment) proponents of writing assessments that often are reductive and at odds with our best understanding of teaching and learning" (Neal 5). In effect, teachers in colleges that employ such tests for placement or other purposes are put in the position of endorsing those tests, whether the teachers were consulted about the adoption or not.

Yet the reasons that these new voices are beginning to have an effect beyond the academy lie in the very origins of the timed writing test. Indeed, timed writing, while inevitably limited in its generalizability and usefulness, is still being practiced in more responsible ways—at universities such as Washington State, Louisiana State, MIT, Hawaii, and others that provide sufficient time for writing; collect multiple samples; engage teachers in constructing the tasks, creating the criteria, and rating the samples; tie the prompts to local curriculum; and in various ways create a community around the assessment so that instead of obtaining only a ranking that allows for a placement, the assessments provide far more information about students' learning experiences than can be produced by the kinds of short, too-tightly controlled, de-contextualized samples favored by the commercial side of writing assessment (see, for example, Condon).

Hill's objections to the quality of the writing in those early Harvard essays, Farrand's call to engage assessment with instruction, these early voices were ignored in the frenzy to develop indirect tests that is described, in a larger context, by Stephen Jay Gould's *The Mismeasure of Man*. In our own field, Elliot's *On a Scale* is essential in understanding the history behind the state of affairs to which so many who teach writing object so strenuously. Elliot's volume is a detailed history of writing assessment as it emerged in the context of the College Board and its eventual testing arm, ETS. The account comes about as close to vivid drama as a thoroughly scholarly work can, as Elliot tells the story of the internal clashes between

the ruling faction, which favored indirect tests because of the reliability of the scoring process, and the sequence of test development teams who promoted investment in essay tests and even, in the 1980's and 1990's, portfolio-based assessments. Elliot had unprecedented access to internal records at ETS, and the resulting account—a social history, as the subtitle notes—provides, for the first time, a fully contextualized history of writing assessment, especially but not solely as it came to be a large commercial enterprise. Still, Elliot's account begins and ends with teachers, teachers who find that rubrics help them communicate expectations to their students, demystify the grading process, and provide a common language for talking about writing in the classroom. Even while telling a story about an assessment juggernaut that consciously and consistently excluded teacher input, this volume addresses what teachers need to know about the assessment enterprise and how teachers can apply the better aspects of that enterprise in the service of students' learning. In doing so, Elliot closes the feedback loop that CEEB and ETS steadfastly refused to close. This book helps teachers understand the history behind the dominance of indirect and direct tests of writing. As teachers grasp the import of separating assessment from what happens in classrooms, they can become more informed advocates of engaging instruction with assessment.

Elliot leads the emphasis among academics to reframe the history of writing assessment, and *On a Scale* remains the most detailed account—though it focuses almost solely on the College Board and ETS. O'Neill, Moore, and Huot's *A Guide to College Writing Assessment* begins with a history of writing assessment in the US that is shorter, if broader than Elliot's, since the authors' account is not limited to the College Board. In roughly thirty pages, O'Neill et al. summarize Elliot's account and reach beyond it to tie that history to events outside ETS. In effect, they extend the history beyond the development of robust indirect tests of writing, which is the effective close of Elliot's history. Huot's chapter in this collection provides a view of "Writing Assessment as a Field of Study" that tells this history from a different point of view. Huot's principal aim in the chapter is to lay out a clearly defined field of research, but in the process he deals usefully with the history of writing assessment from the advent of holistic scoring forward. All these accounts are grounded in the history of the field, yet all reframe that history to open the field for the participation of teachers in classrooms—from K-12 through university—whose aim of engaging assessment with instruction essentially changes the prior orientation of assessment away from the classroom. These accounts recast the enterprise by expanding the borders to include the growing number of assessment experts who ply their trade within schools and universities.

This decade's worth of new books also recognizes the explosion of scholarship about writing assessment within the academy. Indeed, the reassessment of assessment is largely happening inside colleges and universities and within schools. Several books provide a compilation of that emerging wisdom. These books—all three of them collections of essays by various leading scholars—fill at least two valuable roles: they establish a canon of key readings on important issues, and they provide an accessible knowledge base for graduate seminars (much as Victor Villanueva's *CrossTalk in Composition Theory* does for teachers of writing). Brian Huot and Peggy O'Neill's *Assessing Writing: A Critical Sourcebook* collects the canon, beginning with a section on "Foundations," where key articles begin with the transition from indirect to direct testing and move forward to Huot's landmark 1996 *College Composition and Communication* essay, "Toward a New Theory of Writing Assessment." "Foundations" encompasses the decades of scholarship in what I have described above as the second phase of the conversation on writing assessment. Thus, it provides examinations of topics ranging from holistic and primary trait scoring, to validity and reliability, to the beginning of the portfolio movement and the first attempt to create a new history of writing assessment, including a reprint of Yancey's "Looking Back as We Look Forward: Historicizing Writing Assessment." While the "Foundations" section in some ways represents the more distant past of writing assessment, the articles here hold up both in their treatment of persistent issues of test validation and in their presentation of the changes in thinking during the roughly two decades that laid the foundation for and anticipated the present day of writing assessment. The second section, "Models," is a much-needed nod toward essentialism. Newcomers to the field need to see examples of the foundational principles in action, and here the editors have assembled a gallery of programs that changed the way writing assessment works: William Smith's expert rater system at Pittsburgh; Richard Haswell and Susan Wyche-Smith on Washington State's Junior Portfolio; Russell Durst, Marjorie Roemer and Lucille Schultz on Cincinnati's program portfolios; Dan Royer and Roger Gilles' founding article on directed self-placement; and examples of writing assessment in the context of WAC. Finally, "Issues" provides a summary of problems the field continues to struggle with, from the notion of holistic scoring of portfolios, to portfolios and second-language writers, to issues of culture and other contexts that impinge on writing assessment. For now, this collection represents our canon, both in the selection of essays and in the authors included in the volume. These are the voices that have brought us to our current state, from White and Elbow, to Williamson, Huot, and Yancey, to Bob Broad and Haswell. Perhaps the sole weakness of *Assessing Writing*

is its treatment of the move toward using computers to score timed essays. Given the long battle Huot, in particular, has fought to keep the elements of responsible test development on the table, even as colleagues in the field call for discarding some of those elements (Lynne, Wilson), I was surprised that only one essay related to machine scoring is included in that collection.

This deficit is more than remedied, however, by Patricia Freitag Ericsson and Haswell's *Machine Scoring of Student Essays: Truth and Consequences*. This collection, which in some ways acts as a response to the industry-sponsored *Automated Essay Scoring: A Cross-disciplinary Perspective* (Shermis and Burstein, eds), provides examinations of the claims that commercial writing assessment makes about automated essay scoring, finding most of them disingenuous at best. *Machine Scoring* also provides practical tests of the leading systems, such as E-Rater and Criterion, both by scholars and in writing programs. Finally, the collection looks more broadly at what is lost, in terms of curriculum and the community that develops around local assessments, when an institution chooses to employ tests that are designed for computers to score. *Machine Scoring* promises to be an important resource for years to come because the political economy of the commercial testing industry dictates finding cheaper and cheaper ways of delivering rankings of test-takers for the purposes of deriving a course placement on a given campus and, as this initiative evolves, for evaluating faculty, courses, and even academic programs. That devolution of what writing assessment is—merely a ranking that leads to a placement—is the opening that teachers and scholars in the academy are exploiting in order to change the assessment landscape to engage a broader range of purposes and effects. In this way, the conversation about machine scoring can help counteract the reductive notions of writing assessment inherent in legislative efforts such as No Child Left Behind and government statements such as the Spellings Commission's report.

As that broader range becomes more apparent, the academic assessment community will inevitably expand, as indicated by *Assessment of Writing* (Paretti and Powell, eds.), the fourth volume of the Association for Institutional Research's series, *Assessment in the Disciplines*. This collection of essays provides models of writing assessment in a WAC context. As the sponsorship indicates, the book addresses an audience of institutional researchers, so each essay addresses essentially the same set of issues that are important to that audience. For example, most essays respond to the Spellings Commission report, and most also address concerns directly or indirectly related to accreditation. The collection is important within the Rhetoric and Composition community for a couple of major reasons. First, the inclusion of voices from our field indicates that the conversations about

assessment are already reaching across disciplinary and institutional boundaries. Yancey, Terry Zawacki, O'Neill, Joseph Janangelo, Huot, Charles Moran, and Anne Herrington are familiar names that represent model programs within our discipline. These people and programs should be informing practice in institutional research, so their presence here is encouraging. Second, we see the emergence of potential partners in the fight to sustain more robust forms of writing assessment. Two essays in particular address writing in engineering, engaging the Accreditation Board for Engineering and Technology (ABET) standards in the process. ABET's writing requirements for students in engineering programs make them a natural partner in pushing for authentic assessments of many kinds and in resisting the increasingly reductive forms of writing assessment coming from the commercial assessment enterprise.

As a whole, then, these volumes provide resources that promote a deeper and broader understanding of the writing assessment landscape, especially as it exists within academe—in writing programs of several kinds, in accreditation efforts, in pushing back against reductive governmental initiatives and even more reductive forms of assessment coming from the commercial side of writing assessment. These collections demonstrate that in the past three decades, we have built a canon, a body of high-quality research and debate that supports our current efforts to promote better, more authentic, more useful and generalizable assessments, and that aids us in preparing the next generation of academics in the field.

#### REGROUNDING WRITING ASSESSMENT

Another set of volumes builds on that substantial base of research by laying out assessment processes from a perspective that includes and promotes more robust forms of writing assessment, forms that move beyond ranking and placing and provide output that can help us engage assessment with instruction in order to improve both. Beginning with Wolcott and Legg's *An Overview of Writing Assessment* (1998), these volumes lay out the case that teachers need to understand the values represented in "the education and psychometric traditions" (O'Neill, et al. 54), without being trapped in their underlying assumptions. Writing assessment, these volumes argue, should rightfully exist within this larger framework of assessment, but it should also engage writing theory so that the constraints of the larger context can no longer exert a reductive pressure on *writing* assessment. The mistake of the early assessment community was to treat writing assessment the same as other kinds of assessment: sorting military enlistees into categories based on general aptitudes, for example. The stakes in making

a transition to better instruments are high. We can neither withdraw from the field, as Patricia Lynne and Maja Wilson argue, nor can we be the sole assessors of writing, as Chris Gallagher favors. The key to our participation, as O'Neill, Moore, and Huot point out, is that “[b]ecause writing assessment is fundamentally about supporting current theories of language and learning and improving literacy and instruction, it should involve the same kind of thinking we use every day as scholars and teachers” (59).

While Wolcott and Legg arguably marked the threshold of this latest phase of the conversation, it was first and most clearly defined by Huot's 2002 work, *(Re)Articulating Writing Assessment for Teaching and Learning*. If we can say that White is the father of modern writing assessment, then Huot has surely become its godparent, at least. Huot's article “Toward a New Theory of Writing Assessment” (an updated version of which is chapter four in this volume) is the defining statement of the second phase of this new conversation, as I argued above, and his name is everywhere in this third phase. Assessment, Huot argues, is a field of study with its own set of methodologies. Assessment should be regarded as part of the research enterprise—it is, in fact, primary research. If writing assessment engages with writing theory, then the assessment practices that emerge will be consistent with the best that has been thought, researched, and written about writing as a construct, as a set of competencies, and as a social practice. This volume sets the expectation that teacherly readings need to be at the heart of writing assessment. Writing, Huot argues, cannot be learned without assessment (165), but assessment, following his argument, cannot be responsible without engaging the contexts for writing, in particular the classroom context. His chapters three and five most clearly contribute to this argument.

Jumping ahead, we can see how this argument engages assessment with learning in O'Neill, Moore, and Huot's *A Guide to College Writing Assessment*, cited above because of its succinct history of writing assessment in the twentieth century. The book's principal mission, however, is the *praxis* of writing assessment—putting theory into practice in placement, proficiency, program, and faculty assessments. These four activities are the bread and butter of writing programs of all kinds, from first-year composition to WAC and WID. If, as Huot asserted in *(Re)Articulating*, the most important element of writing assessment lies in working methodologies (165), then O'Neill, et al. present working methodologies for the most common kinds of assessment for writing programs, and the appendices provide “best practice” examples for scoring rubrics, classroom observations, portfolio-based assessments, surveys, and more. If assessment was once something we needed to do defensively—lest, as White warned, someone do it *to us*—then this book outlines the practice of proactive assessments, assessments

conceived and run by the people inside writing programs, geared to inform us about our own students, programs, processes, and faculty.

In an allied book, Linda Adler-Kassner and O'Neill's *Reframing Writing Assessment to Improve Teaching and Learning*, the authors state, "Our position is that writing instructors and program directors know a lot about writing instruction and need to be centrally involved in discussions about writing assessment" (9). The twentieth century began with a conscious exclusion of writing teachers from the assessment process. Adler-Kassner and O'Neill's assertion, which echoes Huot's in *(Re)Articulation*, is fast becoming the anthem of twenty-first century writing assessment. Excluding us never made much sense, except in the political economy of twentieth-century assessment (see Neal, above), and this set of books acknowledges, first, the hard-won expertise of writing teachers and writing program administrators in the theory and practice of writing assessment and, second, the political contention that without the input of writing experts and without a thorough recognition of the contexts surrounding writing assessments, those assessments simply cannot be valid. Absurdly, the values composition teachers hold and the theoretical bases upon which they operate have rarely—and never really seriously—been part of commercial writing assessment. Noting that omission, Adler-Kassner and O'Neill argue that "not only can contemporary understandings of psychometric theory accommodate composition's frames but that they can also help us connect to the larger frames about education and assessment that operate in the public, which can help shift these larger frames so that they reinforce—or at least accommodate—composition's values, theories, and pedagogies" (71). I would put the issue this way: beginnings matter. The fields of composition and writing assessment began in a context that emphasized deficit and error (recall Hill's response to such writing, cited above). So, the public frame for what we call Composition Studies is that we tend to issues of correctness, that good writing is the same thing as error-free writing. Resolve the problems that result from severely limiting the construct and, these books claim, we can build our own perspective from which to influence and perhaps gain control of the more important conversation about writing, as well as the one about writing assessment.

Sandwiched between *(Re)Articulating* and *Reframing*, Sara Cushing Weigle's *Assessing Writing* (2002), published in the Cambridge Language Assessment Series, bears mention because together with *A Guide to College Writing Assessment*, it provides a thorough survey of writing assessment as conceived on a world-wide basis. Beyond the United States, writing assessment is just one piece of the larger sphere of language assessment, and in that context, those who teach writing are and have long been well

acquainted with the educational and psychometric methodologies for testing writing within the framework of language assessment. For the international assessment community, Weigle's book summarizes current practice and introduces possibilities that go beyond the timed impromptu test. In doing so, she provides a clear, useful summary of the construct ("the nature of writing ability"), of assessment theory ("basic considerations in assessing writing" and "research in large-scale writing assessment"), of the processes of designing and scoring direct tests of writing, and most useful, of classroom writing assessments and portfolio-based writing assessments. Huot has consistently promoted the importance of assessment theory for writing teachers and administrators, bemoaning the fact that "the emphasis in assessment is on practice without adequate attention to theory" (O'Neill, et al. 35). In a body of work extending from *Validating Holistic Scoring* (1993) to *A Guide to College Writing Assessment*, he and his co-authors have insisted, with good reason, that a knowledge of sound writing assessment theory and practice can save us from the evermore reductive assessments that the for-profit assessors try to push upon us. Weigle's book contains just such a synthesis of traditional theories and methodologies of writing assessment, clearly and cogently presented. Taken together, Wolcott and Legg,; Weigle,; O'Neill et al., and Moore and Huot leave no excuse for anyone within the Rhetoric and Composition community to be uninformed about writing assessment.

#### REINTEGRATING ASSESSMENT WITH INSTRUCTION: CLASSROOMS MATTER

As the conversation has shifted to include the voices of teachers, scholars, and administrators of writing, one wonders what a reviewer in 2111 might write of the past century or so of writing assessment. Looking back from that distant vantage point, surely one shift would be the entry of the writing classroom into the writing assessment arena, but another would be the engagement of writing assessment within the writing classroom. Perhaps the work that marks the beginning of that sector of the conversation is Patricia Lynne's *Coming to Terms: A Theory of Writing Assessment* (2004). In one sense, Lynne's book expresses the discipline's frustration with commercial writing assessment. Tired of reductive assessment instruments and the misconceptions about students' learning that those instruments foment, and annoyed at the commercial firms' resistance to more valid forms of writing assessment—portfolios in particular, but other forms of authentic assessment as well—Lynne argues, basically, for boycotting the old testing order, which clearly is not engaged with what happens in classrooms.

Lynne argues that those in charge of constructing writing theory and of teaching writing must throw off the trappings of traditional writing assessments and develop their own models. That argument expresses a frustration that anyone in the field of Rhetoric and Composition has felt, but Lynne's call to reject *all* aspects of traditional writing assessment goes too far when she identifies concepts such as validity and reliability as hegemonic forces of the commercial enterprises that inevitably undermine attempts to establish better assessments. Lynne's call to resist the ever more reductive, ever less valid samples, scored ever more cheaply, is of course completely sensible and a much-needed addition to the past decade's conversation. However, the best tactic for making that argument is not to abandon aspects of responsible assessments that professional organizations within the academy universally accept, but to point out in how many ways the commercial assessments hide behind concepts such as validity, which they themselves violate or ignore outright. O'Neill, Moore, and Huot, citing the 1999 APA, AERA, and NCME Standards, point out that validity is a unified standard that "not only includes all notions of validity and reliability, it also demands that test consequences and implications for the local educational environment be considered" (27). Thus, while ACT may advance statistical correlations to make claims that COMPASS has concurrent or predictive validity, there is no way to demonstrate construct validity for an extremely reductive test such as COMPASS (which is simply an exercise in editing). O'Neill, et al. conclude, "[u]nless decisions based upon a test can demonstrate educational value for students, it is difficult to make a convincing argument for validity" (27). Instead of disregarding tenets of responsible testing, which Lynne promotes, it is more useful to use those tenets, which actually support the more robust forms of assessment that have emerged from the academy, to demonstrate that the commercial tests are unacceptable and irresponsible. Furthermore, deserting foundational concepts like reliability and validity robs us of the ability to compete on a level playing field with those older, more reductive forms of assessment.

As much as I sympathize with Lynne's displeasure over the way numbers have come to dominate writing assessment, I believe we have to acknowledge the place that statistics have in national conversations about writing, and our assessments need to speak back with numbers of our own. For the foreseeable future, if we cannot mount counterarguments grounded in our own responsibly designed assessments, then the commercial assessments will continue to dominate as they did in the twentieth century. This power shift in the conversation has only begun in this decade. We have attained a voice in a conversation from which we were previously excluded. Our presence in the conversation may—*may*—allow us, in time, to change the

face of writing assessment. However, ignoring the widely accepted bases for responsible assessments will not advance our cause. Instead, Lynne's concept of "meaningfulness" as a condition of responsible assessment does help articulate for the public the same statements that Huot and others advance within the scholarly community: that the kinds of writing we assess need to be meaningful to the writers.

Meaningfulness helps the classroom talk back to the commercial assessments, to begin the process of change by helping the public understand the challenges students and teachers face in classrooms, to assert the authority that comes from our own expertise. Bob Broad has been an important voice in taking back control over important aspects of assessment and in framing discussions about standards in ways that are friendly to the writing classroom. Broad's *What We Really Value* (2003) begins a look at the ways the commercial assessment establishment has influenced the making of rubrics. While Elliot begins *On a Scale* with a sense of the hope a teacher like Katherine derived in 1913 from the newly published Hillegas scale, Broad points out that many widely disseminated rubrics, like the commercial assessments for which they guide scoring, have also become reductive. The ubiquitous "Six Traits" rubric, for example, emphasizes surface features of writing in four of its six traits, leaving only two traits for what the writer is trying to say, for what purpose, and to what audience, how the writing is organized, and other traits that have much more to do with the effectiveness of a piece of writing than surface features. Not that spelling, grammar, and mechanics do not matter, but even the most basic understanding of what constitutes good writing would not weight them as two-thirds of success. And the more robust, detailed, and locally generated rubrics that Broad endorses not only provide teachers with a useful tool for grading, but they also provide a far more useful support than "Six Traits" kinds of rubrics for curriculum development, for constructing assignments, for talking with students about their writing, and for responding to and grading student writing.

Broad, et al. extend and enrich that argument in *Organic Writing Assessment: Dynamic Criteria Mapping in Action*. Eight authors in this edited collection present their institutions' experience with Dynamic Criteria Mapping (DCM), complete with practical examples of applications of DCM and key documents—criteria maps and rubrics, of course, as well as surveys, training materials, etc. The collection is important for at least two reasons. First, it documents the ability of DCM to accommodate the additional complexity of local assessments. Authentic assessments are "messy" in a good way, and that messiness challenges the tight controls of more reductive assessment methodologies. The examples here demonstrate

that DCM allows for local assessments tailored to local curriculum and to local programmatic and institutional needs, yet sufficiently rigorous as to yield outcomes that are comparable with other local assessments. In other words, DCM provides a practical alternative to “one size fits all” assessments. Second, the values expressed in these maps and rubrics come from the classrooms in these eight institutions and speak to the larger context of a national conversation about writing assessment—essentially reversing the direction of impetus of the past century or so, beginning with Hillegas’s rubric, in which the national context dictated what happened in classrooms. Portfolio-based assessment implies that the reverse is possible; DCM shows a system for enacting that reversal.

Maja Wilson’s *Rethinking Rubrics in Writing Assessment* tells something of the same story from the standpoint of a secondary English teacher. Wilson’s core argument is that “our assessments should be based on the same assumptions as our pedagogy” (52). Wilson documents a struggle with rubrics—and here the notorious “Six Traits” rubric demonstrates its chokehold on secondary English teachers. Elliot’s 2004 version of Katherine strives to help her students understand the rubrics being used to rate their writing, downloading rubrics and writing samples from the Internet. Wilson unveils the extreme limitations of that move, illustrating in her own practice how such materials limit the teaching of writing in ways that diminish writing itself to the small portion of the construct that six-trait rubrics and timed essay tests can reach. Instead, Wilson struggles to give up rubrics in favor of making writing meaningful to her students, trusting that engaging in meaningful writing will lead students to develop the skills needed to pass the state-mandated tests. Her work points out the basic discord between the state tests, which view writing as a discrete set of skills, and a sound writing curriculum that recognizes that writing is far more than a set of skills. From the classroom, Wilson talks back to large-scale testing, clarifying the ways that such testing mounts huge obstacles to learning to write in the fuller sense. Writing, as a full construct, cannot be captured on a rubric. Only a reductive sense of writing can be judged in that way. Given that fact, we have a responsibility to develop rubrics that are as robust and non-reductive as possible, following the DCM model.

As writing moves into electronic environments, the construct becomes more, not less complex, raising Broad’s and Wilson’s arguments to an even higher plane. E-mail, social networking spaces, websites, multimodal compositions, all these and more engage writers in more complex ways than mere words on paper. Michael Neal takes on this scenario of writing’s future in *Writing Assessment and the Revolution in Digital Texts and Technologies*, a volume in Columbia University Teachers College’s Language and Literacy

series. Neal echoes Broad's view, in *What We Really Value*, that modern writing assessment has gone astray. As Broad writes, "Rather than seek to understand and carefully map out the swampy, rocky, densely forested terrain of writing assessment found lying before them, they [the large-scale commercial testing corporations] quickly moved to simplify and standardize it" (5) which, Neal observes, "they did by defining seven characteristics of writing that could be isolated and measured" (63). In describing writing assessment *as* a technology, Neal provides not so much a history as an overview of the ways that the current technology, constructed by agencies outside the teaching enterprise, reduces the construct *writing* to the point that teachers and learners of writing can no longer recognize the construct being tested *as* writing. Looking to the more recent past, Neal discusses paper-based and electronic portfolios as complicating that process, though he also provides descriptions of the ways commercially available eportfolio systems actually reduce the possibilities for students to exhibit electronic texts—an irony of epic proportions, and one that completely disqualifies such systems from being legitimate eportfolios in much the same way as reductive assessments like COMPASS cannot meet the validity standards to qualify as systems that test writing. Neal's solution is, in part, a common trope among writing teachers who work with new technologies: rather than wait for others to design the technologies for future writing assessment, teachers need to become involved in the design and construction of these spaces. That many in the profession are already participating in that way seems to have escaped Neal's attention, as he does not mention Texas Tech's ICON system, or the University of Georgia's EMMA; nor does he provide much information on the locally grown eportfolio systems coming out of Barbara Cambridge and Yancey's eportfolio initiatives or Trent Batson's Association for Authentic, Experiential, and Evidence-Based Learning. Still, Neal's volume provides a useful look at the ways new technologies—and in particular new genres of writing in electronic spaces—complicate the construct *writing* and, in turn, challenge existing models for assessing writing.

Overall, this decade of the conversation on writing assessment provides a great deal of encouragement and hope about the future. The commercial assessment industry is engaged in a race to the bottom as it promotes tests that are less and less valid—so much so that the general public is beginning to see the inadequacy of their tests, as evidenced by states such as Oregon and Washington, which have de-emphasized or eliminated such tests. Teachers, scholars, and administrators of writing can stop this devolution by affirming a common thread among the voices represented above: that "our assessments should be based on the same assumptions as our pedagogy" (Wilson 52; echoed in Huot; Neal; O'Neill, et al.; Adler-Kassner

and O'Neill; and Broad). If we are at a tipping point in the conversation, we have arrived there by understanding more clearly what we are about when we teach writing. Beginning with Janet Emig's *Composing Processes of 12<sup>th</sup> Graders*, our field has pushed writing theory forward. Through the decades since, our understanding of the construct *writing* has grown more sophisticated, more complex, more substantiated, and more complete. As that theoretical framework has advanced, so have the assessments *we* have designed. Therefore, we now have a set of practical alternatives to what is happening on the commercial side. As late as 1943, John Stalnaker, writing for the CEEB, wrote,

The type of test so highly valued by teachers of English, which requires the candidate to write a theme or essay, is not a worthwhile testing device. Whether or not the writing of essays as a means of teaching writing deserves the place it has in the secondary school curriculum may be equally questioned. Eventually, it is hoped, sufficient evidence may be accumulated to outlaw forever the "write-a-theme-on"...type of examination. qtd. in O'Neill, et al. 20).

While we can join Stalnaker in his final wish, we do so for different reasons. In the years since Stalnaker, the fields of Rhetoric and Composition, WAC, Cognitive Psychology, and Education have advanced the importance of writing throughout school and university curricula, and of course no one today would question the need for students to write in order to improve their writing, or to evaluate their writing abilities. The CEEB tests, which in 1943 had devolved to indirect tests of grammar and usage, vocabulary, analogies, etc, were that era's race to the bottom—Stalnaker was responding to teachers' objections over the elimination of writing from the CEEB's tests. That the CEEB succeeded in 1943 shows how little regarded were the opinions of teachers. Today, our knowledge of the construct has grown so that no one can responsibly suggest that a writing assessment should not involve writing. Today's devolution toward less and less valid tests of writing ability—including machine scoring of students' writing—occurs in a different context. Today, we know that a major reason timed essay tests produce bad writing is that they ask students to write in circumstances that are not conducive to good writing. And we know that focusing on the deficits and errors in those low-validity, context-poor samples leads to curricula that are focused on deficit and error, and are therefore not sufficient to support students' efforts to become better writers. Thus, as this review of a decade of writing assessment scholarship reveals, we are ready with practical alternatives, assessment instruments that meet the unified standard of validity, that provide far more useful information than merely a ranking

or a placement, and that are grounded in a much more advanced understanding of what writing is and of what is required to further the cause of improving writing instruction. Part of that progress is the need for better assessments, and the past decade is evidence that the conversation is tipping in that direction.

## WORKS CITED

- Anson, Chris, Les Perelman, Maya Poe, and Nancy Sommers. "Symposium: Assessment." *College Composition and Communication* 60.1 (2008): 113-164. Print.
- Bailey, Richard W. and Robin Melanie Fosheim. *Literacy for Life: The Demand for Reading and Writing*. New York: MLA, 1983. Print.
- Belanoff, Pat and Peter Elbow. "Using Portfolios to Increase Collaboration and Community in a Writing Program." *WPA: Writing Program Administration* 9 (Spring 1986): 27-39. Print.
- Broad, Bob. *What We Really Value: Beyond Rubrics in Teaching and Assessing Writing*. Logan: Utah State UP, 2003. Print.
- Calfee, Robert and Pamela Perfumo. *Writing Portfolios in the Classroom: Policy and Practice, Promise and Peril*. Mahwah: Erlbaum, 1996. Print.
- Cambridge, Barbara, Susan Kahn, Daniel P. Tompkins, and Kathleen Blake Yancey, eds. *Electronic Portfolios: Emerging Practices in Student, Faculty, and Institutional Learning*. Washington, DC: American Association for Higher Education, 2001. Print.
- Condon, William. "Looking Beyond Judging and Ranking: Writing Assessment as a Generative Practice." *Assessing Writing* 14.3 (2009): 141-156. Print.
- Cronbach, Lee J. "Five Perspectives on Validity Argument." *Test Validity*. Ed. Howard Wainer and Henry Braun. Hillsdale: Erlbaum, 1988. 3-17. Print.
- Daiker, Donald A., Jeffrey Sommers, Gail Stygall, and Laurel Black. *The Best of Miami's Portfolios*. Oxford: Miami U, 1990. Print.
- Emig, Janet. *Composing Processes of Twelfth Graders*. Urbana: NCTE, 1971.
- Gallagher, Chris W. "At the Precipice of Speech: English Studies, Science, and Policy (Ir)relevancey." *College Composition and Communication* 73 (2010): 1, 73-90. Print.
- Gould, Stephen Jay. *The Mismeasure of Man*. New York: Norton, 1996. Print.
- Hamp-Lyons, Liz and William Condon. *Assessing the Portfolio: Principles for Practice, Theory, and Research*. Cresskill: Hampton P, 2000. Print.
- Haswell, Richard. "The Two-Tier Rating System: The Need for Ongoing Change." *Beyond Outcomes: Assessment and Instruction Within a University Writing Program*. Ed. Richard Haswell. Westport: Ablex, 2001, 39-52. Print.
- Huot, Brian "Toward a New Theory of Writing Assessment." *CCC* 47 (1996); 549-66. Print.
- Messick, Samuel. "Validity." *Educational Measurement*. Ed. R.L. Linn. 3rd ed. New York: Macmillan, 1989. 13-103. Print.
- Moss, Pamela. "Can There be Validity Without Reliability?" *Educational Researcher* 23.4 (1994): 5-12. Print.

- Nye, David E. "Technological Prediction: A Promethean Problem." *Technological Visions: The Hopes and Fears that Shape New Technologies*. Ed. Marita Sturken, Douglas Thomas, and Sandra J. Ball-Rokeach. Philadelphia: Temple UP, 2004. 159-176. Print.
- Sacks, Peter. *Standardized Minds: The High Price of America's Testing Culture and What We Can Do to Change It*. Cambridge: Perseus, 1999. Print.
- Shermis, Mark and Jill Burstein, eds. *Automated Essay Scoring: A Cross-disciplinary Perspective*. London: Routledge, 2002. Print.
- Smith, William L. "The Importance of Teacher Knowledge in College Composition Placement Testing." *Reading Empirical Research Studies: The Rhetoric of Research*. Ed. John R. Hayes. Norwood: Ablex, 1992. 289-316. Print.
- Commission Appointed by Secretary of Education Margaret Spellings U.S. Department of Education. *A Test of Leadership: Charting the Future of U.S. Higher Education*, 2006. Web. <http://www2.ed.gov/about/bdscomm/list/hied-future/reports.html>
- Villanueva, Victor, ed. *Cross-Talk in Comp Theory: A Reader*. 2<sup>nd</sup> ed. Urbana: NCTE, 2003. Print.
- White, Edward M. *Assigning, Responding, Evaluating*. 4<sup>th</sup> ed. New York: Bedford, 2007. Print.
- "The Opening of the Modern Era of Writing Assessment: A Narrative." *College English* 63:3 (2001): 306-320. Print.
- Willard-Traub, Margaret, Emily Decker, Rebecca Reed, and Jerome Johnston. "The Development of Large-Scale Portfolio Placement Assessment at the University of Michigan: 1992-1998." *Assessing Writing* 6.1 (1999): 41-84. Print.
- Williamson, Michael M. and Brian Huot, eds. *Validating Holistic Scoring: Theoretical and Empirical Foundations*. Cresskill: Hampton P, 1993. Print.
- Yancey, Kathleen Blake. "Looking Back as We Look Forward: Historicizing Writing Assessment." *CCC* 50 (1999): 483-503. Print.

