Title

Forum: Issues and Reflections on Ethics and Writing Assessment

Permalink

https://escholarship.org/uc/item/7076v8f0

Journal

Journal of Writing Assessment, 9(1)

ISSN

1543-043X

Authors

Elliot, Norbert
Slomp, David
Poe, Mya
et al.

Publication Date

2016

Copyright Information

Peer reviewed

by Norbert Elliot, **David Slomp**, **Mya Poe**, **John Aloysius Cogan Jr.**, **Bob Broad**, and **Ellen Cushman**

We hope this special issue adds to the body of knowledge created by the writing studies community with respect to the opportunities that can be created when assessment is seen in terms of the creation of opportunity structure. This hope is accompanied by a reminder of our strength as we annually encounter approximately 48.9 million students in public elementary and secondary schools 20.6 million students in postsecondary institutions (Snyder & Dillow, 2015). Our influence is remarkable as we touch the lives of many, one student at a time.

1.0 The Old Road is Rapidly Agin'

(Please Get Out of the New One if You Can't Lend Your Hand)

Another radical change is to come, beginning now and continuing through 2060. This time, the change will be demographic and will come in two forms: educational enrollment and electorate evolution. Preparing for this change has little to do with abrupt exigence and everything to do with principled action.

In its most recent projections, the U.S. National Center for Educational Statistics (NCES, Hussar & Bailey, 2013) estimates that public school K-12 enrollments are projected to be to 5 and 6 percent lower for American Indian/Alaska Native and White students, respectively, in 2022 than in 2011 and a moderate 2 percent increase for Black students during the same time period. For Hispanic, Asians/Pacific Islander, and other students of two or more races, enrollment changes are expected to be more dramatic. In terms of increase against the 2011 benchmark, there will be a 20 percent rise for students who are Asian/Pacific Islander; a 33 percent rise for students who are Hispanic; and a 44 percent rise for students who are two or more races. The number of high school graduates is projected to be 2 percent lower in 2022 than in 2009–10. Specifically, there will be a 29 percent decrease between 2009–10 and 2022 for students who are American Indian/Alaska Native; a 16 percent decrease between 2009–10 and 2022 for students who are White; and a 14 percent decrease between 2009–10 and 2022 for students who are Black. Conversely, there will be a 23 percent rise by 2022 for students who are Asian/Pacific Islander and a 64 percent rise by 2022 for students who are Hispanic.

Along with these shifts for K-12 students, NCES predicts similar race/ethnicity changes in postsecondary enrollment. For American Indian/Alaska Native students, enrollment is projected to be about the same in 2011 and 2022. Against the 2011 benchmark, there will be a 7 percent increase for both White and Asian/Pacific Islander students by 2020. For Black students there will be a 26 percent increase, and for Hispanic students there will be a 27 percent increase.

To document the challenges posed by the rapid demographic evolution in the U.S. from the 1970s to 2060, the Center for American Progress and the American Enterprise Institute have come together to project the race/ethnicity composition of every state to 2060 (Teixeira, Frey, & Griffin, 2015). The demographic changes they project are so significant—from 80 percent White citizens in 1980 to 44 percent in 2060—that they classify these shifts as superdiversification. During that same period, Hispanic citizens are projected to grow from 6 percent to a projected 29 percent; Asian Americans from 2 percent to 15 percent. Only Black citizens will remain stable at 12 percent to 13 percent from 1980 to 2060.

Mya Poe: Reflection on Identification of the Least Advantaged

Demographic changes are convenient data tabulations but often miss identification of the least advantaged. When the least advantaged are not seen, they cannot seek a remedy.

The case of *Bronson v. Board of Education of City School District* usefully illustrates this point. In the *Bronson* court challenge, which spanned from 1974-1984, plaintiffs sought to address de facto segregation as well as the closing of a predominantly African American school in the low income neighborhood of Over the Rhine in Cincinnati, Ohio. The

school served not only African American children (82%) but also white Appalachian students (15%). In their lawsuit, the plaintiffs who were African American and Appalachian argued that school closure "discriminated against low income Blacks and Appalachians for violation of 42 U.S.C. § 1983"—i.e., Title VI of the Civil Rights Act of 1964.

The school district argued Appalachian persons were not within those classes the Civil Rights Act of 1964 was designed to protect. The judge agreed with the defendants and ruled that the plaintiffs did not show Appalachians differ "in some unspecified manner from other groups."

The *Bronson* case illustrates the difficulty in identifying "least advantaged." Appalachians are not merely poor white

people. Although the *Bronson* plaintiffs argued from the point of view that Appalachians are poor whites, Appalachian people are a diverse racial/ethnic group with a distinct cultural identity. However, legal structures ignore cultural identity in favor of racial identity. As such, Appalachians have no legal identity under federal civil rights laws. In the end, as the case of Appalachian people demonstrates, identification of least advantaged is neither a straightforward process nor one that is easily supported by reporting/documenting approaches to fairness.

These trends are but part of global trends. As the United Nations (2015) reported in its annual review of world population prospects, substantial and enduring economic and demographic asymmetries among countries are likely to remain powerful generators of international migration. Projections suggest that, between 2015 and 2050, the top net receivers of international migrants (more than 100,000 annually) will be the United States of America, Canada, the United Kingdom, Australia, Germany, the Russian Federation, and Italy. India, Bangladesh, China, Pakistan, and Mexico will have net emigration of more than 100,000 annually.

That such demographic shifts will be associated with cultural shifts is a certainty. Presently, the world population stands at 7.3 billion, an addition of approximately one billion over the last twelve years. In high income countries, net migration is projected to account for 82 percent of population growth. While no one is certain of the extent of such overwhelming change, that which has served poorly for many in the past will be intolerable for all in the future.

Bob Broad: Reflection on Quantity and Quality

"…the overload is unbearable"

Few could argue with this point that U.S. students are over-tested. However (as I argue in "This is Not Only a Test…"), the main problem is not the quantity of assessments, but rather the adverse impact of even one mass standardized test on the *quality* of students' educations. Good assessments (e.g., robust writing portfolios) motivate and guide the best teaching and learning; requiring lots of those kinds of assessments (within reasonable limits) benefits the educational project at the heart of democracy. The main problem with the status quo is that standardized tests, by their nature, undermine and corrupt the quality of education generally, and *especially* education in writing studies.

Among the list of systems touted to yield universal advancement—processes that, in fact, would have been scrapped generations ago had they served White folks as poorly as they had served others—is educational testing. In the U.S., a report by the Center for American Progress documents that the overload is unbearable (Lazarin, 2014). Among the findings: Students in grades 3-8 in which U.S. federal law requires annual testing take, on average, about ten large-scale, mandated tests throughout the year; students in grades K-2 and 9-12 who do not take or are less frequently tested using federally required state exams take approximately six tests in a year; and urban high school students take three times as many district level tests, and they spend up to 266 percent more time taking them compared with suburban high school students. While students are tested as frequently as twice per month and, on average, once per month, the report finds, many districts and states may be administering tests that are duplicative or unnecessary. "There is a culture of testing and test preparation in schools," the report concluded, "that does not put students first" (p. 19). While it is commonly assumed that these tests are mandated under federal control, the report found many of the tests are under state and district control.

Bob Broad: Reflection on Education in Rhetoric

"…however small and restricted…."

Here (as usual) I am less modest than Norbert Elliot. In the tradition leading from the Greek sophists to Jon Stewart and Frank Lutz, I submit that powerful education in *rhetoric* (the arts of critical and creative consumption and production of public texts) is crucial to our abilities to address social and political problems, whether local or global. This belief drives my passionate concern for the effects of both good and bad writing assessment.

2.0 The Order is Rapidly Fadin'

(And the First One Now Will Later be Last)

Projections are just that: calculations, often using techniques of exponential smoothing and multiple linear regression, based on trend assumptions that do not include interventions. Superdiversification on an international scale will not necessarily result in increased opportunity. While the flood of issues facing nations is overwhelming, writing assessment remains an international field and, however small and restricted, has a role to play in responsible action.

In this special issue, the authors have taken action in a very specific area of research: the assessment of writing ability. In planning for the special issue since we first presented our ideas at the Sixty-Sixth Annual Convention of the Conference on College Composition and Communication (Cushman, 2015), we recognized the significance of preparing for the demographic shifts to come —shifts that will change received views of life. As we reflected on the consequences of our work as researchers interested in written forms of literacy in preparation for this special issue, we realized that a turn to ethics, not solely to the foundational principles of validity and reliability associated with educational measurement, would help us reconceptualize the nature of our work. Efforts were collaborative, though distinct. For those of us who work in quantitative assessment (Elliot and Poe), we sought a way of thinking about decisions that would serve everyone equally well through techniques associated with the examination of bias, and we hoped to ensure that the potential for disparate impact of past practices did not continue in the future. For those of us who are qualitative assessment specialists (Slomp and Broad), we wanted to identify methods that would help us better understand the writing construct and ensure educational equity and quality through that understanding. The legal scholar among us (Cogan) uses the disparate impact approach to highlight the capacity of legal doctrines and methods to inform and enrich the research of non-legal scholars. Disparate impact analysis, which is used to identify and remediate unintentional discrimination, offers non-legal scholars a method to enhance educational opportunity and access while addressing issues of fundamental fairness in high-stakes testing. For those of us who are researchers in language preservation (Cushman), we wanted to identify imperialist legacies of knowledge as they are often associated with validity, and we wanted to identify ways of celebrating multiplicity that resulted in more than token adjustments in writing assessment.

Bob Broad: Reflection on A Role for Qualitative Inquiry

Our profession is already well equipped to design and carry out writing assessments that promote best practices in teaching and assessing writing (starting with robust constructs of what civic and professional writing involve). Both Elliot's and Slomp's articles in this special issue describe in considerable detail approaches to designing writing assessments that are just and ethical because they attend carefully and continuously to the educational consequences that follow from those assessments. More significantly, the assessments Slomp and Elliot describe are what Wiggins (1998) called "educative assessments": assessments designed with the explicit purpose of guiding, shaping, and motivating the best pedagogy and learning.

My training in the early 1990s as a qualitative researcher coincided with my training in the field of rhetoric and composition; the combination of the two spurred my interest in writing assessment as a sub-field. In reading the articles in this special issue, I'm struck by how seamlessly fused just writing assessment and qualitative inquiry can be, and how potentially productive that fusion is for the project of promoting ethics in writing assessment.

However, I can't possibly cheer and support the direction in which the other contributors to this issue are pointing us better than Pamela Moss did twenty years ago in her article, "Enlarging the dialogue in educational measurement: Voices from interpretive research traditions":

> I argue that the theory and practice of educational measurement would benefit by expanding the dialogue among measurement professionals to include voices from research traditions different from ours and from the communities we study and serve: It would enable us to support a wider range of sound assessment practices, including those less standardized forms of assessment that honor the purposes teachers and students bring to their work... (Moss, 1996, p. 20)

We need to take note of something in Moss's exhortation I believe will prove crucial to the current project of developing a new ethics in writing assessment. Not only did she urge the educational measurement community to explore and make use of what she elsewhere calls "interpretive traditions" (which many of us would call "qualitative inquiry"), she also calls for researchers in the measurement community to *listen* to, engage in dialogue with, and reproduce "voices… from the communities we study and serve." The combination of interpretive methods and a democratizing impulse to listen (two methods which I would argue are actually of a piece) fits perfectly with the theories and methods put forth in this journal issue. In my article in this issue, I explore at length the difficulties test-makers have demonstrated both in *seeing* the educational consequences of their products and also *listening to* the

voices of those teachers and students most affected by mass-market evaluations. Qualitative methods are, in one view, highly disciplined systems for promoting listening, and especially listening to those voices (students, teachers, and others) who are often ignored in educational decision-making.

Without going deeply into the history of psychometrics and its role in writing assessment in the U.S. in the 20th and 21st centuries, I believe it's fair to say that psychometrics attempted to render writing assessment simple, stable, and objective by importing into the realm of textuality many quantitative tools and concepts from the physical sciences. If measurement specialists could take the messy, complex, socio-linguistic terrain of textual production and consumption and render it clear, clean, and flat (I picture a jungle cleared with a bulldozer), they could solve many of the ethical and political problems they saw with writing assessment as practiced by educators: variability of designs, unarticulated evaluative frameworks, variability of judgments among teachers, uncertain connections between curricula and assessments, etc. (for an extended discussion on the historical perspectives on writing assessment see Yancey [1999], Elliot [2005] and Huot, O'Neill, and Moore [2010]).

Psychometricians' most powerful tool for achieving "objectivity" in writing assessment was inter-rater reliability, the remarkable accomplishment of getting evaluators to agree in their judgments of students' writing. Unfortunately, the all-consuming drive to establish inter-rater reliability led to two disastrous outcomes. First, it compelled test designers to design out of their tests nearly all the dimensions of rhetorical situations writers and educators prize most highly, including (but not limited to):

- the writer's choice of and interest in her topic resulting in an authentic sense of mission and purpose
- composing processes that occur over weeks and months
- experimentation with and creative transition from one genre and/or composing technology to another
- research, response, and revision, and
- writing for actual audiences who care about what has been written.

Second, the drive for inter-rater agreement led to numerous cases of cheating and fraud within the testing industry (Farley, 2009) to deliver the required statistical evidence of objectivity.

The other problem with the psychometric approach, of course, is that the writing, reading, and evaluation of texts never has and never will obey the laws of Newtonian physics. (Kenneth Gergen has argued something similar regarding the field of experimental psychology.) Understanding human beings and the things they do with language (including writing assessment) requires traditions of inquiry that can cope productively with the interplay of multiple perspectives, meanings, and values that characterize human society and language. As Moss pointed out long, long ago, qualitative research provides these traditions, tools, and concepts. Qualitative inquiry (derived from ethnography) features a disciplined reverence for the validity of various people's lived experience. It treats those people's experiences as real, significant, and worthy of disciplined study. It seeks to understand how people understand their life-worlds. Moss observed: "…social scientists… must understand what the actors—'from their own point of view'—mean by their actions" (1996, p. 21). This subordination of the researcher's worldview to that of the research participants may provide a pathway to achieving the "decolonial validation" for which Cushman calls.

Also very much in harmony with Cushman's discussion of decolonial writing assessment is Maddox's (2014) ethnographic study of "literacy assessment, camels, and fast food in the Mongolian Gobi." Exploring how "standardized literacy assessments travel globally," Maddox participates in a larger project called for by Hamilton (2012). In Maddox's paraphrase, Hamilton "proposed a new direction of research, studying literacy assessment regimes *from the inside* [emphasis added], including ethnographic investigation into the institutional processes of knowledge production, 'procedures of evidence gathering' and how assessment is 'enacted at the local level'" (p. 474). Understanding writing assessment practices (or anything else) "from the inside" requires qualitative research methods.

One of my contributions to the field of writing assessment is dynamic criteria mapping (DCM), a "streamlined form of qualitative inquiry" proposed in *What we really value: Beyond rubrics in teaching and assessing writing* (Broad, 2003) and developed and refined by the co-authors of *Organic writing assessment: Dynamic Criteria Mapping in action* (Broad et al., 2009). In terms of its intellectual roots and soil, DCM can be usefully understood as a hybrid of grounded theory methods (Glaser & Strauss, 1967; Strauss & Corbin, 1994; Charmaz, 2014) and Guba and Lincoln's *Fourth-generation evaluation* (1989). Elliot and Slomp (this issue) implicitly and explicitly point to DCM as a potentially useful tool in the pursuit of ethical writing assessment.

When Elliot (this issue) notes that theories of fairness must (among other things) "account for stakeholder perspectives," he invokes *Fourth-generation evaluation* and the elements of DCM that emulate Guba and Lincoln's (1989) and Inoue's (2004) multi-perspectival approach to assessment. And when Elliot, following Beason, notes how

"empirical research and assessments are ethical obligations" because they "help us be informed enough to determine what a campus community considers valuable," he helps open the way for the empirical and inductive (Broad, 2012) methodological mainstays of qualitative inquiry. Meanwhile, Elliot's "insistence on localism" resonates strongly with the goal of DCM to get evaluators to generate "organically and locally-grown" (as well as empirical and inductive) accounts of how they evaluate texts.

Elliot also quotes scholars who understand

> how learning is built around "recurrent linguistic, cultural, and substantive patterns" in which individuals, as part of communities, become attuned to patterns—and thus, over time, prove capable of creating meaning through participative structures.

Here Mislevy and Duran eloquently describe *coding*, the key technique by which qualitative researchers (such as Slomp and the teachers whose work he describes) take raw discursive data and sift it methodically to find patterns and themes. This intensive, disciplined work of coding is at the heart of grounded theory and other qualitative methods' ability to generate valid findings from previously chaotic piles of data.

Slomp's description of his teacher/co-researchers provides a key link to qualitative methodology in that he implicitly seeks out and values the sorts of difference and diversity more experimentalist methods purposefully avoid in pursuit of standardization and objectivity:

> teachers involved in this project represent a range of schools (urban, rural, distance learning). Teachers involved in the program are all at a range of stages of their careers, from beginning teachers, to experienced vice-principals.

Slomp notes that "teachers will engage in… dynamic criteria mapping" (52). Yet his research group's work in Alberta significantly surpasses most DCM and other projects of qualitative inquiry because they include students (the most profoundly affected assessment stakeholders of all) as co-researchers and decision-makers in their integrated design and appraisal framework (IDAF). These researchers boldly extend the democratic principle of listening by methodically broadening the groups of stakeholders to which we attentively listen, constantly seeking out new truths about writing assessment and what makes it educatively beneficial and fair.

To prevent misunderstandings, I wish to emphasize that much valuable work in writing assessment can, does, and should involve quantitative, number-based inquiry and analysis. Poe & Cogan's article in this special issue vividly exemplifies this fact, and Haswell (2012) has repeatedly reminded our field of the value and importance of serious quantitative inquiry.

But we did not want the censures or threats that often accompany codes of practice attendant to discussions of ethics. We also did not want to call for goals that called for unreachable aims, resplendent with language calling for the-highest-possible-degree-of-moral-conduct. In short, reason must be wedded to the human experience.

With Rachels (1986), we found minimalistic concepts helpful in providing a picture of the conscious moral agent identified in the introduction to this special issue: one whose conduct is guided by reason while giving equal weight to the interests of each individual affected by that conduct. Because we do not hold that assessment itself can right structural inequality, our aims for assessment are modest—and so is our language to describe those aims. Because inflated language often takes the place of praxis, we found ourselves aligned with what Peter Elbow (2012) has called good enough assessment—a pragmatic approach that sidesteps dead-end conflicts and attends to specific instances. With Elbow, we are all "trying to side-step the theoretic impasse between a positivistic faith in measurement or assessment and postmodern skepticism about any possibility of value in measuring, testing, or scores" (p. 306).

David Slomp: Reflection on Structured Ethical Blindness

This is not to say these distinctions are unimportant. Our field's great strength, and its enduring weakness, is that it is informed by a broad array of scholarship from across a range or research traditions, philosophical foundations, and theoretical positions. It is important to recognize that not all viewpoints are commensurate with one another. The challenge for our field, as we move toward a more ethical stance, is to be more explicit and transparent about the

foundational beliefs and assumptions that guide our work, and to grapple with the implications of these assumptions for our practice.

As both Elliot and Broad observe in their reflections within this forum, and as Cushman so eloquently argues in her article for this SI, writing assessment researchers need to understand deeply how our research traditions provide both lenses and blinders to our work. From an economic standpoint, Broad makes the case that structured ethical blindness prohibits testing corporations from recognizing the damage their tests do to students, schools, and systems of education. In Canada, we have a different tradition: Our provincial governments are responsible for the design and use of large-scale assessments. In most jurisdictions these tests are created and administered in house. Turning a profit is not the goal of government and so structured ethical blindness, as Broad describes it, does not explain why negative consequences stemming from the use of large-scale literacy assessments in Canada (Slomp, Corrigan and Sugimoto, 2014) do not get addressed. I would argue, as Broad does, that those involved in developing and administering large-scale literacy assessments in Canada are genuinely motivated by their care for students but that they are blinded by theory, ideology, politics, history (the list could go on) to the harm their tests exert.

In the Masters of Education in Curriculum and Assessment program that I co-direct at the University of Lethbridge, we tackle this issue directly. Our students are compelled to explore how theories of curriculum and assessment both support and challenge one another. An important jumping off point for these discussions is Lori Sheppard's (1992) excellence article, "The role of assessment in a learning culture"—based on her AERA presidential address—which explores how the divide between curriculum and assessment theories have driven a wedge between large-scale assessment practices and classroom pedagogies, a wedge that endures to this today. A failure to attend to these differences, I fear, will only continue to perpetuate this divide.

For this reason, the theory of ethics we are developing in this special issue emphasizes the importance of integration. As we develop this theory further we need to explore in greater detail how points of tension can best be harnessed in efforts to move this project forward.

Along with pragmatism and principled inquiry, we also wanted to tie assessment with instruction in a very specific way. While we realized that assessment creates barriers, our experiences revealed that it also creates opportunities. Turning to the concept of opportunity structure (Merton, 1938, 1996) allowed us to investigate the roles for individuals created by assessment—and the frustration that occurs when opportunity is denied—and then link that important concept to opportunity to learn. Establishing links between writing assessment and opportunity structure allows long-standing calls to connect assessment, instruction, and learning (White, 1985) to be deepened in very specific ways.

The research specialization of writing assessment now finds itself in a situation similar to that that of Opportunity to Learn (OTL) researchers in 1999. During that time, the "Opportunity of Testing" Project became the topic of conversation between Diana C. Pullin and Pamela A. Moss in a Montreal coffee shop. As interest gathered for the project, it soon became clear that OTL was at the heart of the matter regarding the consequences of testing. Following a series of meetings between 2002 and 2005, an edited volume—*Assessment, Equity, and Opportunity to Learn*—was published by Cambridge University Press. This book now stands as a classic in examining relationships between learning and assessment and serves as a guide to direct the consequences of assessment for the policy, educational research, and psychometric communities.

Because OTL may be somewhat unfamiliar, it is worth establishing its parameters. OTL begins with a theory of learning informed by social and cognitive perspectives. Therefore, learning is not understood as the acquisition of information and skills that will, somehow, serve students across time and circumstance; rather, learning is understood as occurring within a complex ecology of interaction in which conceptual understanding, reasoning, and problem solving in a particular domain are the aims of instruction. In order to provide OTL experiences, educational leaders must ensure that the ecology is a dynamic one, a process that includes establishing a range of activities including productive relationships within and beyond the site of formal education, innovative curriculum design, and assessments that bridge boundaries within and beyond the classroom.

Mya Poe: Reflection on Fairness as Access *and* Consequence

September 1976: My first day of first grade in the Mount Healthy school district. My six-year self could not find the bus among the rows and rows of busses that would take me home at the end of the day. I walked up and down the rows, becoming increasingly panicked at being left behind. I stopped to ask for help.

"Where do you live?"

"I live on Hudepohl Lane."

"Oh, you ride #63, the black bus."

It was my introduction to the valuation of racial identity in education (See *Tina Deal v. Cincinnati Board of Education*). At the height of the Cincinnati bussing crisis and the push for desegregation (see *Mona Bronson, et al. v. Board of Education of the Cincinnati School District of the City of Cincinnati, et al.,* Case No. C-1-74-205), vehicles transporting children were the Civil Rights battleground. It was a lesson that would inform my educational experience, evidenced in hallway conversations, parent and teacher interactions, and through the evaluative mechanisms of the U.S. educational system. Desegregation happened, but racial equality did not. And assessment regimes from the pre-Civil Rights era remained unchanged. That system would leave only three African American students in AP courses by the time we reached 12th grade; only two African American students in a standardized testing "talent search" program for early college admission; and a stunning silence of African American student voices in leadership positions—in a school district where 50% of the students were students of color.

In the end, is access—i.e., equality as process—sufficient? No. Equality as a result is also needed to account for the way that the racial valuation of busses and other educational resources limits the outcomes for children of color. The lesson here for fairness: A focus only on "access to construct representation" cannot bring about fair assessment. Fairness—conceptually and methodologically—must be bolder. It must account for access *and* consequence. It is a lesson that Cincinnati Civil Rights activists such as Marian A. Spencer have spent decades trying to teach us.

On the subject of assessment, OTL has developed an elaborated position. Realizing that traditional assessments are out-of-touch with social cognitive perspectives, assessment is advanced to support teachers and other educational professionals in their desire to support student learning. Recognizing that assessment protocols do more than provide information, OTL scholars view assessment programs as occasions to shape understanding about learning itself and the environments that support learning. In this reconceptualized version of assessment—one that informs this special issue—assessment is viewed as a way of assuring that evidence regarding instruction questions is attendant to the ways OTL may inform our views regarding the foundations, operations, and applications of measurement and the decisions we make as a result of those views. This perspective means that questions arising from instructional design are more important than demands for accountability. As practices that shape learning, assessment thus becomes both formal and informal. "All of this suggests" as Edward H. Hartel, Moss, Pullin, and James Paul Gee (2008) concluded, "that developing useful assessment practices—practices that function productively at different levels of the educational system—will depend on richly contextualized understandings of what information is needed, how it is used, and the effects of this use" (p. 10).

For our purposes, identifying and ensuring OTL served to support the aim of fairness in writing assessment. We recognized fairness is a contested term, and we acknowledge there is no magic definition. We nevertheless hold fairness must look beyond construct representation to consequences and social validation. To achieve this aim, the theory of ethics for writing assessment advanced fairness in a very specific way: *as the identification of opportunity structures created through maximum construct representation*. Along with emphasis on cognitive, interpersonal, intrapersonal, and physiologic domains, this link to construct representation allowed identification of resources: Constraint of the writing construct is to be tolerated only to the extent to which benefits are realized for those who may be disadvantaged by the assessment. As the theory demonstrates, score disaggregation by group allows a robust identification of those who may be disadvantaged; this disaggregation also allows OTL to support and strengthen the existing agency of those who must succeed if society is to benefit. The boundary, order, and foundations accompanying the theory demonstrate ways to make this aim a reality.

To return to the idea of "good enough" raised by Elbow, we add the phrase "for what" to emphasize aim. If asked good enough for what, our answer is straightforward: The planned writing assessment must be good enough to identify opportunity for students created by maximum construct representation. If such robust representation of the writing construct is not possible? Then benefits must be realized for those least advantaged by the assessment. If these benefits cannot be allocated? Then do not use an episodic test at all and plan to use other measures of student ability based on longitudinal observation. The chain of causal logic is clear, and its links are intended to advance student learning through the achievement of fairness. The burden has shifted from the oppression of accountability to the celebration of agentic power in each student.

3.0 Keep Your Eyes Wide

(The Chance Won't Come Again)

Just as learning is a situated activity, so, too, is this theory. In this desire for contextualization, the theory of ethics for writing assessment is distinct from most theories. Rawls, to whom the proposed theory is greatly indebted, did not insist on localism, as we do. While the limits of the possible are not given by the actual, as Rawls noted (2001, §1.4, p. 5), beginning with the actual is nevertheless a good place to dig in. There is a great use of the phrase "for example" in this special issue.

Norbert Elliot: Reflection on the Need for Quantitative Techniques in Graduate Education

For the theory of ethics to advance, I believe that it is essential that graduate education include a foundation in quantitative techniques. The range of descriptive and inferential statistical analyses needed to insure educational equity can easily be introduced within a seminar sequence devoted to research methods. Based on my own experiences in offering such a course to both undergraduate and graduate students for twenty-five years—and with collaborations with colleagues in the field during that period—I have found that in-depth exposure to the epistemological frameworks accompanying various empirical techniques is important in the pursuit of equity and fairness. Once this knowledge base is created, multidisciplinary teams can be developed for advanced research.

While Elliot identifies limits to the theory, there are also challenges that need identification. Each is related to the lack of cohesion that runs from our professional identities in English Language and Literature/Letters, through our identities as members of the field of Rhetoric and Composition/Writing Studies, and then to our research specializations as writing assessment scholars. Among the challenges, these are easily recognizable: lack of empirical education in graduate school; potential for slavish importation of measurement concepts without interrogating those concepts due to poor empirical education; absence of research conducted within related referential frames, absence of inclusiveness in research design, score disaggregation, and reporting; and lack of new voices resulting from professional disenfranchisement.

To address these limitations, the ethics of assessment, framed within a philosophical landscape, is promising. This ecology of virtue is germane to every teacher and program administrator. Ethics invites participation, especially the ways we are trying to frame it in relation to civil rights, citizenship, and colonialism. Because it seeks to establish the conditions for *eudaimonia*—defined by Hursthouse (1999) as "flourishing," or "the sort of happiness worth having" (p. 10)—ethics invites inclusiveness of disposition and reliability of action. And, because it seeks to establish principled action, ethics strives to do something beyond establishing "gaping need"—a barrier Horner (2015) has identified as one that shackles our practices to commodification (p. 473). As we pursue ethical dispositions, we come to identify the elements of *phronesis*, or practical wisdom. As we begin to look for the kind of distinctive approaches to particular problems identified in this special issue—such as the technical method of regression analysis identified by Cleary (1968) and broad historical analysis of imperialism by Tlostanova and Mignolo (2012)—we realize that ethics is more than tendency. At play is what Hursthouse (1999) described as a "rationality of morality" (p. 20). As ethics become more doable, diffusion is less likely because it is actionable. There is solace here. As postmodernism has thankfully ensured, absent are notions of censure and threat; present are occasions for agentic action.

Bob Broad: Reflection on Imperatives

"…gone are notions of should and must…"

In some contexts, at least, I'm still a fan of "should-and-must." That's why this excerpt from Rawls figures prominently in my discussion of the ethical problems of the standardized testing industry: "…laws and institutions no matter how efficient and well-arranged *must* [emphasis added] be reformed or abolished if they are unjust."

The theory is just that—a project advanced for further examination. Required for the theory to have force is a research agenda for ethics. Even articulation of that agenda is complex, as these five basic questions demonstrate:

- How can theoretical and historical research support and grow the theory of ethics we have advanced?
- What kinds of empirical methods, both traditional and experimental, will best accompany fairness as the first principle of writing assessment?
- What new voices are necessary for advancement of the proposed theory?
- How will those new voices, in concert with current voices, together foster inclusiveness of multiple points of view while maintaining principled action?

- What knowledges, languages, and learning practices must be recognized to ensure mutually sustaining relationships between reliability and validity within the theory of ethics under the integrative framework of fairness?

In a recent online dialogue with Asao Inoue regarding Dynamic Criteria Mapping, Broad argued that the *desire*, based in a principle of justice, to seek out and understand different ways of valuing language (and anything else) need not be completely constrained by identity politics. Productive border crossings and dwellings can take place through acts of principled intent despite a lack of difference. Indeed, depending on how we conceptualize and operationalize difference, we are all shaped by what Solomon (2012) has termed our "horizontal identities" (p. 33). Under such frameworks of inclusion, it is our hope that others will turn in future studies to questions left unanswered here.

Ellen Cushman: Reflection on Collaborative Assessment Design as Border Dwelling

Though much work toward more socially just theories and methods for assessment has been proposed and undertaken, more work remains in theorizing what assessment looks like when language biases are surfaced and addressed. Much writing assessment focuses on language use, reading, and writing that happens in school or workplace contexts that are created for students working in English only. This is particularly troubling given the rapid decline in diversity and speakers of Native and heritage languages. China, for example, established a language preservation policy in the face of diverse language erosion that seeks to ensue the reading, speaking, and writing practices of regional tribes, peoples, and states. A pluriversal language assessment would seek to honor and test the types of languages spoken outside the schools of students' home communities.

This praxis of assessment might adapt participatory action teacher research outlined here and elsewhere. It could begin with students, teachers, and test designers, working together to create language asset maps in classrooms and communities. Student research papers could be assigned that help to surface the types of language happening in everyday contexts. Follow up assignments could then focus on the border crossing moments where languages come into context side-by-side, especially those that require code meshing and brokering, as these events in particular are believed to lead to highest orders of metalinguistic awareness. Teachers and test designers could then begin to design constructs similar to these language practices that could focus on translation, transferring, and transforming of linguistic assets. Students, teachers, and designers could create rubrics and other assessment measures that best represent the generic conventions found in these community and border crossing moments.

We hope this special issue adds to the body of knowledge created by the writing studies community with respect to the opportunities that can be created when assessment is seen in terms of the creation of opportunity structure. This hope is accompanied by a reminder of our strength as we annually encounter approximately 48.9 million students in public elementary and secondary schools 20.6 million students in postsecondary institutions (Snyder & Dillow, 2015). Our influence is remarkable as we touch the lives of many, one student at a time.

David Slomp: Reflection on Complexity

The work on this special issue has been an enormously challenging undertaking. Through the many stages of this work (conference presentations, internal reviews of articles, external reviews, editor comments and feedback, webinar presentations, group discussions, and our individual efforts to execute our personal research agendas) we have been reminded of the enormous complexity of the project we are proposing. Complexity should never be seen as reason for inaction.

This past summer I walked a cohort of graduate students—all experienced teachers—through a bioecological (Bronfrenbrenner & Morris, 2006) exploration of why change in education was so difficult to achieve. As we explored this question, my students reached a point of despair. "It is too complex! Real, meaningful, substantive change isn't possible," they said. But they could not wallow in their despair—their final assignment was both to articulate a theory of change and to develop a plan to enact that theory—if they wanted to pass the course (in my own way, I agree with Broad's appreciation of *must*).

What my students learned is that coming to terms with complexity is necessarily a step in the process of discovery, imagination, and experimentation. I hope that like the authors of this special issue, our readers too take up the complexities we offer as an invitation to reimagine our field.

References

Broad, B. (2003). *What we really value: Beyond rubrics in teaching and assessing writing*. Logan, UT: Utah State UP.

Broad, B. (2012). Strategies and passions in qualitative research. In L. Nickoson & M. Sheridan-Rabideau (Eds.), *Writing studies research in practice: Methods and methodologies* (pp. 197–209). Carbondale, IL: Southern Illinois UP.

Broad, B., Adler-Kassner, L., Alford, B., Detweiler, J., Estrem, H., Harrington, S., … Weeden, S. (2009). *Organic writing assessment: Dynamic Criteria Mapping in action*. Logan, UT: Utah State UP.

Bronfenbrenner, U., & Morris, P. A. (2006). The bioecological model of human development. In R. M. Lerner & W. Damon (Eds.), *Handbook of child psychology, Vol. 1: Theoretical models of human development* (6th ed., pp. 793–828). New York, NY: Wiley.

Charmaz, K. (2014). *Constructing grounded theory* (2nd ed.). Thousand Oaks, CA: SAGE.

Cleary, T. A. (1968). Test bias: Prediction of grades of Negro and White Students in integrated colleges. *Journal of Educational Measurement, 5*, 115–124.

Cushman, E. (Chair). (2015, March). A theory of ethics for writing assessment: Risk and reward for civil rights, program assessment, and large-scale testing. Panel presented at the Sixty-Sixth Annual Convention of the Conference on College Composition and Communication, Tampa, FL.

Dylan, B. (1964). The times they are a-changin'. New York, NY: Columbia.

Elbow, P. (2012). Good enough evaluation: When is it feasible and when is evaluation not worth having? In N. Elliot & L. Perelman (Eds.), *Writing assessment in the 21st century: Essays in honor of Edward M. White* (pp. 303–325). New York, NY: Hampton Press.

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

Farley, T. (2009). *Making the grades: My misadventures in the standardized testing industry*. Sausalito, Calif. : LaVergne, TN: Berrett-Koehler Publishers.

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Chicago: Aldine.

Guba, E. G., & Lincoln, Y. S. (1989). *Fourth generation evaluation*. Newbury Park, CA: Sage.

Hamilton, M. (2012). *Literacy and the politics of representation*. London, UK: Routledge.

Hartel, E. M., Moss, P., Pullin, D. C., & Gee, J. P. (2008). Introduction. In P. A. Moss, D. Pullin, J. P. Gee, E. H. Haertel & L. J. Young, (Eds.), *Assessment, equity, and opportunity to learn* (pp. 1–16). Cambridge, UK: Cambridge University Press.

Haswell, R. H. (2012). Quantitative methods in composition studies: An introduction to their functionality. In L. Nickoson & M. P. Rabideau (Eds.), *Writing studies research in practice: Methods and methodologies* (pp. 185–196). Carbondale, IL: Southern Illinois UP.

Horner, B. (2015). Rewriting composition: Moving beyond a discourse of need. *College English, 77*, 450–479.

Huot, B. A., O'Neill, P., & Moore, C. (2010). A usable past for writing assessment. *College English, 72*(5), 495–517.

Hursthouse, R. (1999). *On virtue ethics*. Oxford, UK: Oxford University Press.

Hussar, W. J., & Bailey, T. M. (2013). *Projections of education statistics to 2022* (NCES 2014-051). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Inoue, A. B. (2004). Community-based assessment pedagogy. *Assessing Writing, 9*, 208–238.

Lazarin, M. (2014). *Testing overload in America's schools*. Washington, DC: Center for American Progress. Retrieved from https://cdn.americanprogress.org/wp-content/uploads/2014/10/LazarinOvertestingReport.pdf

Maddox, B. (2014). Globalising assessment: An ethnography of literacy assessment, camels and fast food in the Mongolian Gobi. *Comparative Education, 50*, 474–489.

Merton, R. K. (1938). Social structure and anomie. *American Sociological Review, 3*, 672–682.

Merton, R. K. (1996). *Opportunity structure: The emergence, diffusion and differentiation of a sociological concept, 1930s–1950*. In F. Adler & W. S. Laufer (Eds.), *The legacy of anomie theory: Advances in criminological theory* (pp. 3–78). New Brunswick, NJ: Transaction Publishers.

Moss, P. A. (1996). Enlarging the dialogue in educational measurement: Voices from interpretive research traditions. *Educational Researcher, 25*, 20–28,

Moss, P. A., Pullin, D. C., Gee, J. P., Haertel, E. H., & Young, L. J. (Eds.). (2008). *Assessment, equity, and opportunity to learn*. Cambridge, UK: Cambridge University Press.

Rachels, J. *The elements of moral philosophy*. Philadelphia, PA: Temple University Press, 1986.

Rawls, J. (2001). *Justice as fairness: A restatement*. R. Kelly (Ed.). Cambridge, MA: Harvard University Press.

Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher, 29*(7), 4-14.

Snyder, T.D., & Dillow, S.A. (2015). *Digest of education statistics 2013* (NCES 2015-011). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://nces.ed.gov/pubs2015/2015011.pdf

Solomon, A. (2012). *Far from the tree: Parents, children, and the search for identity*. New York, NY: Scribner.

Strauss, A., & Corbin, J. (1994). Grounded theory methodology: An overview. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 273–85). Thousand Oaks, CA: Sage.

Teixeira, R., Frey, W. H., & Griffin, R. (2015). *States of change: The demographic evolution of the American electorate, 1974-2060*. Washington, DC: Center for American Progress, American Enterprise Institute, & Brookings Institution.

Tlostanova, M. V., & Mignolo, W. D. (2012). *Learning to unlearn: Decolonial reflections from Eurasia and the Americas*. Columbus, OH: Ohio State University Press.

United Nations, Department of Economic and Social Affairs, Population Division (2015). *World population prospects: The 2015 revision, key findings and advance tables* (Working Paper No. ESA/P/WP.241). New York, NY: United Nations. Retrieved from http://esa.un.org/unpd/wpp/Publications/Files/Key_Findings_WPP_2015.pdf

White, E. M. (1985). *Teaching and assessing writing: Recent advances in understanding, evaluating, and improving student performance*. San Francisco, CA: Jossey-Bass.

Wiggins, G. (1998). *Educative assessment: Designing assessments to inform and improve student performance*. San Francisco, CA: Jossey-Bass.

Yancey, K. B. (1999). Looking back as we look forward: Historicizing writing assessment. *College Composition and Communication, 50*, 483–503.