

CHAPTER 7.

THE VIOLENCE OF ASSESSMENT: WRITING ASSESSMENT, SOCIAL (IN)JUSTICE, AND THE ROLE OF VALIDATION

Josh Lederman and Nicole Warwick

Research Problem: The negative consequences that accompany writing assessment are investigated—whether large-scale or classroom-based—and attention is given to better understanding of ways to conceptualize and mitigate these consequences.

Research Question: Can writing assessment be understood as a form of violence, and if so, what can be done to avert negative consequences?

Literature Review: We review writing assessment scholarship and peace research (specifically, Johan Galtung [1969, 1990]) in order to make the case that many of the negative consequences that accompany assessment and can be rightly considered violent; then we review validity theory literature to examine ways to discover, identify, and mitigate these consequences.

Methodology: We develop a theoretical framework by constructing a definition of violence. We then examine writing assessment through this theoretical lens. Finally, we argue that validity theory can and should work for the goal of limiting the violence of assessment.

Conclusions: We argue that the argument-based approach to validation proposed by Michael T. Kane (1992, 2001, 2006, 2013, 2015) plants the seeds for a theory of validity and validation that could be more sensitive to the violence of assessment than is usually the case. We make specific suggestions for how to adapt Kane's approach for these purposes. Specifically, we argue that shifting the role of mitigating negative systemic consequences (e.g., structural violence) into the very interpretation/use argument (IUA) of the

assessment will necessitate that such potential violence be investigated as part of the validation process, and not relegated to a secondary role.

Qualifications: Our focus is on locally developed and used writing assessments, not on published tests. Our aim, however, is to provide a theoretical framework that might guide inter-institutional assessment, and/or serve as a guiding principle of local assessments at any variety of sites. In addition, we note that while we focus on Kane’s argument-based approach to validation, and while Kane’s work is perhaps the most widely accepted model/theory of validity and validation, not all educational measurement specialists support this model (e.g., Borsboom et al., 2004).

Directions for Further Study: As noted above, we suggest actionable ways to foreground structural violence in assessment arguments as a necessary part of validation. Future studies might report on validation research that takes the potential violence of assessment seriously enough to frame the disruption of structural violence into the IUA of the assessment, and therefore as a primary area of investigation for the assessment’s validation inquiry.

It is no secret that writing assessment, at all levels, can *feel* violent to those assessed—whether by discouraging their progress, making them feel incapable or unintelligent, reinforcing a history of voices telling them that they “can’t write,” placing them in Basic Writing when they don’t feel that is where they belong, or denying them exit from a writing course or program. On one hand, classroom grading practices, which are intended to help students understand where they stand in a course or on a project, can reify hierarchical power relationships between the teacher/authority and the students/subjects (e.g., Shor, 1992), obscuring students’ views of their writing abilities or paper qualities (e.g., Inoue, 2012, 2014), or encouraging students to engage in inauthentic writing situations, as noted by Karen S. Nulton and Irvin Peckahm (Chapter 9, this collection; see also, Pulfrey, Buchs, & Butera, 2011) when they point out, “Only in school do we write to show others we know how to write—for a grade.” Assessment scholars like Inoue (2015) validate such anecdotes, noting that grades can be indeed “destructive to student learning in writing classrooms” (p. 178; see also Kohn, 2011). On the other hand, large-scale assessment—such as placement and exit assessment—comes with its own set of problems: students are often rendered powerless, receiving life-level decisions out of placement and exit from the nameless, faceless will of the institution (Lucas, 1988; O’Neill, 2011), and

as Marilyn S. Sternglass (1997) pointed out, large-scale assessment of timed-essays like the CUNY Writing Assessment Test (used for both placement and as a “rising junior” exam), pay “scant attention to the quality of the ideas being expressed and no concern with the writer’s purpose in responding to the test question” (p. 143).

Assessment can also feel violent to teachers and writing program administrators, often making them feel that they have to conform to imposed criteria that may not fit their students or their pedagogies, or forcing them to quantify unquantifiable matters just for the sake of satisfying institutional pressures. In classrooms, many of us dread the day that we hand back those first papers and watch our classroom full of critical co-investigators (Freire, 1970) begin to look more like unhappy employees who will do whatever the boss says but will not like it, the pay(off), or the boss. Program assessment can feel equally violent. For example, many commonly used writing assessment methodologies—such as score-based placement and exit assessments—are based upon such outdated assumptions, which often put teachers and placement/exit evaluators in a position where writing assessment actually undermines their philosophies of teaching, learning, and literacy. Even recent innovations, such as outcomes assessment, can be fraught; Chris W. Gallagher (2016) argues that outcomes assessment can take away teachers’ abilities to more genuinely engage in the types of pedagogical praxes that composition theory and practice strives to achieve, later saying that “top-down outcomes assessment regimes can lead to latter-day Taylorization” (p. 257). As history reveals, disenfranchisement is related to violence.

So the suggestion that writing assessment—as an enterprise of sorting and ranking students, particularly via comparison to pre-determined outcomes—can be harmful on several fronts is not likely to shock many. In this chapter, we aim to articulate when the harm caused by writing assessment becomes violence, and further, identify contexts in which that violence impedes social justice. In making this argument, we build a tradition of research in writing assessment on the effects of construct underrepresentation in assessing the writing of diverse students. From White and Thomas’s (1981) study of the English Placement Test to more recent work by Poe and colleagues (2014) on disparate impact in placement testing and Elliot and colleagues’ (2016) study of eportfolios, writing assessment researchers have traced the various ways that writing assessment may disadvantage various groups. With the exception of Inoue (2015), however, this work has not used structural violence as an explanatory framework.

Our main focus in this chapter, then, is on structural violence (Galtung, 1969), which is a less interpersonal or direct kind of violence and one that is laced into social structures and inflicted upon some groups but not others—e.g., when the poor suffer health issues from having less access to quality foods; when

racial or ethnic minorities have differential access to police protection; when, as Kelly J. Sassi (Chapter 10, this collection) points out, “the rate of death by suicide by Native female adolescents is nearly four times that of White females (U.S. Dept. of Health and Human Services, 2011)” (see UNESCO [2015] for other specific details/examples). Although composition scholars may be familiar with the concept of structural violence, the term is often used loosely, and we seek here to trace the history of the term in order to identify a more precise application of it in regards to writing assessment. We thus follow peace researcher Johan Galtung, who coined the phrase structural violence in 1969, and use his theory of violence as a framework for discussing and making visible the types of social (in)justice issues that arise when we assess student writing. In focusing on structural violence, our work follows in the tradition of scholarship in health (e.g., Farmer et al., 2006), sociology (e.g., Bourdieu & Wacquant, 1992), anthropology (e.g., Gupta, 2012), law (e.g., Peresie, 2009; West-Faulcon, 2009) and peace and conflict studies (e.g., Galtung, 1969, 1990), among other fields. We further seek to explore the mechanisms through which such potential violence seeps into our assessment efforts. In particular, we explore the roles that representation and normativity/normalization play in both (a) the types of worldviews that beget structural violence (e.g., informing a colonialist mindset) and (b) in nearly all assessment practices. Finally, while we offer no solution per se, we end with a suggestion that an assessment validation model based on Michael T. Kane’s argument-based approach (1992, 2001, 2006, 2013, 2015), with certain key revisions, could work to make these matters more visible for validity claims.

DEFINING AND THEORIZING VIOLENCE

Johan Galtung, who is often recognized as the founder of peace and conflict studies, coined the term structural violence in a 1969 article (Farmer et al., 2006; Galtung, 1990; Gupta, 2012). In that article, Galtung (1969) uses *structural violence* and *social injustice* interchangeably: “In order not to overwork the word violence we shall sometimes refer to the condition of structural violence as social injustice” (p. 171, italics original). Similar to Galtung’s method, we do not attempt “to arrive at anything like the definition, or the typology—for there are obviously many types of violence” (p. 168, italics original), agreeing instead that what’s “[m]ore important is to indicate theoretically significant dimensions of violence that can lead to thinking, research and, potentially, action” (p. 168). Our goal here is to develop a deeper understanding of what violence is, and how it relates to the larger project of writing assessment, so that we can work toward more nonviolent/socially just assessment practices and methodologies. In what

follows, we suggest three components of violence that will be central to our discussion below: (a) potential-actualization distance, (b) a zero-sum cost-benefit relationship, and (c) the avoidability of the harm inflicted. The conceptual framework informing our search for a deeper understanding of violence is shown in Table 7.1.

Table 7.1. Violence defined and differentiated

Violence	
Potential-actualization distances <i>exist</i>	
Potential-actualization distances are <i>avoidable</i>	
Personal Violence (Direct)	Structural Violence (Indirect)
Clear assailant-victim connection	No clear assailant-victim connection
Intentionality on the part of the assailant	No intentionality on the part of the assailant
Clearly visible	Largely invisible

Galtung (1969) defines violence as

the cause of the difference between the potential and the actual, between what could have been and what is. Violence is that which increases the distance between the potential and the actual, and that which impedes the decrease of this distance. (p. 168)

Rather than defining what a person's or people's potential may be, Galtung asserts that whatever that potentiality is, violence is the cause and impact of being held back from reaching it. This framework is particularly powerful because potentiality can relate to larger social goals such as "freedom, education, autonomy, dignity, and the ability to participate in society" (Deaton, 2013, p. 9), and it can relate to micro-level matters such as academic achievement, job opportunities, and leadership roles. For example, when a person is less likely to get a job interview because their name sounds more Latinx or African American than white (e.g., Bertrand & Mullainathan, 2003; Matthews, 2014), Galtung argues that violence is what causes and what maintains that distance between the potential to be employed and lesser likelihood of actualizing that potential because of these social factors. Similarly, when students from certain zip codes have a greater potential-actualization distance because their school has less qualified teachers (Goldhaber, Lavery, & Theobald, 2015), more physical violence, fewer resources, less likelihood of proper diagnosis of learning disabilities or special education needs (Morgan, et al., 2015) than students from a different zip code experience, we can describe this as violence; all of these students may have similar levels of

potential, however defined, but the chances of actualizing that potential are far lower for the first group. These disparate situations are not just matters of happenstance or bad luck; they are created and maintained by violence. J. Elspeth Stuckey (1991), famous for her work *The Violence of Literacy*, makes a similar point when she argues, “To elucidate the violence of literacy is to understand the distance it forces between people and the possibilities for their lives” (p. 94).

Violence that occurs on the physical, individual level—as opposed to the structural level—seems predicated upon both direct action and intentionality. The existence of an agent directly and intentionally inflicting harm upon another seems to be, in large part, what distinguishes violence from mere or accidental injury. Galtung (1969) uses the word pairs *personal-structural* and *direct-indirect* to delineate this type of direct, physical violence from forms of violence that inflict harm through the smooth functioning of unjust social structures (see also Žižek, 2008). Galtung states, “Violence with a clear subject-object relation is manifest because it is visible as action. . . . Violence without this relation is structural, built into structure” (p. 171). Thus, “in a society where life expectancy is twice as high in the upper as in the lower classes, violence is exercised even if there are no concrete actors one can point to directly attacking others, as when one person kills another” (p. 171). Given this notion that structural violence has no clear direct assailant-victim relation, using the metric of intentionality to identify acts of violence can be problematic because without that subject-object relation between the victim and the assailant, it may seem as if there is no violence when there is no concrete agent intending to benefit from the suffering of another. But Galtung warns that “ethical systems directed against *intended* violence will easily fail to capture structural violence in their nets” (p. 171, our emphasis). In other words, when one looks only for intentional harm as indicating violence, the existence of structural violence becomes invisible. Focusing on intent may actually misdirect our attention from the indirect, structural violence that exists in these systems.¹

Rather than intentionality, then, we focus on the types of violence that entail a system of *beneficiaries* and *casualties* operating in a zero-sum relationship—meaning that those who benefit do so *because of*, and *to the extent that*, those who suffer, suffer. Farmer and colleagues (2006), using Galtung’s framework to explore the violence of differential disease spread and healthcare opportuni-

1 Iris Marion Young’s (2011/1990) discussion of oppression asserts a similar point: she argues that oppression refers not only to “the tyranny of a ruling group over another” but that in the “*structural* sense[,] oppression refers to the vast and deep injustices some groups suffer as a consequence of the often unconscious assumptions and reactions of well-meaning people in ordinary interactions, media and cultural stereotypes,” to the “everyday practices of a well-intentioned liberal society” (p. 42, our emphasis)

ties, point out some of the major instances of structural violence in the modern world, e.g., poverty, racism, and gender inequality. In terms of a zero-sum system of beneficiaries, these matters can be seen as violence in part because (a) the non-poor benefit from the unequal distribution of resources that keep the poor in danger of all kinds of harm; (b) men benefit from the opportunities disproportionately unavailable to women; (c) white people benefit from, among other things, a legal system that is less concerned with their transgressions than with those of racial minorities. These beneficiaries may not wish for these benefits, but as Charles W. Mills (1997) notes in *The Racial Contract*, “All whites are *beneficiaries* of the Contract, though some whites are not *signatories* to it” (p. 11, italics original). Many may even wish to give up their received benefits if it were possible; however, when the suffering of some exists in this zero-sum relationship with benefits experienced by others, regardless of intentionality, we have reason to call it violence.

The final component of our definition of violence stems from Galtung’s (1969) point that violence can be seen as avoidable harm. He uses the example of an earthquake, stating “[T]he case of people dying from earthquakes today [1969] would not warrant analysis in terms of violence, but the day after tomorrow, when earthquakes may become avoidable, such deaths may be seen as the result of violence” (pp. 168-169). As we explore the violence of assessment and of assessment methodologies, we will also focus on the question of avoidability, using the question of whether assessment has to be done in certain ways—or whether certain practices and principles are avoidable—as part of our inquiry.

With this framework in hand, then, we next explore the ways in which two methodological pillars of writing (really, all educational) assessment work to usher in the potential for violence in even our more carefully thought-out and context-sensitive assessment practices: representation and normalization/normativity. Matters of social (in)justice will require the kinds of critical social inquiry traditions that specifically deal with less visible matters of power and systemic oppression—such as feminist, queer, postcolonial, anti-racist traditions which actively seek to problematize historical power-relations (including dominant or assumed/unexamined positionalities), some of which we cite in this section.

VIOLENCE, REPRESENTATION, AND NORMALIZATION

One of the key connections we see between structural violence and writing assessment occurs in the process of representation, or the act of speaking on behalf of others or creating a description or portrayal of others. Representation will always be problematic, as it takes over (at least part of) the decision-making process for the represented; speaks *for* others; chooses what to include (and im-

portantly, to exclude) from any description or portrayal; and delimits agency, in that the represented are at the mercy of the representor in terms of how they will be portrayed to the world and back to themselves. Postcolonial theorist Edward Said (1993) stated:

We have become very aware in recent years of the constraints upon the cultural representation of women, and the pressures that go into the created representations of inferior classes and races. In all these areas—gender, class, and race—criticism has correctly focused upon the institutional forces of modern Western societies that shape and set limits on the representation of what are considered essentially subordinate beings. (p. 80)

Feminist ethnographers Avishai, Gerber, and Randles (2012) similarly describe “the problematic power dynamics of speaking for others” (p. 402).

In terms of violence, it would seem that those with the power to represent others have the power to control the potential-actualization distance of those others. As Said (1993) noted in the above quote, modern Western societies possess the institutional power to determine the boundaries (the shape and limits) of how others (by gender, class, and race) will be represented, a point that echoes Galtung’s (1969) first principle of violence, namely, the increasing of another’s potential-actualization distance. But Said also points out that this “capacity to represent, portray, characterize, and depict is not easily available to just any member of just any society [R]epresentation itself has been characterized as keeping the subordinate subordinate, the inferior inferior” (p. 80). Thus we see that zero-sum system of beneficiaries at play here too. Those with the power to represent, with the power to decide how others are to be portrayed, are invariably the ones who benefit from both the representation of the other, as well as of the self (Coronil, 1996).

As for the third component of our framework, avoidability, Said (2003/1977) says “it needs to be made clear about cultural discourse and exchange within a culture that what is commonly circulated by it is not ‘truth’ but representations” (p. 21). Indeed, Said’s “analysis of the Orientalist text therefore places emphasis on the evidence, which is by no means invisible, for such representations as *representations*, not as ‘natural’ depictions of the Orient” (p. 21, italics original). In other words, the given representation of a group or group member cannot be unavoidable since it is predicated upon a decision made, not upon what is the ‘truth’ or what is ‘natural’—which might then be seen as avoidable.

The connections to assessment here should be clear. Whether we represent a specific paper with a letter grade, a student with a placement decision/identity (e.g., basic writer), or we represent their semester’s or year’s coursework as a pass

or fail via exit assessment, our assessment scores and decisions nearly always reflect an image back to the student/writer of themselves, one which will never capture the totality of their learning/literacies, and one which is always rooted in some type of cultural and/or institutional values that were decided by those with the power to decide. As James Baldwin succinctly put it in a 1963 interview, the white power structure of the modern world was based upon the “possib[ility] for an Englishman to describe an African and make the African believe it . . . [,] for a white man in this country to tell a negro who he is and make the negro believe this” (Mossberg, 2012). Writing assessment—whether classroom-based or large-scale, formative or summative—would seem to entail a very similar power of description/representation. Key is focusing on the interpretation and use of scores.

Further, we know from the brief history of our field that the mechanisms through which we represent students are methodological choice, none of which offers a “natural” depiction of the “truth” of the student/writer/writing. At one time, for example, the notion of direct writing assessment—that is, actually reading student writing instead of assessing their writing via multiple-choice tests—was once viewed as unfeasible; White (1993, 2001) recounts the battle to convince relevant audiences that direct writing assessment could be made reliable enough to replace indirect, multiple-choice testing. Today, the choice to use multiple-choice testing, with its highly reliable but dangerously thin representations of student writing, is very clearly avoidable—from a methodological perspective, if not always from a financial perspective. Since that time, the direct assessment of timed-essays has been largely over taken by portfolios and/or eportfolio assessment as the soundest, richest methodology for assessing student writing/writers. And scholars such as White (2005) proposed specific methods that are far more cost effective than what once seemed plausible. The point is that as time passes and fields develop, the matter of avoidability changes, and that which once seemed unavoidable (and therefore, nonviolent) can indeed become a matter of violence—as Galtung (1969) pointed out in stating that in 1969 a deadly earthquake “would not warrant analysis in terms of violence, but the day after tomorrow, when earthquakes may become avoidable, such deaths may be seen as the result of violence” (pp. 168-169).

Inseparable from representation is another relic of colonialist thinking—perhaps *the* archetype of a structurally violent worldview—is the drive toward normalization and/or normativity, or the rewarding of proximity to a norm and punishment of distance from it. Normalization involves expectations for what student writing will look like by the end of the semester or program; the methodological assumption behind outcomes assessment is one that privileges proximity to a dominant norm, which renders the assessor unable to see anything but that distance/deficit when the text does not line up to the outcomes

as predicted (Gallagher, 2012). Normalization also involves norming readers/raters to those criteria that indicate good writing (Gallagher, 2014). According to Gallagher (2012, 2014), both aspects risk losing what is “surprising or excessive or eccentric” on the part of the reader, not just the student/writing: “[C]onsensus-driven processes [risk] marginalizing, or managing away, outlier views and dissent in favor of shared understandings and normed judgments” (p. 82). In other words, normalization as a fundamental methodological assumption ensures that we only see the already agreed-upon (limiting our responses to the anticipated, scripted, programmed) and hence, we blind ourselves to what is surprising or idiosyncratic, or simply of high quality but in unexpected or non-pre-determined ways.

When we conflate literacy learning with normativity, even if done in the belief that the same definitions of literacy create the fairest assessment for all students, we create or reinforce differential potential-actualization distances for the already (and historically) normative and the already (and historically) marginalized, maintaining a smoother road for the former and a rockier road for the latter. In that way, assessment-as-normativity-check will differentially impact students/writers from socio-historical situations that are already closer to or further from those dominant norms—a clear zero-sum game in that those born into circumstances closer to the expected norms are privileged to the extent that those further from the expected norms are held down. What is more, while some may suggest that the privileging of proximity to predetermined norms is unavoidable, arguing that there is no assurance of fairness outside of such a move, we note that more and more scholarship over the past quarter century—in both writing assessment (e.g., Gallagher, 2012, 2014; Inoue, 2015; O’Neill, 2011; Whithaus, 2005) and educational measurement literature (e.g., Miselvy, 2004; Moss, 1994; Parkes, 2007)—has worked to problematize this premise, suggesting that such methodologies have the type of avoidability that Galtung (1969) noted as a facet of structural violence.

In a similar vein, Nicole I. Caswell and William P. Banks (Chapter 11, this collection) assert that “heteronormativity continues to enact violence on our teachers, students, instructional contexts, and research practices,” and Sassi (Chapter 10, this collection) problematizes the whitestream discourse of the rubrics used to assess Lakota/Dakota Indian student writing at Sitting Bull College, reporting that “[s]teps toward a more socially just writing assessment were possible only after educators proved to themselves that their students were making gains on a whitestream measure.” Furthermore, J.W. Hammond (Chapter 1, this collection) argues, while we may frame such enterprises around notions of inclusivity, in reality, they “pivot around the axis of sameness, assessing difference and deviance against an imagined “native” white norm or ideal,” ultimately

serving as a “linguistic front for advancing white normativity and supremacy” and the “purging foreign difference.”

To be sure, there is a practical counterargument that such critical arguments face: We must have some centralized standard against which every student is evaluated, otherwise the assessment cannot ensure that all students are being evaluated fairly. Writing that displays vastly different norms, but receives the same scores, may seem patently unfair, inequitable. But on the other hand, due to the uneven nature of background social structures, some students are essentially born into situations with norms that already conform to these expectations, while other students—particularly those from non-dominant backgrounds—are born into situations where these norms are foreign or even actively resisted (e.g., Gee, 1989; Ogbu, 1999, 2004). As such, evaluating all students’ writing against the same criteria or norms may be the unfair approach in that it requires both a different path (longer and bumpier for some) from the home language and literacy practices to the tested practices and it ignores all kinds of social and community pressures to resist such conformity to school-based language and literacy practices that only student writers from non-dominant will be likely to experience (Ogbu, 1999, 2004). From a critical perspective, it seems, then, there is no way to establish a norm *a priori* that is politically neutral, as history shows that those with the power to say what is normal always tend to assert their norms, and then require conformity from all others (Mathew Gomes, Chapter 6, this collection; Keith L. Harms, Chapter 3, this collection).

THE VIOLENCE OF ASSESSMENT: A CASE STUDY

Linda Harklau’s (2000) longitudinal study, “From the ‘Good Kids’ to the ‘Worst’: Representations of English Language Learners Across Educational Settings,” demonstrates how representation and normativity leads to the increased potential-actualization distance for a group of long-term US immigrant students. Harklau’s article focuses on three participants, each a long-term immigrant and language minority, who all experienced vastly different institutional relations and projected learner identities after their transition from high school to a local community college. All students had been praised in high school for their work ethic and determination, seen as “an inspiration to everyone” (p. 46). In college, all three were placed in the ESOL program—by means of a “diagnostic test—a commercially published, standardized multiple-choice, grammar-oriented measure designed for nonnative speakers of English” (p. 57)—whereupon Harklau noticed that, for all three, their effort and achievement wilted and their resistance rose. These students had been in U.S. schools for 6, 7, and 10 years, respectively, and the college’s ESOL program consisted mainly of international

students and recent immigrants. Its curriculum was partly dedicated to teaching “American” activities like how to read a newspaper, “on the assumption that they were not accustomed to extensive reading in English” (p. 47). But in fact, two of the three participants were not enrolled in the ESOL program at all in their senior year of high school, and one was literate only in English. By the end of that first college semester, Harklau reports, “these students’ resistance to alienating representations of their identities had become complete rejection” (p. 61).

We suggest that Harklau’s (2000) piece can be seen as a case study of potential structural violence of assessment—and the implications of representation and normativity—as the shift from the good kids to the worst is kicked off by the placement procedure. Harklau’s study itself is not focused on writing assessment, and all we are told about this placement test is that it was “a commercially published, standardized multiple-choice, grammar-oriented measure designed for nonnative speakers of English.” But what Harklau describes about the lived experiences of her participants can be viewed as the consequential fallout from this assessment; the violence that we see enacted upon these students—particularly visible in the elevated potential-actualization distance vis-a-vis their learner-identity trajectories—stems in great part from the representational label ESOL student at this particular college, a label placed upon them via placement assessment.

Part of the problem must be seen in the larger ecology in which that placement test took place. Harklau (2000) notes:

In the context of the high school, these images [‘Ellis Island’ images of immigrants leaving their homes, enduring financial and emotional hardships, and through sheer perseverance succeeding in building a better life for themselves in America] informed a representation of ESOL students as hardworking highly motivated students who had triumphed over adversity. (p. 46)

But Harklau observed that the college context stemmed from more of a deficit model, one that saw ESOL students as deficient and in need of acculturation:

Walking into college ESOL classes, the students in this study found themselves viewed in ways that not only were discontinuous with the predominant representation of their identity as ESOL students in high school but also seemed to cast their experiences with U.S. schooling and society in an unfavorable light. . . . [T]he prevailing representation of ESOL student identity depicted students as in need of socialization into U.S. college norms and behavior as well as to life in U.S. society

more broadly. The curriculum in ESOL classes reflected this image of students as cultural novices. (p. 53)

As Harklau points out, this “prevalent institutional and programmatic representations of what it means to be an ESOL student had material effects on students’ motivation and classroom experiences” (p. 62), as various forms of resistance “began surfacing in all the case study students and escalated over the course of the semester” (p. 60). Some of the forms of resistance were subtle, like one participant “declaring to her reading teacher that she seldom read the assigned newspaper” (p. 60). But one student stopped handing in assignments by the end of the semester and began missing classes. Another student only did the minimum requirements for credit, which “made it clear to her teacher that she regarded language assignments as busy work” (p. 58). On one assignment about clause connectors, the student “used since for every sentence. Because the teacher did not specify which connectors to use, she technically fulfilled the assignment requirements. Nevertheless, she drew her teacher’s ire and a C on the assignment” (p. 58, italics in original). Harklau’s observations in November and December of that semester report on participant spending most of her grammar classes either kicking and fidgeting or putting her head on her desk, and the other two participants “conspicuously for[getting] to bring their books to class” (p. 60).

To be sure, it seems clear that Harklau’s participants were fully capable of doing better work in college, of being the “good” kids, as they had been in their earlier years. Their struggle in these college classes seems less a statement of their capabilities and more a signal of unequal social relations—i.e., the description/representation/label of each student as ESOL college student with the various negative connotations that label brought in this particular setting. Such labeling became another example of Baldwin’s observation that power structures exert the ability to “tell a [student] who he is and make the [student] believe this” (from Mossberg, 2012). We should ask: What does it do to a first-generation college student, a long-term immigrant (recall, 6, 7, and 10 years, respectively), one who has finished their high school career in the US no longer needing any enrollment in an ESOL program (for two of the three participants), when the first institutional statement from college tells you, “No, you are indeed an ESOL student; you need to be taught how to be a real American”? How is such a student to avoid the trap that Baldwin describes of believing the descriptions through which institutional power structure represent you?

In terms of normativity, while we do not know much about the specific assessment by which Harklau’s participants were placed in the ESOL program, it seems clear that a “commercially published, standardized multiple-choice, gram-

mar-oriented measure designed for nonnative speakers of English” (p. 57) will be driven by the methodology of assessing students vis-a-vis proximity to a pre-determined norm. Such exams—again, given the standardized multiple-choice, grammar-oriented nature Harklau describes—en masse, are not meant to observe and evaluate test-takers’ actual language abilities but rather to operationalize the (only) criteria through which those literacies will be acknowledged; one could argue that the premise of such assessment methodologies is to determine that which is to be considered normative, and then to measure students’ proximity/distance vis-a-vis that norm. But while we do not have enough information to speak more about the assessment instrument and decision-making process itself, we can speak of the impact of normalization/normativity at play in the larger ecosystem of the assessment-program-curriculum-instruction as well.

As Harklau observed one participant (Penny) in class, she saw that “while her teacher lectured the class on how to locate stories in sections of the newspaper, Penny could be seen flipping to her horoscope and local department store ads, a small but telling act of resistance as well as a more authentic act of newspaper reading than the class exercises” (p. 57). Ironically, in a program in which the “curriculum was partly dedicated to teaching ‘American’ activities like how to read a newspaper” (p. 47), Penny may have been displaying the most normal, authentic American teenager behavior. She already knew how to read a newspaper fluently, yet her paper reading behavior was clearly not the normal that her teachers or the program sought. Penny was expected to align her reading behaviors with institutional norms—to read in an idealized American way.

Harklau and her participants all noticed this issue as well. Harklau (2000) states, “[I]ronically enough, teachers implicitly rejected the very Americanness of [Harklau’s] students’ educational backgrounds in favor of the class and educational backgrounds of students educated abroad.” Further, she notes that “the case study students recognized and sometimes resented the favoritism shown toward newcomers in their classes” (p. 59). Clearly, there is a certain type of narrow normativity (Americanness) that is valued over others in the larger ecosystem of this ESOL program. Extended to assessment practices used to funnel students into the program, this singular view of American normativity is troubling.

Are these side-effects of the ESOL placement indications of violence? If we consider impact of a college education—or the lack of one—in the 21st century United States, especially for students from historically marginalized social groups, and the decreased likelihood of students like Harklau’s participants making it through college after such a rough welcoming, we can clearly see the seeds sown for increased potential-actualization distance, Galtung’s (1969) baseline definition of violence. We ought to further ask who benefits from this situation, for it is clearly not these three students. In addition, if there are factions of society benefitting

from the potential-actualization distance suffered by students like these—perhaps students whose college readiness is overestimated by similarly shallow measures of grammar (which often really means dialect), or students who come from college educated families and who have more guidance about the consequences of placement tests, and who even know how to speak to the right people to overrule placement decisions—if there are factions of society benefiting from the potential-actualization distance of other student, in a zero-sum relationship, we suggest that students like Harklau’s participants are suffering the violence of assessment. Finally, in terms of avoidability, if there are better ways to determine which students would be best served by which courses and programs—methods more sensitive than commercially developed grammar-oriented multiple-choice exams, methods that would approach the type of maximum construct representation, which Elliot (2016) asserts as the definition of fair assessment—then by the definition laid out in this chapter, the label of violence seems appropriate.

With that said, pinning the ensuing problems entirely on the placement exam is neither fair nor our intention. However, the exam is surely involved; the exam does not exist in a vacuum. Taking an ecological view, we must see that the test-based placement decisions and their consequences do not exist outside of the larger social and institutional systems in which student with certain scores will be subjected to certain types of treatment and expectations—because of those scores. While the exam may be claiming to place students, by measurement of (and only of) their English language abilities, into programs of study that will best serve their academic futures, we can clearly see that the test results foisted Harklaus’ participants into an environment in which (a) their ESOL status was forefronted as a larger part of their identity than it had been in high school, and in which (b) the teachers expectations of ESOL students was far lower than in their previous schooling (see Rubie-Davies, 2006, for a study and a review of the impact of low teacher expectations upon students). In this way, the assessment is surely caught up in a larger system that upholds this type of violence, particularly for students from certain vulnerable populations—language minority students, in this case. At this point, some may question if the assessment can legitimately be called violent, just for working within a larger system the propagates violence, while possibly not enacting violence itself. We hope to address this critically important question below.

MOVING PAST THE VIOLENCE OF ASSESSMENT: THE ROLE OF VALIDITY AND VALIDATION

If the violence of assessment is to be mitigated, it first must be made visible through more vigilant evaluation of specific assessment practices. The process

of evaluating an assessment—including the search for indications of structural violence—is the process of validation. For decades now, validity theorists have argued for the incorporation of social consequences—and in particular, a concern for unintended negative social consequences—into the very definition of validity and, therefore, as central to the validation of assessments. While there is considerable debate about both what role consequences should play in validity and validation (e.g., Linn, 1997; Mehrens, 1997; Popham, 1997; Reckase, 1998; Shepard, 1997) and which negative social consequences should count as evidence for or against validity (e.g., Cronbach, 1988; Kane, 2013; Messick, 1989)², the term negative social consequences (emphasizing the word *social*) clearly echoes the phrases *social injustice* and/or *structural violence*, and so we believe that these matters are unquestionably relevant to the validity or invalidity of writing assessment practices. To suggest otherwise is to argue that structurally violent assessment can still be valid assessment, and there should be no room for such a contradiction in a conceptualization of writing assessment validity. Our aim here is to explore the potential that current educational measurement validity theory holds for spotting and disrupting the violence of assessment, but we also aim to push that theory (particularly as it relates to writing assessment) in directions that will focus more explicitly on the search for and disruption of this violence. In particular, we explore the potential of Michael T. Kane’s argument-based validation model for incorporating social justice into the very validity of writing assessment practices, thereby rendering assessments that propagate social injustice/structural violence, by definition, invalid.

2 A great deal of validity theory regards the development of tests by testing experts, whose goals are to design the best possible test for a particular use, but who will have no first-hand experience of the consequences of the test’s use. Published tests are designed to be used across as many sites and contexts as possible, and so the actual decision made and consequences experienced at any given site of test use are far removed from the daily practice of such test developers. The validation concerns of such test developers do regard the consequences of test use—or at least, they are supposed to (e.g., American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014)—however, the way such consequences impact their validation research would tend to be much less direct. Whereas we might instantly rethink and revise an assessment procedure that was harming our students and their learning, a test publisher would need much more data about the nature of such consequences, the degree to which they resulted from proper or improper use of their product, and evidence that these consequences stemmed from problems with the test itself (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014, p. 21; Messick, 1989, p. 68). But to the extent that our writing assessments remain locally-controlled, site-based, and context sensitive (Huot, 2002), the separation of the validity of an assessment and the consequences of that assessment make little sense. As such, we take for granted that the validity of (locally-controlled, site-based) writing assessment is inseparable from the consequences (social, individual, negative, positive) of the decisions it helps us make.

ADAPTING KANE'S ARGUMENT-BASED APPROACH TO WRITING ASSESSMENT AS SOCIAL JUSTICE

For Kane (2013), validation consists of two key stages: first, “a clear statement of the proposed interpretations and uses,” called the interpretation/use argument (IUA) and second, “a critical evaluation of these interpretations and uses” (p. 64), called the validity argument. For writing assessment purposes, we interpret Kane’s model as requiring (a) an IUA that articulates what we hope to achieve through the assessment—what decisions we want to make and how we plan to draw inferences from the collected writing to make such decisions—and (b) a validity argument that evaluates the theoretical and empirical soundness of this plan. Kane’s approach/model has arguably become the primary model in the educational measurement research and literature over the past decade or so, replacing Samuel Messick’s long-standing but more abstract and philosophical theory of validity. Messick’s (1989) major treatise on validity, along with his other works throughout the 1980s and into the 1990s, emphasized (a) the need to incorporate social consequences as fundamental to the concept of validity, (b) the notion that validity is a unified concept (i.e., that there are not types of validity, such as content- or criterion-validity; there are only various lines of evidence for the overarching judgement of validity for an assessment procedure), (c) and that construct validity is the totality of validity; in other words, that the meaning of the assessment scores vis-a-vis the construct of interest, interpreted through multiple lines of evidence and multiple methodologies, is where validity is to be found. (The notion that a test itself can’t be considered valid, but rather that validity regards the inferences drawn from the test scores—these matters pre-date Messick’s major work [e.g., Cronbach, 1971]). For all the philosophical richness of Messick’s work, particularly Messick (1989), some theoretical contradictions seem to exist between (a) and (c), as well as (b) and (c) above. His work left those charged with validating specific testing instruments a bit uncertain as to how to proceed. This is where Kane’s more pragmatic model has been somewhat of an antidote to the problems of applying Messick’s work. In fact, Kane speaks at length about the processes of validation, but does not himself delineate a theoretical definition of validity itself. Again, Kane (2013) breaks down the process of validating an assessment into two steps: articulating the IUA, and then evaluating the coherence and “evaluating the coherence and completeness of the IUA and the plausibility of its inferences and assumptions” (p. 9). Our belief is that while Messick’s work lays the foundation for the centrality of consequences to the validity of an assessment, Kane’s approach holds the greater potential for our specific goal here of making visible, and rooting out, the potential and actual

violence of specific assessment practices.

At the same time, we call for some adaptation to Kane's approach, particularly centering around one aspect. We want to emphasize the empirical element of the validity argument (the second stage) more than Kane has tended to, and in a different way. Kane repeatedly emphasizes the need for the validity argument to check on the coherence and completeness of the IUA and on the plausibility of its inferences and assumptions (e.g., Kane, 2013, p. 1, p. 9, p. 14, p. 18, p. 64-65; Kane, 2015, p. 5, p. 12), but he provides little emphasis upon the types of social research that Messick and others (e.g., Moss, 1994) have argued are vital to evaluating the consequences of assessment-based decisions. That is, Kane's second phase of the argument-based validation approach may stay too focused on the internal logic and coherence of the IUA, with too little emphasis on exploring the worldly impact of the decisions (the use) of the assessment scores.

To proceed with a social justice agenda, however, the spotting and disrupting of structural violence must be a top priority, and these matters require empirical social inquiry. Rarely, if ever, will structural violence become visible through the assessment developers' *a priori*, rational considerations of the logical coherence of their IUA. As we note above, from a critical perspective, it seems there is no way to establish a norm *a priori* that is politically neutral. In fact, structural violence will hide from many empirical methodologies—particularly those that operate within historically dominant paradigms and positionalities. As such, validation research that is genuinely concerned with these matters of social (in)justice will require the kinds of critical social inquiry traditions that specifically deal with less visible matters of power and systemic oppression—such as feminist, queer, postcolonial, anti-racist traditions which actively seek to problematize historical power-relations (including dominant or assumed/unexamined positionalities).

We are not the first to recognize that traditional measurement-based research methodologies are largely ill-equipped to investigate unintended social consequences like structural violence/social injustice. Edward Haertel's (2013) article "Getting the Help We Need," for example, notes that "most measurement specialists still feel a stronger affinity to the models and methods of psychology than, say, sociology, anthropology, or economics [and] may be ill-equipped, working alone, to investigate fully the systemic effects of testing programs" (p. 87), and he recommends teaming up with colleagues from fields like sociology, anthropology, economics, law, and linguistics. Pamela A. Moss (1998) similarly discusses the need for "studying the consequences of the repeated and pervasive practices of testing," noting that "[w]hile many of us may not have the resources to undertake this kind of work ourselves, we can

at least (initially) seek to develop collaborations with those who do. We have colleagues in AERA who engage in and find funding for this sort of research regularly” (p. 11). Many of us in writing assessment are in a better position to develop and engage in this type of inquiry than those in the measurement community given that our field is so closely tied with non-dominant research traditions. The question is how to make this type of inquiry fundamental to validation research and to the validity of writing assessment practices, and to not leave it as an afterthought.

Perhaps the answer is to reposition the role of these unintended negative consequences, namely structural violence/social injustice—to shift them from plausible threats to otherwise valid assessment, as most validity theory suggests, to part of the intended, expected benefit of the assessment itself. That is, if part of the very IUA of an assessment—the articulation of what the assessment plans to achieve and how it hopes to achieve it—entails a description of how this assessment will work to ameliorate or disrupt existing social injustice, then social inquiry into the structural violence of assessment would be part of ensuring that the intended goals/benefits of the assessment are reached. If we can issue this challenge to ourselves as developers and users of locally-controlled, site-based writing assessments, we may find a way for a concept of validity that truly works against assessment-based social injustice.

Nearly 30 years ago, Lee J. Cronbach (1988) stated, “*Tests that impinge on the rights and life chances of individuals are inherently disputable*” (p. 6, italics original), but outside of certain notable exceptions (e.g., Callahan, 1999; Chudowsky & Behuniak, 1998; Lane & Stone, 2002; Lane, Park, & Stone, 1998; Poe et al., 2014), empirical inquiry into such impingement is far from the norm. Indeed, the measurement community has an equally long (perhaps longer) history of what some term a confirmationist bias (Haertel, 1999, 2013; Kane, 2001; Shepard, 1993) in that those who design and validate assessments tend to place emphasis on confirming how well they work, and less emphasis on searching for evidence of invalidity (and even less when such evidence requires inquiry into social/systemic matters). When we consider the potential violence of assessment and recognize that our definition of validity—and resulting approaches to validation—are critical for identifying and disrupting this violence, the notion of confirmationist biases becomes deeply troubling. When assessment developers only look for evidence that their instruments are working, but we do not push ourselves to actively seek out less visible ways in which problematic consequences are unfolding, we embrace that confirmationist bias. When we rest easy with evidence for the positive impact of an assessment program but do not consciously explore instances of structural violence that either result from, or simply operate smoothly within,

our assessment systems, we are complicit with that violence.³ As Paulo Freire (1985) reminded us, “Washing one’s hands of the conflict between the powerful and the powerless means to side with the powerful, not to be neutral” (p. 122). We need to resist theories of validity and validation that end their investigations with evidence of clean hands.

If writing assessment is to be an agent of social justice—not just something that tries to stay out of the way—then fighting against social injustice (i.e., structural violence) should be central to “the only genuine imperative in testing, namely, validity” (Messick, 1989, p. 20). When our theory of validity allows for assessments to operate within such systemic inequalities, as long as they stem from matters like differential access, opportunity, quality of education, health, or other instances of structural violence—when we find it okay for testing to merely reflect these existing social inequalities, as long as the tests themselves show no evidence of invalidity—we need to consider ourselves agents of that structural violence. Any assessment program could harbor potential violence, and so if we are not conceptualizing assessment and validity/validation practices that seek to disrupt this violence, it would seem we are complicit in it.

CONCLUSION

In discussing structural violence, or violence built into the very systems we have created, like healthcare, education, and the legal system, and because Galtung (1969) characterizes structural violence as having an absence of an actor or actors acting on a direct object, it can be easy to overlook that people participate in and maintain these systems—that people can and do enact structural violence. It may seem that people are only involved in personal or direct violence. Part of the issue is that often the processes of structural violence that they enact are often rendered invisible because they are part of established systems—because they are part of the very fabric of our society. They are normalized and rendered invisible (Žižek, 2008). However, it is important for our work in connecting structural

3 Such complicity can still be seen lingering in the newest Standards revision (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014), which owes its stance on testing consequences to Samuel J. Messick’s (1989) outlook on the testing-consequences dialectic. While Messick is often (rightly) considered the driving force behind the inclusion of social consequences as central to validity, the fact is that he left open a loophole in his theory which can be seen in passages like this: “[I]t is not that adverse social consequences render test use invalid but, rather, that the validation of test use should assure that adverse social consequences do not stem from any source of test invalidity, such as construct-irrelevant variance” (Messick, 1989, p. 68). In other words, for Messick, negative social consequences only threaten the validity of a testing program when they can be traced to problems in the testing instrument itself (e.g., construct-irrelevant variance).

violence and writing assessment that we see that people play a role in enacting structural violence.

As part of his work on examining structural violence, Galtung (1969) points this issue out. He asks: “Is there really a distinction between personal and structural violence at all?” (p. 177). To answer this question, he notes that people can

act violently not only on the basis of individual deliberations but (also) on the basis of expectations impinging on him as norms contained in roles contained in statuses through which he enacts his social self; *and*, if one sees a violent structure as something that is a mere abstraction unless upheld by the actions, expected from the social environment or not, of individuals. (p. 177 , emphasis in the original)

People often become indoctrinated to ways of knowing and doing that are built on supremacist values—racism, patriarchy, value of abled bodies, and so forth—values that systems are built on. We see this very issue play out in Nicole Gonzalez Van Cleve’s (2016) research of racism and injustice in the criminal courts of Cook County-Chicago. As Van Cleve writes, “It is as though attorneys inherit a culture of racism that has existed ‘a priori’ (before) their participation. The *a priori racism* that defines the courthouse culture and the legal habitus existed long before they arrived at the courthouse, and it will sustain itself long after they retire” (p. 4). This court system is an example of structural violence. A dominant (white) group is advantaged over a nondominant group (black and Latinx people). The system creates and maintains distance between what is potential and what is actual and these issues are avoidable. Van Cleve argues that essentially two criminal justice systems exist in Cook County and “they are two systems that are separate and *unequal*—one with a front door and one with a hidden back door where the majority of citizens—the poor and people of color—experience America’s failed promise of fair and equal justice” (p. xii). And *people* make this system run and they maintain its violence because they have inherited its culture and continue to pass it on, it seems without any critical examination—that is until Van Cleve took a critical look at it.

In this chapter we hope that we are contributing to creating tools to help ourselves as a writing community take a critical look at writing assessment. We began this chapter by incorporating a framework of structural violence as a lens through which to view the social impact of writing assessment, particularly on groups that have historically been denied access to higher education, in large part because of the ways that educational assessments past worked as sorting and ranking mechanisms intent more upon selecting the elite and identifying the unprepared, as opposed to finding ways to ensure access to education and op-

portunity to diverse groups and individuals. Our working definition of violence was based upon Galtung's (1969) article in which he defines violence as potential-actualization distance, and in which he offers a definition of structural violence, i.e., violence that is equally deadly but which possess a less direct/visible assailant-victim relationship, and he equates structural violence with social injustice. Our major three criteria for interpreting a situation as violent built upon the notion of potential-actualization distance; and, we added in the matter of benefit—specifically, a system of beneficiaries that exist in a zero-sum relationship to those suffering the cost of the situation; finally, we included Galtung's notion of avoidability, such that, if the harm inflicted upon persons and groups in a given situation is unavoidable, there may be no basis upon which to call the situation violent—but if the harm can be reasonably viewed as avoidable, then violence is indeed at play.

We next sought to theorize the Trojan horses, as it were, through which violence enters into even those assessment practices that are focused upon creating the best learning opportunities for all. We argued that, to some extent, all assessment seems rooted in the representation of the student/writer/writing, in the institutional description of the students/writers back to themselves—and not always in ways that help the student as individuals (in terms of reaching learning potential) or as social-historical group members (in terms of access and marginalization). Hand in hand with representation, we suggested that most if not all assessment methodology, on some level, center around the matter of normalization/normativity—they reward proximity to a norm and punish distance from it. And the delineation of these norms, again in line with colonialist thinking, are never rooted in notions of truth but always reflect certain values of what is expected by certain people in certain situations. We looked at how this violence appears in Harklau (2000), where we see how it affected three particular students; but we note that these particular students are examples of students whose escape from the historical cycles of marginalization is made less possible due to an assessment (the placement test, in this case) and the larger curricular/institutional ecology that drives, and is driven, by that assessment.

Finally, we suggested that, while eliminating violence from assessment is likely impossible, developing validation practices that make this violence more visible, and explicitly call for the reduction of that violence as fundamental to the validity of a writing assessment practice, is indeed possible. Thus, we issue the following challenge: Rather than asking assessment to “watch out for” collateral damage of unintended negative consequences, a stronger commitment to assessment as social justice would require assessment that actively seeks to disrupt the structural violence/social injustice already present in the larger systems in which students write and that writing is assessed. What would follow is a concept of

validity that sees ensuing or recurring structural violence not as the “cost” of the cost-benefit equation of validation research, not as negative unintended consequences, but instead incorporates the disrupting of structural violence as fundamental to the intended consequences, the benefit side, the very *raison d'être* of any nonviolent, socially just writing assessment. An actionable way forward, it would seem, would entail more explicit statements of the types of unintended negative consequences that would be unacceptable results of an assessment program or practice—statements included in the very design of the assessment. As mentioned above, making such statements part of the IUA itself would limit the possibilities for confirmationist biases to open the door for structural violence because validating such an IUA would entail explicit investigation into these unintended consequences; a validity inquiry that failed to perform such investigation would not have fulfilled its obligation to validate the assessment’s IUA, which means that the assessment would not have sufficient evidence of its validity, even if such inquiry provided strong evidence supporting the assessment’s positive intended consequences.

REFERENCES

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Avishai, O., Gerber, L., & Randles, J. (2012). The feminist ethnographer’s dilemma: Reconciling progressive research agendas with fieldwork realities. *Journal of Contemporary Ethnography*, 42(4), 394-426.
- Bertrand, M., & Mullainathan, S. (2003). *Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination*. Cambridge, MA: National Bureau of Economic Research.
- Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review*, 111(4), 1061-1071. <http://dx.doi.org/10.1037/0033-295x.111.4.1061>
- Bourdieu, P., & Wacquant, L. J. D. (1992). *An invitation to reflexive sociology*. Chicago, IL: University of Chicago Press.
- Callahan, S. (1999). All done with the best of intentions: One Kentucky high school after six years of state portfolio tests. *Assessing Writing*, 6(1), 5-40.
- Chudowsky, N., & Behuniak, P. (1998). Using focus groups to examine the consequential aspect of validity. *Educational Measurement: Issues and Practice*, 17(4), 28-38. <http://dx.doi.org/10.1111/j.1745-3992.1998.tb00633.x>
- Coronil, F. (1996). Beyond Occidentalism: Toward nonimperial geohistorical categories. *Cultural Anthropology*, 11(1), 51-87.
- Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed.), *Educational Measure-*

- ment (2nd ed., pp. 443-507). Washington, DC: American Council on Education.
- Cronbach, L. J. (1988). Five perspectives on validity. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 3-17). Hillsdale, NJ: Erlbaum.
- Deaton, A. (2013). *The great escape: Health, wealth, and the origins of inequality*. Princeton, NJ: Princeton University Press.
- Elliot, N. (2016). A theory of ethics for writing assessment. *Journal of Writing Assessment*, 9(1). Retrieved from <http://journalofwritingassessment.org/article.php?article=98>
- Elliot, N., Rudniy, A., Deess, P., Klobucar, A., Collins, R., & Sava, S. (2016). ePortfolios: Foundational measurement issues. *Journal of Writing Assessment*, 9(2). Retrieved from <http://journalofwritingassessment.org/article.php?article=110>
- Farmer, P. E., Nizeye, B., Stulac, S., & Keshavjee, S. (2006). Structural violence and clinical medicine. *PLoS Medicine*, 3(10), 1686-1691.
- Freire, P. (1970). *Pedagogy of the oppressed* (M. B. Ramos, Trans.). New York, NY: Continuum.
- Freire, P. (1985). *The politics of education: Culture, power, and liberation*. Westport, CT: Bergin & Garvey.
- Gallagher, C. W. (2012). The trouble with outcomes: Pragmatic inquiry and educational aims. *College English*, 75(1), 42-60.
- Gallagher, C. W. (2014). Immodest witnesses: Reliability and writing assessment. *Composition Studies*, 42(2), 73-95.
- Gallagher, C. W. (2016). What writers do: Behaviors, behaviorism, and writing studies. *College Composition and Communication*, 68(2), 238-65.
- Galtung, J. (1969). Violence, peace, and peace research. *Journal of Peace Research*, 6(3), 167-191.
- Galtung, J. (1990). Cultural violence. *Journal of Peace Research*, 27(3), 291-305.
- Gee, J. P. (1989). Literacy, discourse, and linguistics: Introduction and what is literacy? *Journal of Education*, 17(1), 5-25.
- Goldhaber, D., Lavery, L., & Theobald, R. (2015). Uneven playing field? Assessing the teacher quality gap between advantaged and disadvantaged students. *Educational Researcher*, 44(5), 293-307. <http://dx.doi.org/10.3102/0013189x15592622>
- Gupta, A. (2012). *Red tape: Bureaucracy, structural violence, and poverty in India*. Durham, NC: Duke University Press.
- Haertel, E. H. (1999). Validity arguments for high-stakes testing: In search of the evidence. *Educational Measurement: Issues and Practice*, 18(4), 5-9. <http://dx.doi.org/10.1111/j.1745.3992.1999.tb00276.x>
- Haertel, E. (2013). Getting the help we need. *Journal of Educational Measurement*, 50(1), 84-90.
- Harklau, L. (2000). From the "good kids" to "the worst": Representations of English language learners across educational settings. *TESOL Quarterly*, 34(1), 35-67.
- Huot, B. (2002). *(Re)Articulating writing assessment*. Logan, UT: Utah State Press.
- Inoue, A. B. (2012). Grading contracts: Assessing their effectiveness on different racial formations. In A. B. Inoue & M. Poe (Eds.), *Race and writing assessment* (pp. 79-94). New York, NY: Peter Lang.
- Inoue, A. B. (2014). Theorizing failure in U.S. writing assessments. *Research in the*

- Teaching of English*, 48(3), 330-352.
- Inoue, A. B. (2015). *Antiracist writing assessment ecologies: Teaching and assessing writing for a socially just future*. Fort Collins, CO: WAC Clearinghouse/Anderson, SC: Parlor Press.
- Kane, M. T. (1992). An argument-based approach to validity. *Psychological Bulletin*, 112, 527-535. <http://dx.doi.org/10.1037/0033-2909.112.3.527>
- Kane, M. T. (2001). Current concerns in validity theory. *Journal of Educational Measurement*, 38(4), 319-342. <http://dx.doi.org/0.1111/j.1745-3984.2001.tb01130.x>
- Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17-64). Westport, CT: American Council on Education/Praeger. <http://dx.doi.org/10.1111/jedm.12000>
- Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50(1), 1-73.
- Kane, M. T. (2015). Explicating validity. *Assessment in Education: Principles, Policy and Practice*, 22, 198-211. <http://dx.doi.org/10.1080/0969594X.2015.1060192>
- Kohn, A. (2011). The case against grades. *Educational Leadership*, 69(3), 38-33.
- Lane, S., Parke, C. S., & Stone, C. A. (1998). A framework for evaluating the consequences programs of assessment. *Educational Measurement: Issues and Practice*, 17(2), 24-28. <http://dx.doi.org/10.1111/j.1745-3992.1998.tb00830.x>
- Lane, S., & Stone, C. A. (2002). Strategies for examining the consequences of assessment and accountability programs. *Educational Measurement: Issues and Practice*, 21(1), 23-30. <http://dx.doi.org/10.1111/j.1745-3992.2002.tb00082.x>
- Linn, R. L. (1997). Evaluating the validity of assessments: The consequences of use. *Educational Measurement: Issues and Practice*, 16(2), 14-16. <http://dx.doi.org/10.1111/j.17453992.1997.tb00587.x>
- Lucas, C. K. (1988). Toward ecological evaluation. *The Quarterly of the National Writing Project and the Center for the Study of Writing*, 10(1), 1-17.
- Matthews, C. (2014). He dropped one letter in his name while applying for jobs, and the responses rolled in. *The Huffington Post*. Retrieved from https://www.huffingtonpost.com/2014/09/02/jose-joe-job-discrimination_n_5753880.html
- Mehrens, W. A. (1997). The consequences of consequential validity. *Educational Measurement: Issues and Practice*, 16(2), 16-18. <http://dx.doi.org/10.1111/j.1745-3992.1997.tb00588.x>
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed.) (pp. 13-103). New York, NY: Macmillan.
- Mills, C. W. (1997). *The racial contract*. Ithaca, NY: Cornell University Press.
- Morgan, P. L., Farkas, G., Hillemeier, M. M., Mattison, R., Maczuga, S., Li, H., & Cook, M. (2015). Minorities are disproportionately underrepresented in special education: Longitudinal evidence across five disability conditions. *Educational Researcher*, 44(5), 278-292. <http://dx.doi.org/10.3102/0013189x15591157>
- Moss, P. A. (1994). Can there be validity without reliability? *Educational Researcher*, 23(2), 5-12.
- Moss, P. A. (1998). The role of consequences in validity theory. *Educational Measurement: Issues and Practice*, 17(2), 6-12. <http://dx.doi.org/10.1111/j.1745-3992.1998.tb00826.x>

- Mossberg, D. (Producer). (2012). *Malcolm X debate with James Baldwin - September 5, 1963* [video file]. Retrieved from https://www.youtube.com/watch?v=-JIp9_IIV3s
- O'Neill, P. (2011). Reframing reliability for writing assessment. *Journal of Writing Assessment*, 4(1). Retrieved from <http://journalofwritingassessment.org/article.php?article=54>
- Ogbu, J. U. (1999). Beyond language: Ebonics, proper English, and identity in a Black-American speech community. *American Educational Research Journal*, 36(2), 147-184.
- Ogbu, J. U. (2004). Collective identity and the burden of "acting white" in black history, community, and education. *The Urban Review*, 36(1), 1-35.
- Parkes, Jay. (2007). Reliability as argument. *Educational Measurement: Issues and Practice*, 26(4), 2-10.
- Peresie, J. L. (2009). Toward a coherent test for disparate impact discrimination. *Indiana Law Journal*, 84(3), 773-802.
- Poe, M., Elliot N., Cogan, J. A., Jr., & Nurudeen, T. G., Jr. (2014). The legal and the local: Using disparate impact analysis to understand the consequences of writing assessment. *College Composition and Communication*, 65(4), 588-611.
- Popham, W. J. (1997). Consequential validity: Right concern-wrong concept. *Educational Measurement: Issues and Practice*, 16(2), 9-13. <http://dx.doi.org/10.1111/j.1745-3992.1997.tb00586.x>
- Pulfrey, C., Buchs, C., & Butera, F. (2011). Why grades engender performance-avoidance goals: The mediating role of autonomous motivation. *Journal of Educational Psychology*, 103(3), 683-700.
- Reckase, M. D. (1998). Consequential validity from the test developer's perspective. *Educational Measurement: Issues and Practice*, 17(2), 13-16. <http://dx.doi.org/10.1111/j.1745-3992.1998.tb00827.x>
- Rubie-Davies, C. M. (2006). Teacher expectations and students self-perceptions: Exploring relationships. *Psychology in the Schools*, 43(5), 537-552.
- Said, E. W. (1977). *Orientalism* (3rd ed.). London, England: Penguin.
- Said, E. W. (2003/1977). *Orientalism* (3rd ed.). London, England: Penguin.
- Said, E. W. (1993). *Culture and imperialism*. New York, NY: Vantage.
- Shepard, L. A. (1993). Evaluating test validity. *Review of Educational Research*, 19, 405-450.
- Shepard, L. A. (1997). The centrality of test use and consequences for test validity. *Educational Measurement: Issues and Practice*, 16(2), 5-24. <http://dx.doi.org/10.1111/j.1745-3992.1997.tb00585.x>
- Shor, I. (1992). *Empowering education: Critical teaching for social change*. Chicago, IL: University of Chicago Press.
- Sternglass, M. S. (1997). *Time to know them: A longitudinal study of writing and learning at the college level*. Mahwah, NJ: Erlbaum.
- Stuckey, J. E. (1991). *The violence of literacy*. Portsmouth, NH: Boynton/Cook.
- United Nations Educational, Scientific and Cultural Organization. (2015). *Education for all, 2000-2015: Achievements and challenges*. Paris, France: United Nations Educational, Scientific and Cultural Organization, Retrieved from <http://unesdoc>.

- unesco.org/images/0023/002322/232205e.pdf
- Van Cleve, N. G. (2016). *Crook county: Racism and injustice in America's largest criminal court*. Stanford, CA: Stanford University Press.
- West-Faulcon, K. (2009). The river runs dry: When Title VI trumps state anti-affirmative action laws. *University of Pennsylvania Law Review*, 157(4), 1075-1106.
- Whithaus, C. (2005). *Teaching and evaluating writing in the age of computers and high-stakes testing*. Mahwah, NJ: Lawrence Erlbaum Associates.
- White, E. M. (1993). Holistic scoring: Past triumphs, future challenges. In M. M. Williamson & B. A. Huot (Eds.), *Validating holistic scoring for writing assessment: Theoretical and empirical foundations* (pp. 79-108). Creskill, NJ: Hampton.
- White, E. M. (2001). The opening of the modern era of writing assessment: A narrative. *College English*, 63(3), 306-320. <http://dx.doi.org/10.2307/378995>
- White, E. M. (2005). The scoring of writing portfolios: Phase 2. *College Composition and Communication*, 56(4), 581-600.
- White, E. M., & Thomas, L. L. (1981). Racial minorities and writing skills assessment in the California State University and Colleges. *College English*, 43(3), 276-283.
- Young, I. M. (2011/1990). *Justice and the politics of difference*. Princeton, NJ: Princeton University Press.
- Žižek, S. (2008). *Violence*. New York, NY: Picador.