

1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
10	10
11	11
12	12
13	13
14	14
15	15
16	16
17	17
18	18
19	19
20	20
21	21
22	22
23	23
24	24
25	25
26	26
27	27
28	28
29	29
30	30
31	31
32	32
33	33
34	34
35	35
36	36
37	37
38	38
39	39
40	40
41	41
42	42
43	43
44	44
45	45
46	46
47	47
48	48
49	49
50	50
51	51
52	52
53	53
54	54
55	55
56	56
57	57
58	58
59	59
60	60
61	61
62	62
63	63
64	64
65	65
66	66
67	67
68	68
69	69
70	70
71	71
72	72
73	73
74	74
75	75
76	76
77	77
78	78
79	79
80	80
81	81
82	82
83	83
84	84
85	85
86	86
87	87
88	88
89	89
90	90
91	91
92	92
93	93
94	94
95	95
96	96
97	97
98	98
99	99
100	100

## The *Standards* Will Never Be Enough: A Racial Justice Extension

Mya Poe, Maria Elena Oliveri & Norbert Elliot

To cite this article: Mya Poe, Maria Elena Oliveri & Norbert Elliot (2023): The *Standards* Will Never Be Enough: A Racial Justice Extension, *Applied Measurement in Education*, DOI: 10.1080/08957347.2023.2214656

To link to this article: <https://doi.org/10.1080/08957347.2023.2214656>



Published online: 31 May 2023.



Submit your article to this journal [↗](#)



View related articles [↗](#)



View Crossmark data [↗](#)



# The *Standards* Will Never Be Enough: A Racial Justice Extension

Mya Poe <sup>a</sup>, Maria Elena Oliveri <sup>b</sup>, and Norbert Elliot <sup>c</sup>

<sup>a</sup>Department of English, Northeastern University; <sup>b</sup>Buros Center for Testing, University of Nebraska, Lincoln, 21 Teachers College Hall Lincoln; <sup>c</sup>Department of Humanities and Social Sciences, New Jersey Institute of Technology

## ABSTRACT

Since 1952, the *Standards for Educational and Psychological Testing* has provided criteria for developing and evaluating educational and psychological tests and testing practice. Yet, we argue that the foundations, operations, and applications in the *Standards* are no longer sufficient to meet the current U.S. testing demands for fairness for all test takers. We propose racial justice extensions as principled ways to extend the *Standards*, through intentional actions focused on race and targeted at educational policies, processes, and outcomes in specific settings. To inform these extensions, we focus on four social-justice concepts: intersectionality derived from Black Feminist Theory; responsibility derived from moral philosophy; disparate impact derived from legal reasoning; and situatedness derived from social learning theories. We demonstrate these extensions and concepts in action by applying them to case studies of nursing licensure and placement testing.

## 1 Introduction

We begin this article with a fact and a question. *The fact*: The Black Lives Matter (BLM) movement began by Patrisse Cullors, Alicia Garza, and AyQ (formerly, Opal) Tometi in 2013 in response to the acquittal of Trayvon Martin's killer. The BLM movement brought international response to the ongoing violence against Black people, including Tamir Rice, Elijah McClain, Ahmaud Arbery, Breonna Taylor, Ruth Whitfield, Keenan Anderson, and so many others. *The question*: What is the responsibility of academic disciplines to address the ways their practices, theories, and methods have perpetuated the inequities that drive such violence?

In educational measurement, the work of redress needs to take multiple forms, from expanding construct models to account for cultural and linguistic variation to boosting stakeholder inclusion in test design. The field must also reexamine its foundational assumptions about validity, reliability, and fairness. To these ends, we wonder if the the foundations, operations, and applications in the *Standards for Educational and Psychological Testing* (American Educational Research Association [AERA], American Psychological Association [APA], and National Council on Measurement in Education [NCME], 2014) are sufficient to meet the current U.S. educational and psychological testing demands for "fairness for all test takers" (p. 22).

Before the murder of George Floyd in 2020, it is likely that the measurement community's answer to our question about the fairness sufficiency of the *Standards* would have been affirmative. On the surface, there appeared to be agreement that, for over 70 years, the AERA, the APA, and the NCME had provided judicious guidance on sound testing and validation practices. The intent (to promote sound testing practices), audience (professionals who develop, interpret, and evaluate tests), and purpose (to provide development and evaluation criteria and validity guidelines) of the *Standards* seemed sufficient. In today's world, however, the *Standards* is no longer

**CONTACT** Mya Poe  [m.poe@northeastern.edu](mailto:m.poe@northeastern.edu)  405 Lake Hall, 360 Huntington Ave, Boston, MA.

This article has been republished with minor changes. These changes do not impact the academic content of the article.

sufficient to meet the current U.S. educational and psychological testing demands for justice and fairness for all populations.

To be clear, while the *Standards* has long addressed fairness, it has never sufficiently addressed justice as a primary virtue of social institutions (Rawls, 1999). The term *justice* does not appear in the document. As Camilli and Newton conclude in their analysis of the treatment of validity in the 2014 edition, the document “pauses at the threshold of justice” (2022, p. 119). Yet, as Hood and Hopson (2008) remind us, Black researchers’ contributions were always focused on justice::

For nearly three quarters of a century, one issue has guided and driven the work of African American scholars of educational evaluation. Issues of fairness and equity were at the heart of their inquiry in the 1930s when the doctrine of the land mandated so-called separate but equal school systems for children of color. The issues of fairness and equity were central in their investigations of segregated schools during the pre-Brown and supposedly desegregated schools of the post-Brown eras. The issue of fairness remains uppermost in our minds today as we investigate our woefully inadequate schools for Black children, other children of color, and children from economically oppressed backgrounds (p. 413).

Given Hood and Hopson’s position, the authors of *Standards* would have attended to justice long ago if Black scholars had been in the room consistently since 1952 when the *Standards* first appeared in *American Psychologist* (American Psychological Association. Committee on Test Standards, 1952).

Decolonial scholars have also signaled the insufficiency of the *Standards* and static actions of the measurement community to account for the long-standing lack of awareness and ambivalence to the experiences of Indigenous peoples. The word Indigenous does not appear in the document, and there is no mention of native people in the “relevant subgroups” identified in discussion of context (p. 6). Ellen Cushman (2016), a member of the Cherokee Nation and scholar of the culture and language of that nation, extended the concept of exclusion to the creation of validity arguments:

It’s as though transparency can correct for bias when it basically and simply says to those created by the colonial difference: “Let me show you how the making of this assessment will exclude you using our exclusive and exclusionary process of building this test.” (2016, Validity & Fairness section, para 1)

For Cushman, the intent to promote sound testing practices erases those who were created by colonial difference. The professionals who execute these practices remain exclusionary, and their aim of eliminating bias is redolent with a sense of self-justification that values abstract reasoning over human consequences.

We stand with these and other scholars, including the contributors to this special issue, who believe that traditional assumptions and practices of educational measurement such as those represented in the *Standards* are outdated with the society in which they are embedded. In this article, we explain why such considerations are worthwhile, and we advance an approach more aligned with the current moment. Our approach acknowledges that the *Standards* will always be a retrospective document – a document that is meant to inform not imagine. Through what we call “extensions”—principled applications of contemporary theory and research to specific standards—the measurement community can deliberately innovate the *Standards* with, for example, greater attention to racial justice.

In this article, we use the concept of “extensions” to consider *racial justice extensions* as principled ways to extend the *Standards* through intentional actions focused on race and targeted at educational policies, processes, and outcomes in specific settings that affect Black, Indigenous, People of Color (BIPOC) test-takers and stakeholders. We focus on race because we aim to avoid slippage into other categories of social construction, such as socioeconomic class, that may be more comfortable for professionals unwilling to consider relationships between BLM and the foundations, operations, and applications of measurement (Randall, 2021; Randall, Poe, & Slomp, 2021). *Racial justice extensions* are meant to open every section of the *Standards* to scrutiny, not merely the section on the foundational concept of fairness.

Through the contributions of Black Feminist Theory, moral philosophy, Civil Rights law, and social theories of learning, we propose four key concepts by which the *Standards* might be extended to incorporate racial justice perspectives: *intersectionality* derived from BFT; *responsibility* derived from

moral philosophy; *disparate impact* derived from legal reasoning; and *situatedness* derived from social theories of learning.

To illustrate our racial justice extensions, we present two case study applications: the Next-Generation National Council Licensure Examination (NG-NCLEX<sup>®</sup>) for nurses; and placement testing at California Community Colleges (CCC) of English Language Learners (ELLs). For the NG-NCLEX we include examples of racial justice extensions to the standards for workplace testing and credentialing in Chapter 11 of the *Standards*. For the CCC ELL placement test, we include examples of racial justice extensions to the standards for educational testing and assessment in Chapter 12 of the *Standards*. We conclude by reflecting on the challenges and advantages of the proposed principles and their application for advancing racial justice.

## 2 General Principles for Racial Justice Extensions to the *Standards*

In the late 1960s and early 1970s, the field of educational measurement began to wrestle with questions of fairness and race when reckoning with the implications of Civil Rights legislation aimed at ending racial segregation policies. In the wake of desegregation laws, psychometricians needed to consider disaggregation and consequence – ways to analyze sub-group data related to access and discrimination at school and workplaces. This period ushered in a series of publications, talks, and reports, including the 1969 American Psychological Association’s Task Force on Employment Testing of Minority Groups on correlation of SAT scores and college grades (Stanley & Porter, 1967) and research on the validity of academic predictors for Black and white students (Pfeifer & Sedlacek, 1971).

One of the most widely referenced methods from this period was provided in T. Anne Cleary’s, 1968 article “Test bias: Prediction of grades for Negro and white students in integrated colleges.” The article, published in *Journal of Educational Measurement*, was based on her previous research, including a 1966 report written for the College Board in which she examined SAT scores at three integrated colleges. In both studies, she argued that, based on predictive SAT score models and first-year GPAs, the SAT overpredicted in favor of Black students when a common regression line was used. While attention is often given to disjuncture between predictor test and criterion variables (Geisinger, 2022), little attention has been given to the sample itself. In one of the schools studied, students’ race was determined by having two individuals independently examine students’ pictures in school files. Based on those pictures, students were placed into racialized categories. When a student’s racial identity was unclear, Cleary turned to NAACP (National Association for the Advancement of Colored People) membership lists. Students on the NAACP membership list were listed as “Negro.” Cleary’s explanation of how students were divided into racialized categories through photos is what we would call today racist “lookism,” i.e., assuming that someone identifies as a particular racial or ethnic group based on their phenotype. Bias based on appearance, as Tietje and Cresap (2005) argue, is a fundamental form of prejudice that creates unjust barriers to equal opportunity in education and the workplace – and, as the Cleary (1966) study reveals, can result in a racist sampling plan from which we will never know if inaccurate interpretations were made. Yet, based on such a potentially flawed design, Cleary’s model became the first psychometric approach to examine if tests were fair.

Beyond Cleary’s work on bias, measurement researchers also point to the 1976 *Journal of Educational Measurement* special issue on bias in selection (Petersen & Novick, 1976). After reviewing work such as Cleary’s (termed the Regression Model) and Darlington’s (the Culture-Modified Criterion Model), Petersen and Novick argue that “the concepts of culture-fairness and group parity are neither useful nor tenable, and the models spawned from them should not enjoy institutional endorsement.” (p. 28). Petersen and Novick advanced a threshold model in which applicants are selected with the highest expected utility. The threshold model’s advantage is that it requires an explicit public statement of utilities for each subpopulation if utility differences exist (or, if not, a statement of equal risk). They conclude, “If I do not like the utilities *You* provide, an informative public debate will no doubt ensue, with neither of us claiming any axiomatic justification for our utilities. The important thing is that the discussion be public with all interested parties participating in

the debate” (p. 28). While we question the existence of axiomatic principles below in our discussion of purism and formalism, we appreciate the role of public stakeholder engagement when linked to evidence of group differences in Petersen and Novick’s model.

Today, if we consider the move to culturally relevant, culturally responsive, and culturally sustaining assessment practices (Del Rosario Bastera, Trumbull, & Solano-Flores, 2011; Hood, Hopson, & Frierson, 2005; Sato, 2017; Solano-Flores & Nelson-Barber, 2001. See O’Dwyer et al., “Enacting a Processing for Developing Culturally Relevant Classroom Assessments” in this special issue), it would seem that Petersen and Novick’s argument holds the potential to carry the day. To the best of our knowledge, a combined articulation of psychometric models and dialogic engagement, such as described by Evans in “Applying a Culturally Responsive Pedagogical Framework to Design and Evaluate Classroom Performance-Based Assessments in Hawai’i” in this special issue, remains limited in large-scale assessments. Articulated values remain elided, often defaulted to a single score. Attention to consequences for historically marginalized stakeholders is viewed as ancillary to upholding foundational principles. In the case of race, the effect is the “marriage of racist policies and racist ideas that produces and normalizes racial inequities” (Kendi, 2019, pp. 17–18). It is Ibram X. Kendi’s very definition of racism.

Because racism and racial injustices continues in ways seen and unseen in measurement, we offer a principled way to address them in the *Standards* through the concept of racial justice extensions. We define *racial justice extensions* as principled (ethically bound and situationally-alert) ways to extend views of measurement (such as those found in the *Standards*) through intentional actions *focused on race*. Functionally, racial justice extensions are targeted at educational policies, processes, and outcomes, with special attention to the complex ways they are embedded, interact, and operationalized in specific settings. In debt to, and informed by Randall (2021), we propose that these actions focus on race, including its manifestations through colonialism, “exclusively and explicitly” (p. 83). We believe that an explicit racial focus has three benefits:

- allows granular examination of ways that power structures manifest in unique contexts based on putative classifications that are, in fact, socially constructed;
- allows new educational directions to be determined and implemented with a precision often absent when classifications such as socioeconomic status are used;
- opens doors to justice for all through processes of specific change that both reveal the ways that interlocking structures of domination disenfranchise many and leverage educational reforms at both group and individual levels.

Racial justice extensions acknowledge that justice involves both the intention to enact racial justice and the agency to achieve it; recognize the presence of interlocking structures of dominance that appear as institutional norms but are, in fact, systems that frustrate racial justice; posit that these interlocking structures of dominance have causal relationships to structural racism; name these structures, analyze them, and examine their causal effects in specific contexts; identify specific strategies to achieve equity in these contexts; and monitor these contexts to ensure that racial justice is maintained.

In turn, racial justice extensions specifically applied to the *Standards* can be enacted according to the following principles:

- acknowledge that the foundations, operations, and applications in the *Standards* have a potential to be directed toward racial justice;
- insist that social movements such as Black Lives Matter and social theories such as Black Feminist Theory have transformative roles to play in educational measurement;
- recognize the presence of interlocking structures of dominance (Collins, 1989) that appear in the *Standards* themselves, such as the fact that Indigenous people are not merely invisible in the *Standards* but that such invisibility is interwoven with genocide—i.e., the use of statistical

approaches that demand “significance” testing for groups then erasing those groups when their population does not meet sample size requirements;

- posit that these interlocking structures in the *Standards* have causal relationships to structural racism in terms of foundational measurement concepts involving evidence of validity and reliability, including interpretation and use arguments that may, in fact, reinforce racist validity arguments reliant on the construction of those understood to be inferior;
- name these structures in the *Standards* in terms of specific contexts in which BIPOC (Black, Indigenous, People of Color) test-takers may experience negative effects of interpretation and use arguments;
- identify specific strategies for the achievement of equity in these contexts, with specific attention to strategies available by broadening the *Standards* Management Committee as well as the Joint Committee on the *Standards*, with BIPOC scholars, researchers, and educational community members, including educational administrators, teachers, and parents, and guardians; and
- monitor the *Standards* sponsoring institutions – the AERA, APA, and NCME – as an obligation to ensure that racial justice is maintained through position statements, guidelines, and case studies emphasizing justice and fairness.

### 3 Four Intellectual Traditions That Inform Racial Justice Extensions in the *Standards*

In previous work on fairness in educational measurement, we have invited measurement colleagues to consider expanded traditions of fairness (Oliveri, Poe, & Elliot, 2023). These expansions include both discussions of technical concepts, such as bias and reliability/precision, as well as concepts from culturally sustaining assessments, moral philosophy, legal precedent, and social theories of learning. We now demonstrate how those ideas can inform the entire *Standards*, not just the foundational concept of fairness.

#### 3.1 Black Feminist Theory (BFT): Intersectionality and Coalition Building

We are guided by BFT’s long tradition including scholars, teachers, and community leaders – such as Jessie Redmon Fauset (1924), the Combahee River Collective (1977), Davis (1981), Hooks (1992), Patricia Hill Collins (2019), Kimberlé Crenshaw (1989; 1991), and so many other Black women – who have given voice to “the historical reality of Afro-American women’s continuous life-and-death struggle for survival and liberation” and helped us to understand “how their sexual identity combined with their racial identity to make their whole life situation and the focus of their political struggles unique” (Combahee River Collective, 1977). BFT centers on the multi-dimensional facets of Black women’s identities within social systems that have historically denied and denigrated their existence. Collins (1989), for example, argued that “Black women’s political and economic status provides them with a distinctive set of experiences that offer a different view of material reality than that available to other groups” (pp. 746–747). She used the term *matrix of domination* to name the interlocking structures of domination through gender, race, socioeconomic status, and other identity markers.

Taylor (1998; see also Collins, 1989, 2000, 2019) advanced four tenets at the heart of BFT: positive identity formation that rejects negative stereotypes of Black women; dismantling the interlocking structures of domination that harm Black women; combining intellectual theory and political activism; and recognizing and drawing upon [Black women’s] own lives and histories as a source of power. Today, a new generation of Black feminist theorists such as Moya Bailey (2021) and Jamila Lyiscott (2019) continue to give voice to the ways Black women’s bodies and speech are surveilled in a society dominated by white supremacy.

Of the tenets of BFT, the notion of “intersectionality,” coined by Crenshaw (1989; see also Crenshaw, 1991) – and indebted to Hill Collins’ *matrix of domination*—is especially valuable for measurement researchers. Crenshaw argued that intersectionality is a

provisional concept linking contemporary politics with postmodern theory. In mapping the intersections of race and gender, the concept does engage dominant assumptions that race and gender are essentially separate categories. By tracing the categories to their intersections, I hope to suggest a methodology that will ultimately disrupt the tendencies to see race and gender as exclusive or separable (1989, p. 1244).

Significantly, Crenshaw (1989) proposed the concept of intersectionality based on her study of battered women's shelters in Los Angeles. From that study, she realized that the convergence of race, gender, and class domination was unique for battered women of color. She stated that "intervention strategies based solely on the experiences of women who do not share the same class or race backgrounds will be of limited help to women who because of race and class face different obstacles" (p. 1246). Today, as Carbado and Harris (2019) note, intersectionality has been taken up internationally in the pursuit of human rights, political philosophy, and the law.

Intersectionality has also been taken up in educational research (Codioli McMaster & Cook, 2019). Within these discussions of intersectionality, intersectionality has taken on multiple meanings. As Sibbett (2020) notes, "Intersectionality has been understood by some as a theory of multiple marginalization, and by others . . . as a theory of multiple identities" (p. 5). For the purposes of educational assessment, we see the *multiple identities* version of intersectionality as useful (Russell, Szendey, & Kaplan, 2021). We see the *multiple marginalization* version of intersectionality as transformative (Jang, 2018). Why? Because an approach to intersectionality that explores multiple marginalization seeks to look beyond a summation of individual identity markers to the "dynamic interaction between individual and institutional factors" (Hancock, 2007, p. 64). Furthermore, reporting solely by (sub) group differences does not contextualize individuals within larger histories. Failure to recognize that Black women do not share the same historical realities as white women, we believe, is a form of violence associated with white logic (Garcia, López, & Vélez, 2017).

In raising issues of violence, we agree with Adams and Salter (2019) that epistemological violence results from hegemonic science and its purist aims. When measurement posits a formal existence – as it often does in the *Standards* with reliance on "professionals in the testing process" (p. 4) who stand apart from individuals affected by the testing process, institutions, and society – violence occurs in the gap between formalism and situatedness. Within this gap, violence is likely to occur for an individual who will be subject to claims of inability to demonstrate proficiency on a given construct represented in a test. These individuals are constructed as what Teo (2010; see also Spivak, 1988) has termed "the Other:" those understood to be inferior or problematic – the targets – of knowledge theories that are, in fact, one among many. One need only read a sentence in the *Standards* such as "Candidates who fail may profit from information about the areas in which their performance was especially weak" to understand why many feel that the forms of evidence regarding validity and reliability have been assembled to characterize inferiority (AERA et al., 2014, p. 176). The *Standards* is also silent on how disaggregated data may be used to advance claims of injustice, rather than merely identifying performance "gaps" between assessments and classroom performance (AERA et al., 2014, p. 184). It is no wonder that a plausible interpretation of the *Standards* suggests that it is, in many ways, a document that ultimately reveals how the making of an assessment will use multiple forms of evidence to exclude certain individuals and groups and create the Other. It is hard to ignore the presence of epistemological violence.

To combat the violence and marginalization experienced by Black women, Black Feminist theorists have called for coalition building – a sentiment captured in the Combahee River Collective (1977): "Although we are feminists and Lesbians, we feel solidarity with progressive Black men and do not advocate the fractionalization that white women who are separatists demand" (What We Believe, para 4).

To define a coalition, a concept significant for racial justice extensions, we first recognize that "the organized identity groups in which we find ourselves are in fact coalitions, or at least potential coalitions waiting to be formed" (Crenshaw, 1991, p. 1229). Second, we do well to realize that these identity categories do not fragment collaboration. Rather as Roberts and Jesudason (2013) argue, "Attending to categorical differences [such as identity] enhances the potential to build coalitions between movements and makes them more effective at organizing for social change" (p. 315). Third,

we recognize that coalitions are aimed at identifying and eliminating structures of domination that, even with common origins, impact groups and individuals differently. Fourth, following Randall (2021), we agree that coalition actions that focus on race hold the potential to avoid race hegemony. Fifth, we focus on coalitions to call attention to the importance of individual differences that are “dynamic, interacting, and influenced” by their learning context (Gregersen & MacIntyre, 2014, p. xxvii), as well as in their emotional, cognitive, and motivational reactions to injustice (Stavrova & Schlösser, 2015).

In sum, by using a *multiple marginalization* definition of intersectionality, the racial justice extension we propose moves beyond single category analyses to ask what is distinct to specific groups, such as Black women, while using a coalition approach to make dynamic linkages across groups. An intersectional approach to coalition building provides an invaluable lens through which we can understand evidence related to fairness, validity, and reliability as it is interpreted by new coalitions organized with attention to identity.

### 3.2 Moral Philosophy: Responsibility

In its ethical dimension, philosophy – the investigation of basic principles that underlie all branches of knowledge – offers abstract systems that provide perspectives necessary to grapple with dominant cultures and the nature of virtue within them (MacIntyre, 2007). In “Philosophical perspectives on fairness in educational assessment,” Zwick, Dorans, and Cook (2016) provide three case studies to demonstrate how three philosophical systems – Aristotelian pre-structuring, Nozick’s libertarian viewpoint, and Rawls (1999) social justice framework – would apply to educational measurement. After reviewing the applications, Zwick and Dorans reach the following conclusion:

Can our three, or any, philosophical perspectives provide conclusive answers to our fairness dilemmas? We don’t think they can. Reading Aristotle will not tell us what attributes are most important for college students, nor will libertarian doctrine tell us what acquisitions are just or how to compensate for those that aren’t. Rawls cannot tell us what inequalities will ultimately better the situation of the least advantaged in society. However, these philosophical perspectives can provide us with alternate lenses through which to view assessment fairness issues and can at least encourage us to ask the right questions (p. 279).

Whether one agrees or not with these generalizations, it is significant in an analysis of racial justice that the three selected philosophers for case study analysis are white men. Their default selection leaves us wondering what conclusions might be drawn had other philosophers such as Iris Marion Young or Martha Nussbaum been selected for analysis? What would have been demonstrated had the later work of the Caribbean-American philosopher Charles W. Mills on racial liberalism been on Zwick’s and Dorans’s bookshelves in 2016?

Young died in August of 2006, leaving her manuscript *Responsibility for Justice* (2011) in a completed yet unfinalized form. Recognizing the reality of a life cut short, in her foreword for the published manuscript Nussbaum described Young as “one of our era’s most creative and influential political theorists” (2011a, p. ix). For Young, relationships between guilt and responsibility were at the heart of her philosophy. Whether one is guilty or not of a given offense, each individual nevertheless bears responsibility. Since the privatized culture in which most of us live, she holds, produces a political vacuum in which there is little sense of collective action, we must work for public action to intervene in offense. Rather than settle into indifference, Young believes that each of us is under obligation to adopt a forward-looking attitude to organize collective action for change. In terms of educational measurement as represented in the *Standards*, Young’s call for defined responsibility becomes significant as we consider, for instance, the meaning of evidence of fairness, validity, and reliability for individuals. Her work is equally important in its connection to BFT coalition formation.

In her own right, Nussbaum’s uptake of responsibility takes a very specific form in her capabilities project (2011b). Dissatisfied with the deficit model of Gross Domestic Product (GDP) as a crude measure emphasizing strictly economic growth, Nussbaum and her colleagues created



a highly influential human development model that has had an important impact on a variety of international agencies, including the World Bank and the United Nations Development Programme. While a description of her model is beyond the scope of this paper, there are important connections to educational measurement to be pursued through three essential capability principles: each person is taken as an end, with attention given to opportunities available for each person; a focus on choice, with attention given substantial freedom; and pluralist value, with attention to the fact that capability differs in quality, not mere quantity, and that capability achievements cannot without distortion be reduced to a numerical scale (Nussbaum, 2011a, pp. 18–19). Although she grants that assignment of responsibilities should be provisional, Nussbaum does believe – as does Young in her rejection of blame as a backward-looking concept – that there can be a defined forward-looking articulation of capability:

“What is each person able to do and to be?” In other words, the approach takes each person as an end, asking not just about the total or average well-being but about the opportunities available to each person (2011b, p. 18).

Applied to educational measurement, Nussbaum’s insistence on the individual can help measurement researchers consider “the extent to which the construct being assessed has equivalent meaning across the individuals and groups within the intended population of test takers” when assumed equivalence fails (AERA et al., 2014, p. 52). For test developers, insistence on the individual would mean a reconsideration of terms such as “relevant subgroups.” For those who interpret scores, capability invites differentiation between advocacy of sound testing practices and erasure of those silenced by those very practices as forms of evidence become, however unwittingly, the construction of failure.

As is the case with Young and Nussbaum, Mills (2017) is critical of what might be called liberal imperialism – a form of social engineering in which principles appear emancipatory but, in reality, are assimilatory. As he writes, when you add together those who are intentionally excluded in “liberalism’s most celebrated manifestos, treatises, and declarations of human rights, you get a litany of oppressions rather than a list of emancipations” (2017, p. xiii). Yet, seeking to transform systems from within – again, as do Young and Nussbaum – Mills calls for Black radical liberalism. In its aim, this new political philosophy “seeks to transform liberalism to make it responsive to the realities of the black diasporic experience in modernity and the correspondingly necessary reordering of liberal normative priorities” (p. 204). While traditional liberalism has restrictively attended to capitalism and class exploitation, Black radical liberalism places attention on race, white supremacy, and racial exploitation to achieve social justice.

Indeed, Mills has very specific directions rethinking traditional liberalism to achieve justice, from rejecting concepts of an ideal theory of justice to narrowing Rawlsian concept of compensatory justice to that of rectificatory justice aimed to correct wrongful distribution. We believe that there are good reasons to consider the *Standards* a manifestation of liberal imperialism and propose, as we have throughout this paper, modifications such as racial justice extensions that may transform its contents. We believe that such extensions are congruent with the emphasis on responsibility by Young, the call for a capability framework by Nussbaum, and the possibility of rectificatory justice by Mills.

Returning to Zwick, Dorans, and Cook (2016), we ask, Can our three philosophical perspectives provide provisional, case-specific answers to racial injustice? We think they can. Reading Young can help us identify specific responsibilities for justice that extend beyond blame and allow us to organize collective action for change. Nussbaum can tell us how we will ultimately better the situation of the least advantaged in society through a fierce devotion to capabilities. And through applications of Black radical liberalism, Mills can tell us, with some specificity, how to compensate for harm through rectificatory justice aimed to correct wrongful distribution. For all three philosophers the very idea of a universal set of attributes, the essence of theory idealization and a devotion to white universalism, is best replaced by attention to individuals and their diverse ways of ways of knowing and thinking and doing.

### 3.3 Law: Local Civil Rights and Disparate Impact

U.S. Civil Rights legislation has brought various possibilities to address fairness and justice through legal means in the areas of education, employment, voting, and housing. Rightfully, many critics argue that the promise of U.S. Civil Rights laws have been mixed (Beerman, 2001). As Bell (1991; see also Bell, 1980) noted, “the commonly held view of racial advancement as a slow but steady surge forward is wrong . . . . In fact, what is deemed ‘progress’ is cyclical rather than linear” (p. 597).

The shortcomings of the Civil Rights movement to advance racial justice gave birth to movements such as Critical Race Theory (CRT), which has sought to openly critique (but not jettison) Civil Rights legal doctrine and its role in sustaining racism (Crenshaw, Gotanda, & Peller, 1995). Notably, CRT aims to advance “new, more nuanced approaches . . . [These approaches are] required to combat the types of subtle, unconscious, or institutional racism that [are] developing and that [are] more deeply entrenched and difficult to combat than the former overt variety” (Delgado, 2009, pp. 1510–1511).

For the purposes of fairness and racial justice, if we rely only on existing national legal precedent – the *minimum* of compliance – we do not make sufficient progress in the achievement of rectificatory justice. Legal scholar Olutade C. Johnson (2016) explores how subnational efforts to redress discrimination can provide examples of state and local laws can be insightful in seeing “the limitations of the federal civil rights model in addressing subtle bias and structural exclusion” (p. 117). For instance, subnational governments have passed laws “limiting employers from taking into account an applicant’s credit history or arrest and conviction record before making employment decisions” (Johnson, 2016, p. 119). Other state and local laws have addressed discrimination related to sexual orientation, the source of an applicant’s income in housing, and employment status.

Likewise, if legal notions such as disparate impact – defined as unintended racial differences in outcomes resulting from facially neutral policies or practices and evaluated in terms of practical and statistical significance – are taken up outside the courts in local settings, they can be used for methodological innovation. Such an approach combines local civil rights approaches and disparate impact reasoning. Poe and Cogan (2016), for example, argue that the process of *burden-shifting* used in Civil Rights cases can provide a useful heuristic. Briefly, if statistical data obtained through methods such as those used by Petersen and Novick (1976) and discussed in Geisinger (2022) raise a *prima facie* case of discrimination (i.e., the data demonstrates that the challenged practice or policy results in significant disparities between groups based on race), the burden shifts to the defendant. Then, the defendant is to articulate a nondiscriminatory reason for the apparently discriminatory outcome. If the defendant fails to articulate a nondiscriminatory reason, the plaintiff prevails. If the defendant can articulate a nondiscriminatory reason, the burden shifts back to the plaintiff to demonstrate either existence of a less discriminatory approach or that the articulated nondiscriminatory reason is simply a pretext for discrimination. The burden-shifting approach links statistical evidence with contextual analysis and, as such, shifts the collection of evidence gathered in support of claims to a process of dialogue and reassessment. As such, this approach potentially offers a partial augmentation to Cushman’s claim that the creation of validity arguments too often may be distilled to exclusion and deficit (2016, Validity & Fairness section, para 1). If used before an assessment with the assumption that disparate impact may occur, burden-shifting allows extended thought experiments to be performed by stakeholders as part of the anticipatory frameworks described below in the discussion of our case studies.

In sum, as long as the measurement community uses a model of legal precedent that relies on compliance rather than advancement, the project of racial justice will remain stalled. Consequences, as Messick (1980) long ago argued, matter. And consequence is not merely about test interpretation; consequence extends to decision-making and policy impact (Newton & Shaw, 2014). Through attention to local innovations in discrimination law and the dialogic process enacted through the burden-shifting approach, measurement researchers may better anticipate the uneven impact of assessment practices on different populations. As well, collaborating with legal scholars, researchers can investigate the varied ways that innovation and dialog provide forward-looking means to advance racial justice through law.

### 3.4 Education: Social Theories of Learning and Situatedness

Social learning theories may be identified by, at least, the following three perspectives:

- Sociocognitive, in which individuals are understood as agentic in terms of behaviors that are self-organizing, proactive, self-reflecting, and self-regulating (Bandura, 1986; Mislevy, 2018).
- Sociocultural, in which individuals are understood as embedded in situations inseparable from cultural contexts, that individuals learn through social interaction, and that learning is mediated through language and action (Prior, 2006; Vygotsky, 1978).
- Bioecological, in which individuals are understood as existing within the phenomenon of continuity and change, both as individuals and as groups, over individual life courses and across generations (Bronfenbrenner & Morris, 2006).

Such perspectives trace the way human development is mediated by cultural and social contexts, with special emphasis on “tools” of participation (ranging from language to physical objects) as they are related to individuals, communities, and wider societies with which learners interact. Social theories of learning also include recognition of the types of social and community resources. These resources include concepts such as “community cultural wealth” defined by Yosso as “the repertoire of knowledge, skills, abilities, and contacts possessed and utilized by Communities of Color to survive and resist oppression” (2005, p. 77). They also include “funds of knowledge” defined by Moll and colleagues as “the historically accumulated and culturally developed bodies of knowledge and skills essential for individual functioning and well-being that culturally diverse populations bring to educational contexts” (1992, p. 133).

In their applications to educational measurement, social theories of learning are useful because of their attention to the situatedness of learning and assessment (Sato, 2017). In each of the three theoretical perspectives above, relationships among individuals and their environments are interconnected; learning cannot be separated from context. For example, in cognitive science, Robbins and Aydede (2009) define situated cognition as the foundational premise that the mind is embodied in the body, that cognitive activity is embedded in the natural and social environment, and individual cognitive boundaries extend beyond the individual. The mind is therefore seen as embodied, embedded, and extended. This sense of embodiment, in turns, is augmented by further specification of culture.

In the case of racial justice, situatedness also requires responsibility because student learning in the U.S. is undeniably shaped by the historical legacies of racism and colonialism, as well as the inadvertently racist practices within current classrooms. The Nā Hopena A‘o (HĀ) framework in Hawaii provides a promising model of situated responsibility. Sponsored by Hawaii Board of Education, the Indigenous framework focused on HĀ, or breath, and includes belonging, responsibility, excellence, aloha, total well-being, and strengthened culture. As Kasten (2022) demonstrates in her account of the framework since its 2015 origin, it is possible to design and delivery decolonial assessment methods that promote critical reflection and intentionally distribute responsibility within large educational systems. Shultz (this special issue) compellingly writes about the harm of irresponsibility:

When we, [BIPOC researchers], do not see ourselves in the standards that guide our test development work, when we are told that we do not need to be intentional in our integration of culture and language in our work because it is “understood,” when our communities do not see themselves in the curriculum or assessments that are integral parts of their children’s education, when society tells our families that “the reality is” their children need to adapt to a western and English-only world, and when items on an assessment do not hold resonance for students in marginalized communities, then we should not be surprised when the education system, assessments, and their results do not hold any value in our communities.

In the case of Hawai‘i, situatedness requires responsibility for colonial harm and recovery of Indigenous ways of knowing, sovereign pedagogies, and community and cultural validity.

In addition to the critical work in Hawai'i, we can also point to other possibilities for situatedness, including next-generation empirical work on fairness evident in the third generation of DIF analyses, and accompanying statistical methodology, described by Zumbo et al. (2015) as ecological modeling. In this approach, differential item functioning (DIF) is conceived as occurring because of characteristics that are not relevant to the ability of interest and, therefore, the test purpose. By highlighting the test situation over a search for the best statistical mousetrap, Zumbo et al. (2015) propose that DIF theory and praxis can be expanded to matters beyond test structure.

Understanding of the situated nature of learning, intentional responsibility, and potential recovery allows communities of assessment designers and community stakeholders to bridge technical and public concerns regarding fairness. Such awareness gives new life to the need for greater attention to what have previously been considered the technical aspects of measurement. Concepts such as construct representation, construct-irrelevant variance, and bias – as well as technical analysis processes such as DIF – are seen in new ways when a score must be interpreted according to an explicit interpretative argument that is meaningful to a specific community (Ercikan & Oliveri, 2013).

#### 4 Discussion: Sufficiency

In our discussion of these four traditions, we are aware that our original critique regarding insufficiency of the *Standards* comes home to roost: Are these four traditions sufficient to meet the current U.S. educational and psychological testing demands for justice and fairness for all populations? Will the traditions be distilled by whiteness perspectives that turn them toward more comfortable subjects – such as multiple identity versions of intersectionality, superficial engagement with moral responsibility, minimum compliance disparate impact, and essentialized social theories of learning?

The simple answer is “no” and “it depends.” Are the four traditions sufficient to meet the moment? *No*. As we suggested earlier, these traditions, as well as the extensions that follow, are not intended to be a horizon of innovation for test design and use. The idea of racial justice extensions, as well as the four traditions that inform them, are one of many bridges that may be built. Other racial justice extensions may look, for example, to Black Queer Studies (Johnson & Henderson, 2005) or may more fully embrace decolonial evaluation (Waapalanexkweew, 2018).

Will the traditions be distilled by whiteness perspectives that turn them toward more comfortable subjects? *It depends*. There is a long tradition in measurement in which innovation is distilled in favor of tradition. For example, while Maul, Iribarra, and Wilson (2016) argue that various versions of constructive realism appear to be well aligned with current views of the situated nature of cognition used by educational and psychological researchers in assessment development, these current views of situated cognition used by educational and psychological researchers are often vague about the precise interpretation of measurement claims. How is it possible to claim that “Higher stakes may entail higher standards of evidence” (AERA et al., 2014, p. 22), when the meaning of “higher standards,” as well as the implied category of lower standards, is so imprecise as to raise questions about the sufficiency of validity evidence itself. Stakeholders may equally wonder about the callousness of those who believe that any test-based inference made about a student may be considered trivial and, thus, in need only of inferior evidence. We would characterize such callousness as violence.

While there is no guarantee that the promise of extensions to the *Standards* will become another exercise in cooption and absorption, we do think there is a possibility the concept of multipurpose assessments as related to our four traditions and racial justice extensions. Raising issues of *purpose pluralism* in measurement, Newton (2017) proposes that assessment design should be simultaneously driven by a multiplicity of assessment purposes that, in turn, require diverse stakeholder perspectives. As a form of realism in action, Newton casts *pluralism* in contrast to *purism*. As an example of purism, we return to the *Standards* (2014) and this claim: “Most educational tests will serve one purpose better than others; and the more purposes an educational test is purported to serve, the less likely it is to serve any of these purposes effectively” (p. 188). In contrast to this form of purism – axiomatic statements incorporating received views understood as self-evident – pluralism emphasizes the following:

emphasis on assessment-based information to make decisions regarding both test use and test impact; expertise emphasizing fidelity to a specific and prespecified cluster of learning outcomes; and engagement focusing on the consequences arising from anticipating assessment impact for various participants. Pluralism invites polyvocality through community and stakeholder input on test use and invites a range of perspectives. As Newton admits, purpose pluralism “*from a measurement perspective*” seems to be quite wrong. Yet from other legitimate perspectives, such as those of teachers (see Shultz in this special issue), pluralism seems quite right (p. 13, author italics).

Awareness of the situationally-fluid nature of extensions allows them to be seen as ever-evolving processes capable, we hope, of achieving racially just assessments. Extensions will forever remain insufficient and deeply necessary.

## 5. Case Studies: A Racial Justice Extension of the *Standards*

In this final section, we provide case studies of racial justice extensions to the *Standards* for nursing licensure and large-scale community college English Language Learning (ELL) course placement. Randall, Slomp, Poe, and Oliveri (2022) have shown that applying targeted principles associated with fairness and justice require us to ask a basic question: How may components, processes, and inferences of a given test or assessment be restructured to provide evidence that racial justice is under consideration in the assessment design process, its implementation, and the claims and inferences that are ultimately drawn? Our case studies, attempt to answer that question by identifying what the *Standards* states and what the racial justice extensions ask.

To suggest interactions between racial justice extensions and the *Standards*, we offer [Tables 1 and 2](#) as a form of racial justice inquiry heuristic.

In the first two columns of both tables, we have identified each of the four traditions discussed above and selected key elements from each. For BFT, for example, we have selected intersectionality as multiple marginalization because of its centrality to other theory elements. The same is true for responsibility as key to moral philosophy, disparate impact as critical to civil rights legislation, and situatedness as essential to social learning theories. The third column identifies the subject of each case study, with specific reference to relevant sections in Part III (testing applications) of the *Standards*. [Table 1](#) therefore targets Chapter 11 (workplace testing and credentialing) of Part III with attention to the accompanying standards, while [Table 2](#) focuses on Chapter 12 (educational testing and assessment) of Part III and the standards for that application. The third column is further divided into two column spanners: What the *Standards* states (containing the specific standard) and what racial justice extensions ask (containing both concept extension questions and potential actions).

These extensions and actions are intended to yield intentional actions contributing to achievement of racial justice. Each question is based on a specific standard in a specific cluster. In designing each question, we have used interrogative pronouns stressing agency (*how*) and agents (*who*) to emphasize strategies by which justice might be achieved in the case at hand and identify those responsible for its contextual achievement. For each action, we have suggested specific strategies directly related to the key elements of the four traditions. We have specified Black women for questions and actions related to BFT and intend that *exclusively emphasis on race* be used throughout the case studies. To demonstrate the scalability of our racial justice heuristic, we turn to the case studies.

### 5.1 Case Study 1: National Council Licensure Examination

The NCLEX<sup>®</sup> is used for licensure of registered and practical nurses in the United States and is historically understood to be a model of evidence-based, electronically delivered licensure aimed fair design and development for more than 15 years. In 2019, the National Council of State Boards of Nursing (NCSBN) announced that a special section of the traditional National Council Licensure Examination (NCLEX) would be implemented to collect data on new items designed to test clinical judgment through a defined measurement model (Dickison, Haerling, & Lasater, 2019). Intended to

**Table 1.** Racial Justice Extensions of Fairness Evidence for Nursing Licensure in the Next-Generation NCLEX.

		Licensure Testing (AERA et al., 2014, pp. 178–182)	
Principles	Key Element	What the <i>Standards</i> states.	What racial justice extensions ask.
		Standard 11.11 Test Validity Across Job Situations	
Black Feminist Theory	Intersectionality	“If evidence based on test content is a primary source of validity evidence supporting the use of a test for selection into a particular job, a similar inference should be made about the test in a new situation only if the job and situation are substantially the same as the job and situation where the original validity evidence was collected.” (p. 181)	<i>Extension Question:</i> Since evidence based on test content is a primary source of validity evidence supporting the use of the Next Generation NCLEX (NGN) for licensure, how can useful pedagogies using BFT be developed for BIPOC students to ensure sound nursing clinical judgement? <i>Possible Action:</i> Pedagogies informed by BFT, such as those that affirm the identities of Black women in clinical settings, should be designed and examined in their ability to improve nursing clinical judgment.
		Standard 11.9 Predictor-Criterion Relationships	
Moral Philosophy	Responsibility	“Evidence of predictor-criterion relationships in a current local situation should not be inferred from a single previous validation study unless the previous study of the predictor-criterion relationships was done under favorable conditions (i.e., with a large sample size and a relevant criterion) and the current situation corresponds closely to the previous situation.” (p. 180)	<i>Extension Question:</i> Who is responsible for ensuring that evidence regarding predictor-criterion relationships is collected under clinical conditions (i.e., with sufficiently large intersectional sample sizes and NGN criterion) and used to determine whether the test corresponds to clinical judgment situations, including those that can account for the various ways that different students may experience a situation? <i>Possible Action:</i> State boards of nursing can play an important role in establishing predictor-criterion relationships using O*NET job titles and promoting culturally relevant and anti-racist curricular pedagogies in local nursing education programs.
		Standard 11.1 Intended Subgroup Interpretations	
Law	Disparate Impact	“Prior to development and implementation of an employment or credentialing test, a clear statement of the intended interpretations of test scores for specified uses should be made. The subsequent validation effort should be designed to determine how well this has been achieved for all relevant subgroups.” (p. 178)	<i>Extension Question:</i> How can a meaningful statement of the intended interpretations of NGN test scores for specified uses be made informed by disparate impact frameworks? <i>Possible Action:</i> Under disparate impact frameworks, the subsequent NGN validation effort should be designed to determine how well the intended interpretations have been achieved for the entire nursing student sample, intersectional nursing student samples, and individual nursing students. If it is found that relevant subgroups do not meet the disparate impact threshold, the burden for investigation of score differences and exploration for alternative test measures is shifted to NCSBN.
		Standard 11.2 Content Domain of Interest	
Social Learning Theories	Situatedness	“Evidence of validity based on test content requires a thorough and explicit definition of the content domain of interest.” (p. 178)	<i>Extension Question:</i> How is NCSBN-CJM aligned with the ways that nursing practice is performed in specific settings? <i>Possible Action:</i> Plans should be in place to gather content domain evidence in terms of alignment and possible disjuncture between NCSBN-CJM and forms of clinical judgment used in specific settings, including settings that serve historically marginalized communities.

**Table 2.** Racial Justice Extensions of Fairness Evidence for English Language Learner Course Placement.

Educational Testing and Assessment (AERA et al., 2014, pp. 195–201)			
Principles	Key Element	What the <i>Standards</i> states.	What racial justice extensions ask.
			Standard 12.6 Computer and Multimedia Tests
Black Feminist Theory	Intersectionality	“Documentation of design, models, and scoring algorithms should be provided for tests administered and scored using multimedia or computers.” (p. 197)	<p><i>Extension Question:</i> When writing assessments are delivered in digital formats, how do those formats mediate the targeted writing construct for BIPOC students? Historically, are there particular barriers for Black ELL writers?</p> <p><i>Possible Action:</i> Plans should be in place to determine how construct mediation impacts the entire student sample, intersectional student samples, including BIPOC and Black ELL students. Attention should be paid to present and historical circumstances – in present, for ELL Black immigrant students and, historically, for Black Language speakers in the U.S.</p>
			Standard 12.13 Placement Decisions
Moral Philosophy	Responsibility	“When test scores are intended to be used as part of the process for making decisions about educational placement, promotion, implementation of individualized educational programs, or provision of services for English language learners, then empirical evidence documenting the relationship among particular test scores, the instructional programs, and desired student outcomes should be provided.” (p. 199)	<p><i>Extension Question:</i> How can a variety of stakeholders be involved in providing empirical evidence documenting relationships among writing assessment test scores, instructional programs, and desired student outcomes?</p> <p><i>Possible Action:</i> Stakeholders should have processes in place to demonstrate alignment among test scores, instructional programs, and desired student outcomes. If alignment is not demonstrated, stakeholders should pause the assessment until alignment is achieved.</p>
			Standard 12.17 Group Differences
Law	Local Civil Rights	“In educational settings, reports of group differences in test scores should be accompanied by relevant contextual information, where possible, to enable meaningful interpretation of the differences. Where appropriate contextual information is not available, users should be cautioned against misinterpretation.” (p. 200)	<p><i>Extension Question:</i> How can various forms of evidence be used to support meaningful interpretation of writing assessment score differences using disparate impact frameworks?</p> <p><i>Possible Action:</i> Informed by local civil rights, plans should be in place to enable meaningful interpretation of traditional group, intersectional group, and individual student differences that are responsive to local contexts.</p>
			Standard 12.19 Instructional Recommendations
Social Learning Theories	Situatedness	“In educational settings, when score reports include recommendations for instructional intervention or are linked to recommended plans or materials for instruction, a rationale for and evidence to support these recommendations should be provided” (p. 201).	<p><i>Extension Question:</i> How is the assessed writing construct aligned with institutional mission, departmental learning objectives, and instructor pedagogies in each California Community College location?</p> <p><i>Possible Action:</i> Plans should be in place for analyses demonstrating both alignment and possible disjuncture among writing assessment score reports in each California Community College location according to the following: institutional mission, departmental learning objectives, and individual course syllabi. Special attention should be paid to how shifts in local student populations invite the opportunity for curricular and assessment innovations that advance justice.</p>

be implemented in the spring of 2023, the Next-Generation NCLEX (NGN) will measure an “observed outcome of critical thinking and decision-making. It is an iterative process that uses nursing knowledge to observe and assess presenting situations, identify and prioritize client concerns, and generate the best possible evidence-based solutions, in order to deliver safe client care” (Currier, 2019). Because of its sponsoring agency leadership status, development processes, and design transparency, the new test is an excellent candidate for racial justice extension to existing practices for a test that is psychometrically sound and consistent with current U.S. nursing practice (Assessment Technologies Institute, 2022; Betts, Muntean, Kim, Jorion, & Dickison, 2019; Nunn-Ellison, Ard, Farmer, & Beasley, 2020).

We begin [Table 1](#) with a question of alignment between the NGN and clinical judgment models. Following *Standard 11.11*, Dickison, Haerling, and Lasater (2019) used a multidisciplinary team to demonstrate that the NCSBN clinical judgment model (CJM) was closely related to the three leading frameworks currently used to provide clinical judgment education to entry-level registered nurses. As such, the judgment model serves as a primary source of validity evidence within nursing curricula. To ensure that useful pedagogies are developed for Black women to leverage clinical judgment success, we posit that BFT be considered. As Jefferies, Goldberg, Aston, and Tomblin Murphy (2018) have shown, Black feminist poststructuralist frameworks allow discussions of the roles of power, social and institutional discourse, language use, values, and agency within clinical judgment contexts. We see these discussions are directly related to the NCSBN-CJM in terms of environmental and individual factors – the “contextual factors that may impact the performance of the cognitive operations within the model” (Dickison, Haerling, & Lasater, 2019, p. 73).

In turning to predictor–criterion relationships, we ask a question related to the responsibility of those who will extend the work of Jefferies, Goldberg, Aston, and Tomblin Murphy (2018) beyond the classroom into actual clinical practice. Since establishing predictor–criterion relationships is at the heart of credentialing tests, we wonder how intersectional sampling plans might be designed to investigate predictive validity according to the O\*NET OnLine (2023) reporting titles (such as Certified Operating Room Nurse or Oncology Nurse) to supplement current published reports of pass rates according to educational background (National Council of State Boards of Nursing [NCSBN], 2022a). Enlisting state boards of nursing for analysis based on workforce surveys could lend a detailed sense of predictive relationship between the test and nursing practice (NCSBN, 2022b; see Florida Center for Nursing, 2020, Appendix A, for a sample workforce survey). Combined with BIPOC student sampling, such a state-based initiative would allow test developers to examine intersectional relationships among factors identified in the workforce survey (such as questions on highest level of nursing education, weeks worked per year, current nursing practice position, specialization, and future employment plans found in the Florida workforce survey) and the NCSBN-CJM. A coalition approach led by state boards would allow interpretation of test results in terms of BIPOC student performance in specific job titles; as a result, nursing programs within individual states could continue to improve curricula along racial justice lines and NGN test developers would be able to modify the test along those same lines.

Our questions regarding disparate impact research are well answered by Banks et al. (2022). In a longitudinal study of factors that predict NCLEX-RN test success at a Historically Black College and University, the authors applied multivariate logistic regression modeling to show that male students were 69% less likely to pass NCLEX-RN exam on first attempt compared to females. There were also differential relationships between test success and course grades for traditional and accelerated students. Their findings remind us that course grades are an important aspect of intersectionality in academic settings and that ensuring success for male students, in this case, is an important facet of equity.

Our final question, related to the content domain of interest, asks for further study of the NCSBN-CJM used specific settings. Analyzing information from an international review of literature on critical care nursing practice, Macey, O’Reilly, Williams, and Cameron (2022) demonstrated



that this form of nursing is unique to regional and socioeconomic contexts. Such questions are related to those on responsibility and would provide valuable information to test designers and nursing stakeholders.

## **5.2 Case Study 2: California Community Colleges ELL Placement**

The second case is drawn from current efforts at California Community Colleges (CCC) to review assessments related to course placement of students who are English Language Learners (ELLs). Operating under a 2021 charter from the Office of the Chancellor, the Assessment Advisory Committee reviews candidate placement tests under its own *Standards for Assessment Test Instrument Review* involving fairness, validity, reliability, scaling, and test administration (CCC, 2021a). CCC works in collaboration with the Buros Center for Testing at the University of Nebraska to investigate a range of assessment issues, including evidence related to criterion and consequential-related validity, as well as inter-rater reliability estimates. In turn, this evidence is used to engage in a recurrent process in which multiple measures – including school transcripts and military experience – are mandated for student placement. An updated *Vision for Success* emphasizes a new process of equitable placement and support aimed to remedy “past unfair practices that over-placed students into remedial education, especially students of color, creating obstacles to their momentum and success” (CCC, 2021b, p. 11). As is the case with the Next-Generation NCLEX, we take the CCC efforts as excellent candidate for the addition of racial justice existing practices. Under an equity framework established legislatively by the California Education Code (2022), California institutions must maintain the following: maintain equity plans; provide matriculation services; adopt and implement assessment plans sensitive to cultural and language differences, adapted as necessary to accommodate differently abled students; provide students with an education plan; and provide annual reports to the chancellor’s office including an assessment of progress toward meeting equity-focused frameworks. This statewide commitment to equity, addressing placement, retention, and progression issues in term of fairness reveals a state-wide commitment to racial justice. Our extensions to the CCC ELL placement test are presented in [Table 2](#). There, we focus on ELL writing placement.

We begin with proposing an extension question and a related action concerning computer and multimedia tests. In previous work, we have argued that technology mediates constructs in ways that may be unanticipated in abstract psychometric models (Katz & Elliot, 2016). Viewed through an intersectional lens, disjuncture between models and their technological delivery may result in unique, anticipated negative consequences for BIPOC, specifically Black, ELL students. If, for instance, a construct model specifies that students must have knowledge of writing conventions and an automated assessment of writing is used to score writing samples, questions should be raised about the adoption of standard edited American language as representational of White Mainstream English (WME) within the scoring algorithm.

Further questions might be raised regarding criteria related to knowledge of conventions that, in fact, may be coded language for the sole promotion of WME (Alim, Rickford, & Ball, 2016; Baker-Bell, 2020; Green, 2004; Rickford, Sweetland, & Rockford, 2013). Such scoring restrictions result in the erasure of language diversity for Black English-speaking students in the U.S. (Hankerson, 2023; Perryman-Clark, 2013) and, hence, raise questions about biased construct coverage in the placement test itself. In raising questions about placement decisions, we extend the *Standards* to identify those responsible for documenting relationships among writing placement test scores, instruction, and student outcomes. Unless a clear relationship between writing placement scores, classroom instruction, and institutional outcomes can be documented, we recommend that the assessment stakeholders pause the assessment. There is little reason to use a score if a student is placed into a curriculum that cannot support writing improvement. In terms of group differences, racial justice extensions allow stakeholders to remain alert to the

potential of a placement test to violate the California Education Code (2022) itself in its acknowledged sensitivity to cultural and language differences. It may, for instance, be that case that the writing task itself is biased and that institutional alignment is therefore impossible (Elder, McNamara, & Congdon, 2003).

In turning to group differences discussed in the *Standards*, we believe that local civil rights frameworks allow assessment stakeholders to shift evidence perspectives. Graphenreed and Poe (2022) have proposed that antiracist and trauma-informed pedagogies can be vectored through the syllabus and related classroom-level artifacts, including assessments, to affirm marginalized student identity and promote individual well-being.

In the same way that advocates of local civil rights focus on benefitting the least advantaged in specific communities, a situated lens provides sources of evidence by which group differences may be understood. In terms of pedagogical interventions resulting from writing placement, curricular alignment – of both the general kind in which mission, departmental learning objectives, and individual course syllabi are aligned, and the specific anti-racist genre proposed by Graphenreed and Poe (2022)—is of paramount importance. Put simply, if the design, models, and scoring methods are not aligned with local pedagogies, the assessment will lack the content-related evidence needed to provide instructional interventions.

## 6 Discussion: Extensions

Conceptualized through racial justice extensions to the *Standards*, each case emphasizes the need to align the assessment with the complex environments in which the candidates presently live and into which the candidates will matriculate. In that both cases already acknowledge the power of educational measurement to produce foundational forms of evidence related to fairness, validity, and reliability, and the racial justice extensions allows those evidence forms to be extended.

In terms of agency, augmented racial justice extensions allow us to realize that both programs appear to rely solely on race and gender to investigate evidence of concurrent validity, predictive validity, and disparate impact. Investigation of intersectionality along the identity lines identified by BFT does not appear to be an option in these exemplar programs at the present time. Such absence may result in data analysis and subsequent investigations so general that they simply reinforce, however unwittingly, structural inequity through invisibility. Strategically, racial justice extensions are best used when they are part of anticipatory design frameworks deployed before the first assessment is given, such Evidence-Centered Design (Mislevy, Steinberg, & Almond, 2003) and expanded Evidence-Centered Design (Arieli-Attali, Ward, Thomas, Deonovic, & von Davier, 2019). The extensions may have special value when used with design frameworks intended to advance justice and fairness such as Standpoint Theory of Action (Hazelton, Nastal, Elliot, Burstein, & McCaffrey, 2021), the Integrated Design and Appraisal Framework (Slomp, 2016) and the Justice-Oriented, Antiracist Validity Framework (Randall, Slomp, Poe, & Oliveri, 2022).

In terms of agents, we propose that the best way to achieve local aims of fairness and justice is to include key stakeholders in all phases of an assessment. Here, (Klitzman's, 2012) study of community board members in Institutional Board Review (IRB) is enlightening. While Section 46.107 (a) of 45 CFR 46—the federal regulations governing the protection of human subjects – requires IRB members to be experts, Sections 46.107 (b) and (c) state that IRBs will include “at least one member whose primary concerns are in nonscientific areas” (US Department of Health and Human Services, 2019). Often called community members, these individuals play key roles as primary reviewers of research protocols to ensure absence of scientific bias and provide invaluable perspectives on vulnerable populations. If defined as intersectional coalitions, these new IRB groups can form the divergent identity groups necessary to advocate for varied groups who are disadvantaged by the assessment. As well, community members can help us understand that such categorical differences afford strength in measurement design and development, and thereby reject race hegemony and embrace individual difference.

## 7 Conclusion

Our article opened with facts about Black Lives Matter and this question: Are the foundations, operations, and applications in the *Standards for Educational and Psychological Testing* (AERA et al., 2014) sufficient to meet the current U.S. educational and psychological testing demands for justice and fairness for all populations? As we have argued, the *Standards* is necessary but insufficient. Racial justice extensions offer a set of general principles, intellectual traditions, and key concepts that can inform demands for justice and fairness. And, as we have acknowledged, racial justice extensions are not sufficient in and of themselves. Addressing measurement purism requires acknowledgment of social-historical conditions and commitment to ever-evolving process aimed at achieving – not merely pursuing – justice and fairness.

If we are to raise the kind of intentional questions as we have in this article – questions that can be used to identify existing structures of inequity within existing criteria and to modify them through racial justice principles – then we must be committed to challenging majoritarian mind-sets. CRT scholars have long used counterstories, parables, and anecdotes to challenge status quo narratives. CRT scholars such as Delgado (1989) have acknowledged the significance of counterstories – accounts that directly challenge assimilatory narratives, thereby rejecting institutional premises (p. 2430). The counterstory creates new narratives; “as if” reasoning allows those narratives to be viewed as plausible reasoning alternatives, acknowledging social-historical condition, to absolutist narratives and premises. Mislevy (2018) reminds us similarly about the ways measurement models work as majoritarian mind-sets and that the link between model and interpretation can be disrupted by recognizing the contingency of such thinking: “*we approximate individuals’ patterns of proficiency through the lens of the model, AS IF the proficiency variable were a property of an individual, in order to interpret the performance and to pass information on to assessment uses*” (italics and bold in original, p. 137).

Because they provide options for racial justice extensions, the creation of counterstories and deployment of “as if” reasoning both play an important role in racial justice. One need not ascribe, for example, to BFT in exclusion of anti-racist theory (Davis, 2018) or queer theory (Edelman, 2004). As our case studies demonstrate, one need only imagine the potential of racial justice principles to a specific context (the counterstory) and then apply (through “as if” reasoning) key concepts of these principles to the case at hand. If individuals are put at the center of testing and assessment efforts, the presence of counter-narrative and rules plausible reasoning become clear. That is, it becomes increasingly evident that racial justice extensions in educational measurement can be achieved only if the field acknowledges a much broader historical disciplinary identity and its roles within that history. We must be willing to accept that normative understandings may be, in fact, vehicles for structural racism. We must be willing to accept, in turn, that the accompanying logic may be an artifact of that very racism.

When we began drafting this article for the special issue in 2022, the *New York Times* assembled a list of mass shootings for that year. By the end of the year, there were 692 mass shootings. The losses include ten victims murdered on May 14 in a racially motivated shooting at the Jefferson Avenue Tops supermarket in Buffalo, New York, 19 children and two teachers killed by a gunman at Robb Elementary school in Uvalde, Texas, on May 24, and five murders following shooting at an L.G.B.T. Q. nightclub in Colorado Springs on Nov. 19. In the time that we were finalizing this article for the special issue, the *New York Times* assembled a list of mass shootings in 2023. As of mid-April 2023, there were 160 mass shootings so far this year. The losses include three children and three adults killed on March 27 by a heavily armed assailant in a Nashville school. The losses also include 11 killed in an Asian American community as they celebrated the Lunar New Year on January 21. The distance between the *Standards* and these shootings is, depending on your sense of justice, far away or right next door.

## Acknowledgments

The authors thank our reviewers for their careful comments on earlier versions of this manuscript and Susan M. Brookhart for her final review. We also would like to thank Emi Iwatani, Kyndra Middleton, and Jennifer Randall, the editors of this special issue, for their generosity, leadership, and commitment to justice.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## ORCID

Mya Poe  <http://orcid.org/0000-0002-3349-9093>

Maria Elena Oliveri  <http://orcid.org/0000-0002-7606-5324>

Norbert Elliot  <http://orcid.org/0000-0002-3933-2008>

## References

- Adams, G., & Salter, P. S. (2019). They (color) blinded me with science: Counteracting coloniality of knowledge in hegemonic psychology. In K. Crenshaw, L. C. Harris, D. M. HoSang, & G. Lipsitz (Eds.), *Seeing race again: Countering colorblindness across the disciplines* (pp. 271–292). University of California Press. doi:10.2307/j.ctvcwp0hd.18
- Alim, S.H., Rickford, J. R., & Ball, A. F. (Eds.). (2016). *Raciolinguistics: How language shapes our ideas about race*. UK: Oxford.
- American Educational Research Association (AERA), American Psychological Association (APA), and National Council on Measurement in Education (NCME). (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- American Psychological Association. Task Force on Employment Testing of Minority Groups, Washington, DC. (1969). Job testing and the disadvantaged. *The American Psychologist*, 24(7), 637–650. doi:10.1037/h0037806
- American Psychological Association, Committee on Test Standards. (1952). Technical recommendations for psychological tests and diagnostic techniques: Preliminary proposal. *The American Psychologist*, 7(8), 461–475. doi:10.1037/h0056631
- Arieli-Attali, M., Ward, S., Thomas, J., Deonovic, B., & von Davier, A. A. (2019). The expanded evidence-centered design (e-ECD) framework for learning and assessment systems: A framework for incorporating learning goals and processes within assessment design. *Frontiers in Psychology*, 10. doi:10.3389/fpsyg.2019.00853
- Assessment Technologies Institute (2022). *Next Generation NCLEX guidebook 2022*. <https://www.atitesting.com/docs/default-source/next-gen-nclex/2022-next-generation-nclex-guidebook9f6037dd80cb405a808bba44f60d9c52.pdf>
- Bailey, M. (2021). *Misogynoir transformed: Black women's digital resistance*. New York: New York University Press.
- Baker-Bell, A. (2020). *Linguistic justice: Black language, literacy, identity, and pedagogy*. New York: Routledge & National Council of Teachers of English. doi:10.4324/9781315147383
- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall.
- Banks, J., Onsomu, E. O., Hall, S., Hargis, J., Ketner, D. H., Tyson, T. . . McCullough, E. (2022). Factors that predict NCLEX-RN success at a Historically Black College and University: A four-year retrospective study. *Journal of Professional Nursing*, 39, 177–186. doi:10.1016/j.profnurs.2021.12.011
- Beerman, J. M. (2001). The unhappy history of Civil Rights Legislation, fifty years later. *Connecticut Law Review*, 34, 981–1036. [https://scholarship.law.bu.edu/faculty\\_scholarship/1997](https://scholarship.law.bu.edu/faculty_scholarship/1997)
- Bell, D. (1991). Foreword: The final civil rights act. *California Law Review*, 79(3), 597–611. doi:10.2307/3480830
- Bell, D. A. (1980). Brown v. board of education and the interest-convergence dilemma. *Harvard Law Review*, 93(3), 518–533. doi:10.2307/1340546
- Betts, J., Muntean, W., Kim, D., Jorion, N., & Dickison, P. (2019). Building a method for writing clinical judgment items for entry-level nursing exams. *Journal of Applied Testing Technology*, 20(S2), 21–36. <https://www.jattjournal.com/index.php/atp/article/view/143674>
- Bronfenbrenner, U., & Morris, P. A. (2006). The bioecological model of human development. In R. M. Lerner & W. Damon (Eds.), *Handbook of child psychology: Theoretical models of human development*. (6<sup>th</sup> edition). (pp. 793–828). Hoboken: John Wiley. doi:10.1002/9780470147658.chpsy0114
- California Community Colleges (2021a). *Standards for assessment test instrument review*. <https://www.cccco.edu/-/media/CCCO-Website/About-Us/Divisions/Educational-Services-and-Support/Academic-Affairs/Assessment-advisory-committee/ccassessmentstandards2017wupdatesa1ly.pdf>

- California Community Colleges (2021b). *Update to the Vision for Success: Reaffirming equity in a time of recovery*. <https://www.cccco.edu/-/media/CCCO-Website/Reports/vision-for-success-update-2021-a11y.pdf>
- California Education Code, §78220–78222 (2022). <https://law.justia.com/codes/california/2022/code-edc/title-3/division-7/part-48/chapter-2/article-1-5/>
- Camilli, G., Newton, P. (2022). Doing justice to fairness. In J. L. Jonson & K. F. Geisinger (Eds.), *Fairness in educational and psychological testing: Examining theoretical, research, practice, and policy implications of the 2014 Standards* (pp. 111–130). Washington, DC: American Educational Research Association. doi:10.2307/j.ctv2kzv0fw.9
- Carbado, D. W., & Harris, C. I. (2019). Intersectionality at 30: Mapping the margins of anti-essentialism, intersectionality, and dominance theory. *Harvard Law Review*, 132(8), 2193–2239. [https://harvardlawreview.org/wp-content/uploads/2019/06/2193-2239\\_Online.pdf](https://harvardlawreview.org/wp-content/uploads/2019/06/2193-2239_Online.pdf)
- Clearly, T. A. (1966, June). “Test bias: Validity of the scholastic aptitude test for negro and white students in integrated colleges. RB-66-51. *ETS Research Report*, 1966(2), i-23. doi:10.1002/j.2333-8504.1966.tb00529.x
- Clearly, T. A. (1968). Test bias: Prediction of negro and white students in integrated colleges. *Journal of Educational Measurement*, 5(2), 115–124. doi:10.1111/j.1745-3984.1968.tb00613.x
- Codiroli McMaster, N., & Cook, R. (2019). The contribution of intersectionality to quantitative research into educational inequalities. *Review of Education*, 7(2), 271–292. doi:10.1002/rev3.3116
- Collins, P. H. (1989). The social construction of Black feminist thought. *Signs: Journal of Women and Culture in Society*, 14(4), 745–773. doi:10.1086/494543
- Collins, P. H. (2000). *Black feminist thought: Knowledge, consciousness, and the politics of empowerment* (2nd ed.). New York and London: Routledge.
- Collins, P. H. (2019). *Intersectionality as critical social theory*. Durham: Duke University Press. doi:10.1215/9781478007098
- Combahee River Collective. (1977). The combahee river collective statement. <https://web.archive.org/web/20170224021117/http://circuitous.org/scraps/combahee.html>
- Crenshaw, K. (1989). Demarginalizing the intersection of race and sex: A Black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics. *University of Chicago Legal Forum*, 1, Article 8. <http://chicagounbound.uchicago.edu/uclf/vol1989/iss1/8>
- Crenshaw, K. (1991). Mapping the margins: Intersectionality, identity politics, and violence against women of color. *Stanford Law Review*, 43(6), 1241–1299. doi:10.2307/1229039
- Crenshaw, K., Gotanda, N., Peller, G. et al. (1995). *Critical race theory: The key writings that formed the movement*. New York: New Press.
- Currier, J. (2019). NGN Talks: Clinical judgment model video transcript. Retrieved April 20, 2023, from [https://www.ncsbn.org/public-files/videos/Transcript\\_NGNTalk\\_Episode06.pdf](https://www.ncsbn.org/public-files/videos/Transcript_NGNTalk_Episode06.pdf)
- Cushman, E. (2016). Decolonizing validity. *Journal of Writing Assessment*, 9(1). <https://escholarship.org/uc/item/0xh7v6fb>
- Davis, A. Y. (1981). *Women, race, and class*. New York: Random House.
- Davis, A. Y. (2018). The past, present, and future of Assata’s message. *WSQ: Women’s Studies Quarterly*, 46(3&4), 232–234. doi:10.1353/wsq.2018.0044
- Delgado, R. (1989). Storytelling for oppositionists and others: A plea for narrative. *Michigan Law Review*, 87(8), 2411–2441. doi:10.2307/1289308
- Delgado, R. (2009). Liberal McCarthyism and the origins of critical race theory. *Iowa Law Review*, 94, 1505–1545. <https://digitalcommons.law.seattleu.edu/faculty/153/>
- del Rosario Basterra, M., Trumbull, E., & Solano-Flores, G. (Eds.). (2011). *Cultural validity in assessment: Addressing linguistic and cultural diversity*. New York: Routledge.
- Dickison, P., Haerling, K. A., & Lasater, K. (2019). Integrating the National Council of State Boards of Nursing clinical judgment model into nursing educational frameworks. *The Journal of Nursing Education*, 58(2), 72–78. doi:10.3928/01484834-20190122-03
- Edelman, L. (2004). *No future: Queer theory and the death drive*. Durham: Duke University Press. doi:10.2307/j.ctv11hpkpp
- Elder, C., McNamara, T., & Congdon, P. (2003). Rasch techniques for detecting biasing performance assessments: An example comparing the performance of native and non-native speakers on a test of academic English. *Journal of Applied Measurement*, 4(2), 181–197.
- Ercikan, K., & Oliveri, M. E. (2013). Is fairness research doing justice? A modest proposal for an alternative validation approach in differential item functioning (DIF) investigations. In M. Chatterji (Ed.), *Validity and test use: An international dialogue on educational assessment, accountability and equity* (pp. 69–86). UK: Emerald Publishing.
- Fauset, J. R. (1924). *There is confusion*. New York: Boni & Liveright.
- Florida Center for Nursing. (2020). *2020 technical documentation for licensure and workforce analysis*. <https://floridasnursing.gov/forms/supply/2020%20Technical%20Documentation.pdf>

- Garcia, N. M., López, N., & Vélez, V. N. (2017). QuantCrit: Rectifying quantitative methods through CRT. *Race, Ethnicity, & Education*, 21(2), 149–157. doi:10.1080/13613324.2017.1377675
- Green, L. (2004). Research on African American English since 1998: Origins, description, theory, and practice. *Journal of English Linguistics*, 32(3), 210–229.
- Geisinger, K. F. (2022). The history of bias assessment in educational and psychological testing. In J. L. Jonson & K. F. Geisinger (Eds.), *Fairness in educational and psychological testing: Examining theoretical, research, practice, and policy implications of the 2014 Standards* (pp. 13–31). Washington, DC: American Educational Research Association. doi:10.2307/j.ctv2kzv0fw.5
- Graphenreed, T., & Poe, M. (2022). Antiracist genre systems: Creating non-violent writing classroom spaces. *Composition Studies*, 50(2), 53–76. <https://compositionstudiesjournal.files.wordpress.com/2022/11/graphenreed-poe.pdf>
- Gregersen, T., & MacIntyre, P. (2014). Capitalizing on language learner's individuality: From premise to practice. *Multilingual Matters/Channel View Publications*. doi:10.21832/9781783091218
- Hancock, A. M. (2007). When multiplication doesn't equal quick addition: Examining intersectionality as a research paradigm. *Perspectives on Politics*, 5(1), 63–79. doi:10.1017/S1537592707070065
- Hankerson, S. (2023). "The world has to stop discriminating against African American language" (AAL): Exploring the language ideologies of AAL-speaking students in college writing. *Written Communication*, 40(2), 587–619. <https://doi.org/10.1177/07410883221146484>
- Hazelton, L., Nastal, J., Elliot, N., Bursstein, J., & McCaffrey, D. F. (2021). Formative automated writing evaluation: A standpoint theory of action. *Journal of Response to Writing*, 7(1), 37–91. <https://scholarsarchive.byu.edu/journalrw/vol7/iss1/3>
- Hood, S., & Hopson, R. K. (2008). Evaluation roots reconsidered: Asa Hilliard, a fallen hero in the 'Nobody Knows My Name' Project, and African educational excellence. *Review of Educational Research*, 78(3), 410–426. doi:10.3102/0034654308321211
- Hood, S., Hopson, R., & Frierson, H. (2005). *The role of culture and cultural context: A mandate for inclusion, the discovery of truth and understanding in evaluative theory and practice*. Greenwich, CT: Information Age Publishing.
- hooks, b. (1992). *Black looks: Race and representation*. Boston: South End Press.
- Jang, S. T. (2018). The implications of intersectionality on Southeast Asian female students' educational outcomes in the United States: A critical quantitative intersectionality analysis. *American Educational Research Journal*, 55(6), 1268–1306. doi:10.3102/0002831218777225
- Jefferies, K., Goldberg, L., Aston, M., & Tomblin Murphy, G. (2018). Understanding the invisibility of black nurse leaders using a black feminist poststructuralist framework. *Journal of Clinical Nursing*, 27(15–16), 3225–3234. doi:10.1111/jocn.14505
- Johnson, O. C. A. (2016). The local turn: Innovation and diffusion in Civil Rights Law. *Law & Contemporary Problems*, 79, 115–144. <https://scholarship.law.duke.edu/lcp/vol79/iss3/6>
- Johnson, E. P., & Henderson, M. G. (2005). *Black queer studies: A critical anthology*. Durham & London: Duke.
- Kasten, K. (2022). Beloved community: Teacher reflection on Hawai'i-based teaching and learning framework, Nā Hopena A'o (HĀ), and possible implications for decolonisation in peace studies and peace education. In K. T. Maihāroa, M. Ligaliga, & H. Devere (Eds.), *Decolonising peace and conflict studies through indigenous research* (pp. 131–158). Singapore: Springer Nature Singapore. doi:10.1007/978-981-16-6779-4\_8
- Katz, I. R., & Elliot, N. (2016). Information literacy in digital environments: Construct mediation, construct modeling, and validation processes. In B. D'Angelo, S. Jamieson, B. Maid, & J. R. Walker (Eds.), *Information literacy: Research and collaboration across disciplines* (pp. 93–110). Fort Collins and Boulder: The WAC Clearinghouse; University Press of Colorado. doi:10.37514/PER-B.2016.0834.2.05
- Kendi, I. X. (2019). *How to be an antiracist*. New York: One World.
- Klitzman, R. (2012). Institutional review board community members: Who are they, what do they do, and whom do they represent? *Academic Medicine*, 87(7), 975–981. doi:10.1097/ACM.0b013e3182578b54
- Lyiscott, J. (2019). *Black appetite, white food: Issues of race, voice, and justice within and beyond the classroom*. New York and London: Routledge.
- Macey, A., O'Reilly, G., Williams, G., & Cameron, P. (2022). Critical care nursing role in low and lower middle-income settings: A scoping review. *British Medical Journal Open*, 12(1), e055585. doi:10.1136/bmjopen-2021-055585
- MacIntyre, A. (2007). *After virtue: A study in moral theory* (3rd ed.). Notre Dame: University of Notre Dame Press.
- Maul, A., Irribarra, D. T., & Wilson, M. (2016). On the philosophical foundations of psychological measurement. *Measurement*, 79, 311–320. doi:10.1016/j.measurement.2015.11.001
- Messick, S. (1980). Test validity and the ethics of assessment. *The American Psychologist*, 35(11), 1012–1027. doi:10.1037/0003-066X.35.11.1012
- Mills, C. W. (2017). *Black rights/white wrongs: The critique of racial liberalism*. New York: Oxford University Press. doi:10.1093/acprof:oso/9780190245412.001.0001

- Mislevy, R. J. (2018). *Sociocognitive foundations of educational measurement*. New York and London: Routledge. doi:10.4324/9781315871691
- Mislevy, R. J., Steinberg, L. S., & Almond, R. A. (2003). On the structure of educational assessments. *Measurement: Interdisciplinary Research & Perspectives*, 1(1), 3–67. doi:10.1207/S15366359MEA0101\_02
- Moll, L. C., Amanti, C., Neff, D., & Gonzalez, N. (1992). Funds of knowledge for teaching: Using a qualitative approach to connect homes and classrooms. *Theory into Practice*, 31(2), 132–141. doi:10.1080/00405849209543534
- National Council of State Boards of Nursing. (2022b, May). Report findings from the 2021 RN Nursing Knowledge Survey. *Research Brief*, 83. [https://www.ncsbn.org/public-files/21\\_RN\\_KSA\\_FINAL.pdf](https://www.ncsbn.org/public-files/21_RN_KSA_FINAL.pdf)
- National Council of State Boards of Nursing (NCSBN). (2022a). NCLEX pass rates. <https://www.ncsbn.org/16826.htm>
- Newton, P. E. (2017). There is more to educational measurement than measuring: The importance of embracing purpose pluralism. *Educational Measurement: Issues & Practice*, 36(2), 5–15. doi:10.1111/emip.12146
- Newton, P. E., & Shaw, S. D. (2014). *Validity in educational and psychological assessment*. Sage. doi:10.4135/9781446288856
- New York Times. (2022, Nov. 24). A partial list of mass shootings in the United States in 2022. *The New York Times*. <https://www.nytimes.com/article/mass-shootings-2022.html>
- New York Times. (2023, April 17). A partial list of mass shootings in the United States in 2022. *The New York Times*. <https://www.nytimes.com/article/mass-shootings-2023.html>
- Nunn-Ellison, K., Ard, N., Farmer, S., & Beasley, S. F. (2020). Next Generation NCLEX®: Using the ACEN standards and criteria to prepare. *Teaching and Learning in Nursing*, 15(2), A5–9. doi:10.1016/j.teln.2019.12.008
- Nussbaum, M. C. (2011a). *Creating capabilities: The human development approach*. Cambridge: Harvard University Press.
- Nussbaum, M. C. (2011b). Foreword. In I. M. Young (Eds.), *Responsibility for justice*. New York: Oxford University Press. ix. ixxxv.
- Oliveri, M. E., Poe, M., & Elliot, N. (2023). Fairness. In R. Tierney, F. Rizvi, & K. Ercikan (Eds.), *International encyclopedia of education Quantitative research and educational measurement* Vol. 14 (4th ed). pp. 45–50. Elsevier. 10.1016/B978-0-12-818630-5.10006-5
- O\*NET OnLine. (2023). Registered nurses, 29-1141.00. Retrieved April 20, 2023, from <https://www.onetonline.org/link/summary/29-1141.00>
- Perryman-Clark, S. M. (2013). African American language, rhetoric, and students' writing: New directions for SRTOL. *College Composition and Communication* 64(3), 469–495.
- Petersen, N. S., & Novick, M. R. (1976). An evaluation of some models for culture-fair selection. *Journal of Educational Measurement*, 13(1), 3–29. <https://doi.org/10.1111/j.1745-3984.1976.tb00178.x>
- Pfeifer, C. M., & Sedlacek, W. E. (1971). The validity of academic predictors for black and white students at a predominantly white university. *Journal of Educational Measurement*, 8(4), 253–261. doi:10.1111/j.1745-3984.1971.tb00934.x
- Poe, M., & Cogan, J. A. (2016). Civil rights and writing assessment: Using the disparate impact approach as a fairness methodology to evaluate social impact. *Journal of Writing Assessment*, 9. <https://escholarship.org/uc/item/08flc307>
- Prior, P. (2006). A sociocultural Theory of writing. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 54–66). New York: Guilford Press.
- Randall, J. (2021). Color-neutral is not a thing: Redefining construct definition and representation through a justice-oriented critical antiracist lens. *Educational Measurement: Issues & Practice*, 40(4), 82–90. doi:10.1111/emip.12429
- Randall, J., Poe, M., & Slomp, D. (2021). Ain't Oughta Be in the dictionary: Getting to justice by dismantling anti-black literacy assessment practices. *Journal of Adolescent & Adult Literacy*, 64(5), 594–599. doi:10.1002/jaal.1142
- Randall, J., Slomp, D., Poe, M., & Oliveri, M. E. (2022). Disrupting white supremacy in assessment: Toward a justice-oriented, antiracist validity framework. *Educational Assessment*, 27(2), 170–178. doi:10.1080/10627197.2022.2042682
- Rawls, J. (1999). *A theory of justice* (Rev. ed.). New York: Cambridge University Press. doi:10.4159/9780674042582
- Rickford, J.R., Sweetland, J. & Rickford, A. E. (2013). *African American, Creole, and other Vernacular Englishes in education: A bibliographic resource*. New York and London: Routledge.
- Robbins, P., & Aydede, M. (2009). A short primer on situated cognition. In P. Robbins & M. Aydede (Eds.), *The Cambridge handbook of situated cognition* (pp. 3–10). New York: Cambridge University Press.
- Roberts, D., & Jesudason, S. (2013). Movement intersectionality: The case of race, gender, disability, and genetic technologies. *Du Bois Review*, 10(2), 313–328. doi:10.1017/S1742058X13000210
- Russell, M., Szendey, O., & Kaplan, L. (2021). An intersectional approach to DIF: Do initial findings hold across tests? *Educational Assessment*, 26(4), 284–298. doi:10.1080/10627197.2021.1965473
- Sato, E. (2017). Culture in fair assessment practices. In H. Jiao & R. W. Lissitz (Eds.), *Test fairness in the new generation of large-scale assessments* (pp. 147–160). Charlotte, NC: Information Age Publishing.
- Sibbett, L. (2020). Intersectionality in U.S. educational research. In G. W. Noblit (Ed.), *Oxford research encyclopedia of education* (pp. 1–42). <https://doi.org/10.1093/acrefore/9780190264093.013.403>

- Slomp, D. H. (2016). An integrated design and appraisal framework for ethical writing assessment. *Journal of Writing Assessment*, 9. <https://escholarship.org/uc/item/4bg9003k>
- Solano-Flores, G., & Nelson-Barber, S. (2001). On the cultural validity of science assessments. *Journal of Research in Science Teaching*, 38(5), 553–573. doi:10.1002/tea.1018
- Spivak, G. C. (1988). Can the subaltern speak? In C. Nelson & L. Grossberg (Eds.), *Marxism and the interpretation of culture* (pp. 271–313). Urbana: University of Illinois Press.
- Stanley, J. C., & Porter, A. C. (1967). Correlation of scholastic aptitude test scores with college grades for negroes versus whites. *Journal of Educational Measurement*, 4(4), 199–218. doi:10.1111/j.1745-3984.1967.tb00589.x
- Stavrova, O., & Schlösser, T. (2015). Solidarity and social justice: Effect of individual differences in justice sensitivity on solidarity behaviour. *European Journal of Personality*, 29(1), 2–16. doi:10.1002/per.1981
- Taylor, U. (1998). The historical evolution of Black feminist theory and praxis. *Journal of Black Studies*, 29(2), 234–253. doi:10.1177/002193479802900206
- Teo, T. (2010). What is epistemological violence in the empirical social sciences? *Social and Personality Psychology Compass*, 4(5), 295–303. doi:10.1111/j.1751-9004.2010.00265.x
- Tietje, L., & Cresap, S. (2005). Is lookism unjust? The ethics and aesthetics and public policy implications. *The Journal of Libertarian Studies*, 19, 31–50.
- US Department of Health and Human Services (2019). 2018 Requirements (2018 Common Rule), \$46.107 IRB membership. <https://www.hhs.gov/ohrp/regulations-and-policy/regulations/45-cfr-46/revised-common-rule-regulatory-text/index.html#46.107>
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge: Harvard University Press.
- Waapalaneekweew. (Nicole R. Bowman-Farrell, M). (2018). Looking backward but moving forward: Honoring the sacred and asserting the sovereign in indigenous evaluation. *American Journal of Evaluation*, 39(4), 1043–1046. doi:10.1177/1098214018790412
- Yosso, T. J. (2005). Whose culture has capital? A critical race theory discussion of community cultural wealth. *Race, Ethnicity & Education*, 8(1), 69–91. doi:10.1080/1361332052000341006
- Young, I. M. (2011). *Responsibility for justice*. Oxford and New York: Oxford University Press. doi:10.1093/acprof:oso/9780195392388.001.0001
- Zumbo, B. D., Liu, Y., Wu, A. D., Shear, B. R., Olvera Astivia, O. L., & Ark, T. K. (2015). A methodology for Zumbo's third generation DIF analyses and the ecology of item responding. *Language Assessment Quarterly*, 12(1), 136–151. doi:10.1080/15434303.2014.972559
- Zwick, R., Dorans, N. J., & Cook, L. (2016). Philosophical perspectives on fairness in educational assessment. In N. J. Dorans (Ed.), *Fairness in educational assessment and measurement* (pp. 267–281). Routledge. doi:10.4324/9781315774527-17