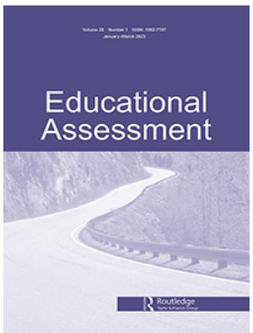

Routledge
Taylor & Francis Group

Educational Assessment

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/heda20

It Ain't Near 'Bout Fair: Re-Envisioning the Bias and Sensitivity Review Process from a Justice-Oriented Antiracist Perspective

Jennifer Randall

To cite this article: Jennifer Randall (2023): It Ain't Near 'Bout Fair: Re-Envisioning the Bias and Sensitivity Review Process from a Justice-Oriented Antiracist Perspective, Educational Assessment, DOI: 10.1080/10627197.2023.2223924

To link to this article: https://doi.org/10.1080/10627197.2023.2223924

Published online: 12 Jun 2023.

Submit your article to this journal

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=heda20

EDUCATIONAL ASSESSMENT
https://doi.org/10.1080/10627197.2023.2223924

Routledge
Taylor & Francis Group

Check for updates

It Ain't Near 'Bout Fair: Re-Envisioning the Bias and Sensitivity Review Process from a Justice-Oriented Antiracist Perspective

Jennifer Randall

University of Michigan, Ann Arbor, MI, USA

ABSTRACT

In a justice-oriented antiracist assessment process, attention to the disruption of white supremacy must occur at every stage – from construct articulation to score reporting. An important step in the assessment development process is the item review stage often referred to as Bias/Fairness and Sensitivity Review. I argue that typical approaches to the item and test review process miss the opportunity to actively disrupt white supremacist and racist logics – in other words, to be anti-racist. Using Critical Race and Critical Whiteness Theory as a frame, this paper challenges the field to re-envision the purpose and outcomes of the bias and sensitivity review process by (a) identifying common themes and/or recommendations found in bias and sensitivity guidelines that, even if unintentionally, center whiteness and/or the paradigm of white dominant culture; and (b) recommending a set of bias and sensitivity principles that promote an antiracist approach to assessment design, specifically item review.

Introduction

> Dominator culture has tried to keep us all afraid, to make us choose safety instead of risk, sameness instead of diversity. Hooks (2003)

The sensitivity review is a critical step in the assessment development process and most large-scale assessment systems engage in some form of sensitivity, bias, and/or fairness review (e.g., the SAT, ACT, GMAT, NAEP). These reviews have several interrelated purposes including making sure the assessment (1) reflects the cultural backgrounds of all test takers (Educational Testing Service, 2002), (2) does not include any offensive or anger-inducing content (Educational Testing Service, 2002; Smarter Balanced Assessment Consortium, 2021), (3) is free of any unnecessary barriers, and (4) does not advantage or disadvantage any test taker through presentation or content (Thompson, Johnstone, & Thurlow, 2002). For example, Smarter Balanced notes that the purpose of their *Sensitivity and Bias Guidelines* is "to support the process of developing and reviewing Smarter Balanced assessments that are fair for all groups of test takers, despite differences in characteristics including, but not limited to, disability status, ethnic group, gender, regional background, native language, race, religion, sexual orientation, and socioeconomic status" (p. 6). ETS' *Guidelines for Fair Tests and Communications* (2016) are designed to help assessment developers "obtain a better understanding of fairness, take fairness into account as materials are designed, avoid the inclusion of unfair content as materials are developed, find and eliminate any unfair content as materials are reviewed, represent diverse people in materials, and reduce subjective differences in decisions about fairness" (p. 3). Ultimately, the sensitivity review process is designed not only to address issues of fairness, but also improve the

CONTACT Jennifer Randall jennrand@umich.edu University of Michigan, 610 E. University Avenue, Ann Arbor, MI 48109, USA

© 2023 Taylor & Francis Group, LLC

psychometric quality of the assessment and, consequently, its legal defensibility (Golubovich, Grand, Ryan, Schmitt, & Schmitt, 2014).

I argue that because (1) racist logics provide the historical framing for all of our assessment and measurement systems and (2) the compulsion to elevate and protect whiteness has led to the dehumanization and/or erasure of marginalized identities and experiences on assessments, an explicit commitment to antiracist[1] processes and guidelines must be made to rupture these patterns. Important to this rupturing process is a reframing of our policies and practice in assessment as they relate to the bias and sensitivity review process. To that end, the primary purpose of this paper is to acknowledge (and provide recommendations/principles that reconcile) the tensions between a justice-oriented antiracist approach to assessment and typical bias, sensitivity, and fairness guidelines employed by the field of educational measurement as part of the assessment development process. For the sake of simplicity, I limit the treatment of the topic to educational (diagnostic, formative & summative) assessments that target P–12 learners but maintain the principles presented apply across all assessment sectors.

I ask that the reader consider these recommendations while holding five assertions in mind. First, these recommendations should be considered in the context of all large-scale assessments (to mean any assessments not created by teachers) including those assessment systems tied to specific curriculum or used for diagnostic or formative purposes including through-course assessments. Second, principles and guidelines – even when intended to inform large-scale assessment development – inform the design of teacher-created assessments as well. Third, although this reconceptualization of the bias and sensitivity process would be most impactful in conjunction with instructional experiences that are also antiracist, it is not a requirement. Fourth, these recommendations should not be considered exhaustive or almighty, but rather provisional[2] concepts as a starting point for thoughtful discussion and solution-seeking. Finally, although the focus of this manuscript is the bias and sensitivity review process, the underlying principles can, and should, be applied across the entire assessment development process. As Audrey Qualls (1998) noted – in referring to culturally responsive assessment – this work requires collaboration across all stages of development.

I begin with a description of the critical theories that provide the framework for both the critique and recommended principles – Critical Race Theory and Critical Whiteness Theory. Then, I provide the reader with a set of general principles/recommendations that support a justice-oriented, antiracist perspective providing examples of specific modified guidelines to that end.

Critical race theory & antiracist assessment

Discussions of race, racism, and/or antiracism are quite appropriately situated within the framework of Critical Race Theory (Bell, 1995; Delgado & Stefancic, 2017) and its history. Critical Race Theory (CRT), which was originally articulated in the legal studies, has been found to be applicable in other fields including the social sciences, specifically education. CRT rests on several tenets that are used to explain the role and impact of race in everyday society. First, race is socially constructed and that there is no biological basis by which it exists. Second, this socially constructed notion of race has a history of being shifted and differentially applied based on the needs of the dominant culture (i.e., whiteness). Minoritized racial/ethnic groups can evolve from trusted agricultural workers to dangerous security threats bound for internment camps (replaced by other minoritized workers in the fields) to an

[1]Here, I rely on Kendi's (2019) conceptualization of anti-racism. In this case referring to the use of language that produces or sustains racial equity between racial groups.

[2]In a footnote, Crenshaw (1991) described intersectionality as a provisional concept linking contemporary politics with postmodern theory. She described her work as challenging dominant assumptions about race and gender as essentially separate categories. Carastathis (2016) describes Crenshaw's provisional concept in this way: "The notion of a provisional concept reflects the intuition that in order to transform our thinking let alone institutionalized practices, our current axiomatic assumptions, cognitive habits, and unreflective premises have to be at once engaged and disrupted" (108). This work similarly seeks to challenge and transform our thinking about the assessment development process.

essential tech labor force in less than two generations. A third and important tenet asserts that racism, despite widespread claims to the contrary, is not unusual or aberrant; but rather typical and sewn into the very fabric of our everyday lives. Any claims, processes, and/or guidelines that operate under the assumption of a racism-free or racism-lite system are inherently flawed. As Toni Morrison (1992) wrote, "... racism is as healthy today as it was during the Enlightenment. It seems that it has a utility far beyond economy, beyond the sequestering of classes from one another, and has assumed a metaphorical life so completely embedded in daily discourse that it is perhaps more necessary and more on display than ever before" (p. 63).

The fourth tenet explains how the white dominant class has failed to enact justice-based processes, laws, guidelines without a clear understanding of the benefits such a practice would offer them (i.e., interest convergence, Bell, 1980). Finally, the importance of storytelling (narratives/counternarratives) is emphasized in CRT. Delgado (1995) writes that storytelling is used to "analyze myths, presuppositions, and received wisdoms that make up the common culture about race and that invariability renders blacks and other minorities one down" (p.xiv). To be sure, the ways in which the ideas and experiences of racially minoritized (especially Black) persons have historically been perverted or ignored must be acknowledged. Mills (2007) wrote: "What people of color quickly come to see – in a sense, the primary epistemic principle of the racialized social epistemology of which they are the object – is that they are not seen at all" (p.18). Given this history, the narratives/experiences of these marginalized voices must be centered in order to provide the necessary and critical context that can go, and has gone, ignored within spaces claiming objectivity, or color neutrality, is the goal or even possible.

Ladson-Billings (1998) described CRT as an important intellectual and social tool for deconstruction, reconstruction, and construction: "deconstruction of oppressive structures and discourses, reconstruction of human agency, and construction of equitable and socially just relations of power" (p.9). To those ends, CRT provides a useful framework to guide and inform antiracist approaches to assessment. An antiracist[3] approach requires an explicit confrontation of racism in our assessment development processes and the assessments themselves actively seeking to reveal and disrupt these systems of oppression (Randall, 2021). Antiracist content on an assessment seeks to (a) explicitly disrupt conventional racist stereotypes, as they relate to any marginalized group; (b) reveal oppressive sociopolitical inequities and injustices, while simultaneously empowering students to enact change (e.g., protesting differential disciplinary and/or dress code policies in schools); (c) provide complete historical and contemporary perspectives that go beyond celebrating and/or protecting whiteness; and/or (d) allow for multiple ways of knowing/understanding and performing the content that extend beyond white-centric values (Randall, 2021). Through this process, antiracist assessment, itself, becomes assessment *for* learning and, if necessary, unlearning. It is unapologetically political, seeking at all times to rupture and "reconstruct hierarchical racial power arrangements that have been historically (re)produced by assessments" (Randall, 2021, p. 1). To be clear, CRT and antiracist framings serve to eliminate racialized oppression as part of a broader, more comprehensive goal of eliminating all forms of oppression.

Critical whiteness theory/whiteness studies & protecting whiteness

Related to CRT, critical whiteness, or whiteness studies, provides a frame for understanding the pervasiveness of white supremacy in all of our everyday lives. It interrogates the [social] construct of whiteness and how it ("it" being whiteness) simultaneously operates as both the absence of race (viewed as neutral or objective) and the dominant social (i.e., racial) identity. Whiteness is "perpetuated and maintained through networks and relations of power and privilege within and across societies" including educational spaces and contexts (Styres, 2019, p. 31). As McIntyre

[3]Kendi (2019) has defined antiracist policies (to include procedures, processes, regulations, & guidelines) as any measure that produces or sustains racial equity between racial groups.

(1997) notes all white people, even if not equally, benefit from whiteness and the ways in which whiteness enables racism. Moreover, whiteness is protected through the ways in which it allows for the discussion of race or mandates avoiding its reference/existence completely (Bonilla-Silva, 2013). McIntyre (1997) refers to this compulsion to avoid naming race as "white talk" which "serves to insulate white people from examining their/our individual and collective roles in the perpetuation of racism" (p.46). Examples of white talk, as described by Rogers and Mosley (2006), include "using humor to avoid difficult conversations, using the passive voice to remove responsibility from a person or group, using distancing pronouns and silences, and changing the topic" (p. 467).

Critical Whiteness Theory (Frankenberg, 1993) maintains that whiteness – which represents a set of dimensions including racial advantage, egocentrism, and obliviousness to whiteness as a race – impacts and shapes the lives of both White people and people of color. One cannot assume that a person of color (e.g., Black or Hispanic) has not been impacted by or convinced of notions of white supremacy and/or whiteness as the norm (thereby marking all else as abnormal/aberrant in need of assimilation and/or transformation). Indeed, just as Critical Race Theory purports that racism is typical and unavoidable, so does Critical Whiteness Theory (Frankenberg, 1993) with respect to white supremacy. As Zuberi and Bonilla-Silva (2008) write "White logic and White methods can be-and have been – used by members of all racialized groups and the critique (and defense) of them comes from all quarters" (p. 18).

Researcher positionality statement

The concept of researcher positionality generally refers to how a researcher positions themselves in relation to their inquiry. Within the fields of educational research and practice, positionality encompasses how a researcher's perceptions, knowledge-making, and the position they adopt might influence their inquiry. Alcoff (1988) asserts that researcher positionality illuminates the researcher–subject relationship by interrogating the relational and power dynamics that might exist between the researcher and the participants, context of the inquiry, and the processes involved in framing the inquiry. Knowledge-making, including the development, review, and implementation of assessment standards and guidelines, are subjective and are often shaped by assessment developers' and reviewers' onto-epistemological assumptions. Arguments arising from Feminist Thought, Decolonial and Anti-Colonial Scholarship, Black Studies and Anti-Racist Frameworks (Dillard, 2000; Foote & Bartell, 2011; Holmes, 2020; Hooks, 1984) suggest that a researcher brings with them their subjective self to their inquiry. Subjective researcher positionality influences research design process, analysis, and interpretation of data. Therefore, to engage in any form of inquiry, researchers must first and foremost *research* themselves (Dillard, 2000). That is, conducting an introspection of the self to unpack the biases likely to creep into the inquiry. I maintain that this process of introspection is particularly critical when discussing issues related to race and injustice. To that end, the author is prompted to conduct an introspection of herself in relation to her analysis of the content of bias and sensitivity guidelines.

I am a black woman, measurement specialist, former classroom teacher, and current professor. I recognize that my prior experiences with small-scale classroom and large-scale accountability, licensure, and selection assessments (as both administrator and test-taker) bound my perceptions about the utility and impact of these assessments. These experiences, as well as my belief in the underlying principles of Critical Race Theory, are reflected in my analyses of the bias/sensitivity guidelines. Moreover, my work and analyses are inextricably bound by the sociopolitical context in which I must conduct it. The rise in anti-Black racism realized through, for example, police brutality/murder, school disciplinary policies (so-called zero tolerance), and the seemingly increasing instances of the weaponization of white tears (see Hamad, 2020), inform and influence my position/engagement with these topics. Indeed, it was my dismay with the current sociopolitical context that drove me to this work. I rely on and

build upon the previous labor of other Black scholars who – over twenty years ago – posited the need for a shift in our assessment practices to be more culturally responsive (e.g., Lee, 1998; Qualls, 1998).

Guiding principles/recommendations

Within an CRT and CWT context, here I recommend a series of general principles for immediate changes/revisions to sensitivity guidelines for assessment developers to consider that reflect a justice-orientation. Rawls (1999) argued that equality of opportunity should be "to the greatest benefit of the least advantaged members of society" (p.35). He writes, with respect to justice:

> Laws and institutions no matter how efficient and well-arranged must be reformed or abolished if they are unjust . . . justice denies that loss of freedom for some is made right by a greater good shared by others. It does not allow that the sacrifices imposed on a few are outweighed by the larger sum of advantages enjoyed by many. (Rawls, 1999, p. 3)

Mills (1999, 2013) critiqued and extended Rawls' framework to include racial justice specifically (defined by Mills as corrective measures to rectify injustices that have already occurred). Mills points out a flaw in Rawls' assumption of an ideal well-ordered society (where his principles of distributive justice reside) in that it fails to provide a pathway to transition our current ill-ordered [white supremacist] society into a well-ordered one. And with this omission, one cannot account for the need for rectificatory justice, the correction of the past. In other words, Mills argued that one cannot merely focus on the conditions of an ideal world; but must also acknowledge the long history of oppressive practices/injustices that has led to the existence of the current ill-ordered world.[4] To limit one's focus ignores the injustices of the current society treating these injustices as incidental rather than structural.

I argue that a justice-orientation within the test-development process broadly, and the bias and sensitivity review process specifically, will require a willingness to reconsider our well-arranged processes and guidelines designed to protect the many (and typically the most powerful) and embrace processes that center the most marginalized populations (and the long history of that marginalization) in our decision-making. With Rawls (1999) conceptualization, and Mills (2013)much needed extension, of justice in mind, I recommend (1) a shift from a fear-oriented to a justice-oriented perspective in the development of guidelines; (2) a re-envisioning of what is meant by barriers and construct irrelevant variance; and finally (3) the need to facilitate the development of the collective critical consciousness of assessment developers and reviewers.

Shift from fear-orientation to justice-orientation

First, the entire assessment design process – but particularly bias and sensitivity review process- must shift from a fear-orientation to a justice orientation. Even a cursory review of most sensitivity and bias guidelines suggests that these guidelines are written – nearly exclusively – from a place of *fear* and not from a place of *justice*. Without question, assessment developers must satisfy multiple stakeholders (e.g., local education agencies, state legislators) throughout the assessment development process. To be sure, any assessment that does not meet the approval of its most powerful stakeholders is subject to replacement by an assessment that does. I argue that this reality has led to an assessment process – particularly with respect to bias and sensitivity guidelines – that prioritizes not offending or upsetting stakeholders for fear of losing contracts due to negative reactions (e.g., making the news, or being the subject of a viral blog post). It is this concern of offending, encountering criticism/pushback, and potentially losing contracts that position test developers to operate from a position of fear. This orientation (toward undoubtedly safe and unremarkable content), however, is often in conflict with

[4]I have argued elsewhere (Randall, 2022) that equity-based approaches to fairness fail to take into account the long history of (and current) oppressive, unjust barriers/practices; whereas justice-oriented approaches to fairness seek to make amends/reparations for those barriers ever having existed.

a justice-oriented assessment process. Not only does this fear-based approach to fairness result in the development of assessments that have been described as irrelevant, de-motivating, and/or anxiety-provoking by many students (Amrein & Berliner, 2003; Au, 2011; Rhone, 2006), such fearfulness limits the potential of what assessments can do *for* learning and in pursuit of social change. Elliot (2016) has defined fairness in assessment as "the identification of opportunity structures created through maximum construct representation under conditions of constraint – and the toleration of constraint only to the extent to which benefits are realized for the least advantaged – expressed in terms of its tradition, boundary, order and foundation" (p. 1). Currently, however, fairness in most guidelines is defined in terms of offending the fewest number of people – if any – at the cost of engagement, comprehensive truthfulness, and equity. In the text below, I provide some examples of this fear-oriented approach and suggest modifications that move us closer to a justice-oriented approach.

Experimentation on People/Animals. Smarter Balanced, for example, notes that experimentation on people or animals that is dangerous or painful as an example of an upsetting topic to be avoided. Indeed, *experimentation on people or animals that is dangerous or painful* may initially – through an uncritical lens – seem inappropriate for a large-scale assessment. A critical lens, however, recognizes that the inability to address this content prevents the comprehensive inclusion, or mention, of many critically important world events that shape current policies and practices. For example, the Tuskegee and/or Guatemalan syphilis experiments can be directly linked to current practices requiring experimental review boards and it should be wholly appropriate to include these events on a secondary social studies or science exam. Similarly, J. Marion Sims' painful gynecological experiments on Black slaves provides the historical basis for many false/racist beliefs held by medical professionals about Black patients today (e.g., Black people feel less physical pain). Furthermore, a comprehensive and accurate depiction of the Holocaust would be impossible in any context without a discussion of internment camp experiments. Consequently, assessments should be free to include corresponding content. I propose that instead of a sweeping prohibition of the inclusion of this type of content, bias and sensitivity standards should *caution against the gratuitous use of content that focuses on experimentation and focus on deliberate and intentional references to animal and human experimentation that serve a larger commitment to name injustice and the current consequences – both good and bad – of those injustices*.

Pregnancy. The pregnancy of human beings is also included as a topic to be avoided (Smarter Balanced Assessment Consortium, 2021). This type of erasure feeds into the false narrative that pregnancy is a disease and that pregnant women should isolate themselves from society until the disease has passed. Adding pregnancy to the list of unallowable topics means that students will never see representations of pregnant women running a business meeting, standing in line in a coffee shop, speaking with friends in a park (i.e., doing the things that humans do) on their assessments. Given 86% of the women between the ages 40 and 44 have given birth (Livingston, 2018), I find it difficult to excuse such a blatant and intentional exclusion of an entire class of humans as if their behavior (i.e., being pregnant) is inappropriate, hurtful, shameful, or wrong. A more appropriate justice-oriented approach would *(a) remove pregnant women from the list of topics to avoid completely and (b) encourage assessment developers to include representations of pregnant women engaged in a variety of activities to the extent to which they are represented (across all racial and ethnic groups) in the communities the assessments are intended to serve.*

Genocide. Genocide has also been identified as an emotionally charged subject that should be avoided or included only when directly related to the curriculum (see Educational Testing Service, 2016, p. 22). One could easily argue that it is simply impossible to address accurately any content standard related to the history of indigenous peoples in the world without explicitly mentioning genocide. Moreover, this history of the world – which includes colonialism and imperialism – includes meaningful, and most importantly, relevant examples of genocide. Yet, within our social institutions including the school system, curriculum offerings, and content standards, the orientation of the indigenous people to the world continues to be redefined and excluded (Smith, 1999). Thus, avoiding historical events such as genocide within the context of an assessment perpetuates another false

narrative of the world's history and present which is, at its core, white supremacist and relies on the devaluation of indigenous people. What is more, the impact of genocide does not simply reside in history curricula, and this guideline – seemingly – discourages item writers to go beyond the most strict and obvious mentions of the content. Antiracist frameworks acknowledge that the historical impact of genocide of North American indigenous peoples reverberates into multiple content areas and contexts and should be represented in assessment content accordingly. For example, the topic of genocide would be appropriate in the context of any human geography assessment (e.g., the forced migrations of people) as well as secondary math assessments (e.g., calculating the negative economic impact/wealth loss due to indigenous land seizures by the government from 1865 to 1965 based on current market values) and/or science assessments (e.g., the impact of land seizures and development on habitats). Indeed, because the white dominant culture may be uncomfortable acknowledging the impact of white supremacist practices beyond the narrowly defined time frame in which the initial oppressive practice was sanctioned, sensitivity guidelines encourage test developers to avoid this acknowledgment. A justice-oriented guideline would not only *remove the restrictive boundaries excluding content related to genocide, but also encourage test developers to include content that acknowledges the long-term consequences and interconnectedness of genocidal practices/policies (both domestic and abroad) on multiple sectors*. In other words, topics such as genocide could serve as opportunities to use assessment *for* learning and not simply of learning.

Slavery. Similarly, to the topic of genocide, American chattel slavery greatly defined hundreds of years of the United States' history. Still, bias and sensitivity guidelines typically restrict the discussion of this topic to specific contexts (see Data Recognition Corporation, 2003; Educational Testing Service, 2016; Smarter Balanced Assessment Consortium, 2021). For example, the Florida Department of Education (2012) standards note that "This topic may be included in historical or literary documents. A focus on graphic, upsetting aspects of enslaving people should be avoided" (p.17). To limit the discussion of chattel slavery to American history and literature assessments, while also simultaneously warning assessment developers to avoid any "upsetting aspects of enslaving people" is an egregiously white supremacist practice. An antiracist approach to assessment requires developers to provide complete and accurate historical perspectives that go beyond celebrating and/or protecting whiteness. Any treatment of the narrative(s) of enslaved peoples should represent the truth of those people (and note that white people should never be the arbitrators of that particular truth). Indeed, to apply this limitation under the guise of protecting children from upset is a blatant manifestation of white supremacy. The legacy of the complete history of enslavement permeates every aspect of American society and should be addressed – and not ignored – in both curriculum and assessments across multiple content areas (e.g., history, music, literature, science, and mathematics). Instead of limiting the discussion of American slavery to specific contexts, I recommend these guidelines *encourage the accurate depiction of slave conditions, honor the dignity and humanity of enslaved persons, and demand/require the critical interrogation of the legacy of slavery within current contexts.*

Historical and Contemporary Biographies. Bias and sensitivity guidelines routinely caution against the use of (a) historical biographical passages that acknowledge known prior bad behavior and (b) contemporary biographical passages for fear that bad behavior may be revealed. In fact, the Smarter Balanced guidelines note "Narratives related to historical figures that explicitly or implicitly point to those figures' involvement in negative contexts such as criminal activity or racism, for example, should be treated with care. This is also applicable to historical events and places and warrants additional attention to the extent to which those events or places detail controversial content and appropriateness of the content in general" (p.18). Similarly, Educational Testing Service (2016) writes, "It is generally best to avoid passages that focus on individuals who are readily associated with offensive or controversial topics, unless important for valid measurement. It is prudent to avoid biographical passages that focus on celebrities who are still living; their future actions are unpredictable and may result in fairness problems." (p.21). Moreover, Data Recognition Corporation (2003) lists "biographies of controversial figures whether or not they are still alive" (p.12) as a topic to be generally avoided. These types of standards unapologetically demand that whiteness be protected. To

encourage assessment developers to avoid the narratives of these figures that include reference to racist or criminal behavior is simply another way to say that the white supremacist hegemony must be maintained through/within our assessments. Such standards do not even allow for the accurate and comprehensive presentation of the world's history; and, instead, encourage a "white-washing" to protect and preserve whiteness. Moreover, the inability to share the narratives and counternarratives of contemporary marginalized persons (a critical tenet of CRT) makes it considerably more difficult to disrupt widely held racist logics about these communities and their experiences.

Adichie (2009) warned about the danger of a single story: . . . "that is how to create a single story, show a people as one thing, as only one thing, over and over again, and that is what they become." No history/person is all positive or all negative, yet assessments routinely identify white historical figures as near-flawless heroes (e.g., Thomas Jefferson, George Washington), indigenous persons as savages (e.g., narratives of scalping), and enslaved Africans as weak, powerless, and oppressed. Bias and sensitivity guidelines must shift away from encouraging the dangerous single story through assessments and instead *encourage assessment developers to include a wide range of biographies (contemporary and historical) that provide comprehensive and truthful stories and perspectives*.

Racial Justice and Social Problems. Both ETS and Smarter Balanced caution assessment developers/item writers about the inclusion of items that address racial injustice or social problems. For example, Smarter Balanced writes "content related to racial injustice, pandemics, or natural disasters needs to undergo thorough reviews to ensure that those topics do not provoke any feelings that have the potential to traumatize students" (Smarter Balanced Assessment Consortium, 2021, p. 12). Although the apparent intent of such guidelines is to protect students from experiencing trauma, I maintain there are several issues with this guideline and similar guidelines. First, the very notion that an item/task addressing racial violence would be traumatizing for students should actually be interrogated. Assessment developers caution that these types of items should go through thorough review, but then fail to engage in an active research and development process to confirm/disconfirm their fears. Instead, item writers are directed to simply avoid the content and make assumptions about students based on the fears of assessment developers and not any actual empirical evidence. In fact, I posit that the trauma was in the event; and that further trauma is imposed through the erasure and dehumanization of ignoring the event. Second, I argue that the expression *potential to traumatize students* is coded language for *potential to upset parents and other adult stakeholders*. Although there is a dearth of evidence suggesting that items addressing social injustices are traumatizing to students, the assessment community has considerable anecdotal evidence that some parents/adults are opposed to engaging students in any critical thinking around issues of justice (e.g., racial, economic). In fact, one set of guidelines states explicitly "It is important to avoid including items that would be deemed inappropriate by parents and *other citizens*" (emphasis added; Florida Department of Education, 2012, p. 4).

To be clear, I am not suggesting that the assessment field move forward with reckless abandon. I am suggesting, however, that (1) we actually investigate the impact of these types of items for students (not their parents) and not simply make assumptions based on fear; (2) agree as a field that, when the research does not yet exist, to lean into the content that is justice-oriented, and not fear-oriented, in the interim; and (3) acknowledge and prepare ourselves for remarkable and extensive resistance from stakeholders who would choose – at all costs – to uphold white supremacist logics in our schools and through our assessments.

To be sure, with an uncritical eye, avoiding topics such as racial injustice and social problems may appear to be a reasonable approach to protecting children. I argue, however, that this kind of erasure is simply racism disingenuously cloaked as a concern for the emotional well-being of students; and encourage assessment developers to reconsider excluding these topics by placing them on restricted lists. In fact, we must reconceptualize our fear-based framing of burdening students with stories of oppression and injustice and recognize/accept that minoritized students are already burdened. Antiracist assessments, then, hold the potential to lighten that burden by instead of rendering the experiences of minoritized students as invisible, non-existent, or imagined, acknowledging and

facilitating the right-making process. Through this process (of highlighting sociopolitical injustices and, importantly, *empowering* students to enact change), those have benefited from the long history of injustice can actively engage in facilitating the un-doing, or the healing process, and those who have endured the injustice can feel themselves seen. We know that assessments have the potential to influence and, in many cases, drive instruction. Instead of designing our assessments in such a way that they uphold the unjust status quo, they can serve as powerful levers of disruption.

In short, I propose that instead of simply warning item writers to handle important topics with care (a fear-based approach) that they be provided real guidelines about how to develop assessment tasks that are antiracist and justice-oriented. For example, the criteria could be does the task (1) empower students – through agency and/or allyship – to address a real issue of justice?; (2) present information/data that are accurate and comprehensive?; (3) disrupt a negative and/or false stereotype about a minoritized group?; (4) address some community-based, authentic need? (Lee, 1998); (5) draw on culturally based funds of knowledge (Lee, 1998)?; or (6) elevate a minoritized group – affirming their values, hopes, and understandings of the world? And, if an item does not meet one of these criteria, then reviewers could/should be encouraged (and trained how) to identify opportunities to integrate them.

Re-envisioning/defining what is meant by barriers and construct irrelevance

A consistent theme throughout sensitivity guidelines is the need to remove any barriers that might impede student success on an assessment; and these barriers can generally be classified as barriers related to (a) upsetting content and (b) construct irrelevant variance. I argue that, as a field, we need to re-envision/re-define what we consider barriers and construct irrelevance from an antiracist lens. The reader is encouraged to refer to Randall (2021) for a more comprehensive treatment of the inherent issues with current conceptualizations of construct irrelevance as it relates to a justice-oriented, antiracist approach to assessment broadly; and to Qualls (1998) for a discussion of construct under-representation (as it relates to issues with culturally responsive assessment). Here, I argue that for any justice-oriented assessment system, antiracism should be considered a central and integral component of the entire design process. As such, antiracist content (however uncomfortable) would not, in and of itself, automatically be considered a barrier or construct irrelevant.

For example, language- and the restrictions typical bias and sensitivity guidelines (as well as the Standards for Educational and Psychological Testing, AERA, APA, NCME, 2014) place on the acceptable language use through arguments of construct irrelevance – serves as an important example of why we must reconceptualize what we mean by/understand to be barriers – especially when considering racially minoritized students. Language is vital to a community's definition and understanding of their experiences in relation to their environments – both natural and social; and the entire universe (Wa Thiong'o, 1986). One's language represents so much of one's self and culture (Gelman & Roberts, 2017; Shashkevich, 2019), and assessments that choose to ignore the language formations of certain sociocultural identities in favor of elevating the language formations of the dominant (read: white) sociocultural identity (often referred to as formal due to white supremacist logics) is, without question, racist. Yet, for example, Smarter Balanced warns developers to "avoid language that is mainly used among people who know each other well or is more appropriate for relaxed and unofficial contexts. This includes slang (e.g., frenemy, brb), colloquialisms (e.g., gonna, sorta), [and] dialectal language (e.g., y'all, you betcha). Some ELA literary texts may include instances of formal language as measured by the corresponding standards" (p.13). Similarly, Florida Department of Education (2012) requires assessment designers to "Use vocabulary in test items that is widely accessible to all students and avoid unfamiliar vocabulary that is not directly related to the construct" (p.20).

This type of language commonly found in bias and sensitivity guidelines represents a white supremacist perspective to/ideology about language. It assumes (within the white supremacist hegemony) that there is one formal language and this language is standard edited American English (i.e., white mainstream English) despite linguists' assertions (Green, 2000; Rickford, 1999) that other

linguistic formations, such as African American English (AAE) are rule-based, complicated, linguistic systems and there is nothing inherently superior in WME. Such guidelines, however, in an effort to remove construct irrelevant variance related to language, in fact, establish barriers for students who do not represent whiteness. In other words, these guidelines imply that everything but "whiteness" should be considered a barrier, thereby marginalizing what whiteness has deemed as "other" – in this case, anything other than white mainstream English. In reconceptualizing what we mean by barriers with respect to language, we must interrogate our assumptions for evidence of white supremacist logics. An antiracist approach would acknowledge the multiple linguistic systems (especially those employed in racially minoritized communities) and afford them equal value to linguistic systems valued/employed via whiteness. Such an approach for evaluating the language used on an assessment would require that all test-takers' linguistic systems be respected and valued maintaining that *care should be taken to ensure that the language used in any assessment reflects the linguistic formations of the intended population. In cases in which multiple linguistic formations are represented within a population, assessment developers should work to ensure that no one linguistic system is privileged over another; but rather multiple systems are acknowledged, included, and affirmed*.

The uncritical restriction of the use of illustrations/artwork on assessments serves as another example of how the use of concerns about construct-irrelevance perpetuates, unintentionally, the white supremacist hegemony. Although this restriction is commonly articulated as a way to avoid distracting the test taker, I argue this logic supports and enables a white supremacist approach to test review and should be more critically interrogated. As Randall (2021) noted a context-free item is, typically, a white-centered item. When test developers opt not to show representations of people of color in conjunction with the item – even if the representation is not considered critically necessary to respond to the item – the implied representation will always be whiteness. In other words, each time an item refers to a person/group of people and no illustration accompanies this text, the assumption is the item is referring to a white person/people. Moreover, I argue that – when used – illustrations should be employed to disrupt – and not merely perpetuate – stereotypes and biases regardless of context. For example, Educational Testing Service (2016) writes, when referring to international assessments, "Illustrations that are intended to aid understanding may be a source of construct-irrelevant difficulty if the depictions of the people do not meet the cultural expectations of test takers in countries other than the United States. People intended to be professors, for example, should look older than the students depicted and should be dressed conservatively" (p.19). I maintain that illustrations should be used to rupture white supremacist, paternalistic logics (disguised as traditional or conservative) and bias and sensitivity guidelines should not be encouraging assessment developers to reinforce them. A guideline with respect to illustrations framed in antiracist logics would instead encourage test developers *to employ the liberal use of illustrations to rupture negative stereotypes, increase representation of historically minoritized persons, and rupture notions of whiteness as neutral and/or superior*. In fact, in an effort to combat (compensate for) the long history of inaccurate or nonexistent depictions of historically minoritized persons, the assessment industry should consider employing the use of illustrations in a manner that overrepresents these previously marginalized sociocultural identities (see prior text referring to Mills (2013) rectificatory justice).

As a final example, I refer to a more subtle employment of the concept of construct irrelevance to further marginalize sociocultural identities that have been othered. Smarter Balanced cautions: "refer to people by orientation only when it is relevant to the construct being measured while using widely-accepted inclusive terminology and allowing room for self-identification wherever possible" (p.18). A cursory read of this guideline might suggest that is simply seeking to remove a [supposed] barrier, or distraction; however, a more critical review recognizes that heteronormativity requires that the default assumption one makes about orientation will always be straight and cis-gendered – and that assumption is rarely relevant to the construct being measured, but is allowed to exist. That lesbian, gay, or queer identities are allowed to "exist" on an assessment only when the identity is directly related to a construct (likely defined with heteronormative ideals at the center) is an example of the assessment field's inherently problematic approach to defining what constitutes (or does not) a barrier. Representations

of white, cisgendered, heterosexual men are rarely, if ever, considered to be a barrier on assessments; yet every other sociocultural identity must be immediately scrutinized for *degree of possible distraction* during the item development/review process. The issue here goes beyond the acceptance and inclusion of LGBTQ+ content – relevant or irrelevant – on an assessment. The broader problem is bias and sensitivity guidelines leveraging construct irrelevance as a tool of erasure. Because racism (Bell, CRT) and the supremacy of whiteness (Frankenberg, CWT) is so deeply embedded in our thinking, practices, policies and norms, a shift toward antiracist assessment will require a comprehensive revision/shift in the ways in which the field evaluates and interrogates its assessments (i.e., Freire's (1973) critical consciousness) for oppressive content. To be sure, in any justice-oriented assessment system rooted in an antiracist lens, we cannot continue to marginalize (or create barriers for) certain sociocultural identities in favor of not upsetting, confusing, and/or discomforting dominant sociocultural identities. With a justice-oriented, antiracist lens, we would concern ourselves with barriers that (a) result from an underlying construct that is, in and of itself, racist and white supremacist; (b) present content that is white-washed, incomplete, and/or skewed in an effort to protect whiteness; and (c) focuses on surface-level, meaningless depictions of marginalized groups failing to capture the complexities and inherent value of these groups.

The broader point I am making here requires us to hold three truths articulated in prior scholarship elsewhere in mind: (1) Field (or context)-independent approaches to learning (and subsequently assessment design) represent Eurocentric conceptualizations (see Shade, 1982; Willis, 1989 for a broader discussion); (2) Racially minoritized students must be allowed to engage with learning and assessments that are socially situated and that address the needs of their communities (see Lee, 1998); and (3) The struggle against race-based oppression is a shared struggle (Hooks & West, 2016) that must acknowledge and disrupt all systems of oppression. Given these truths, any assessment review process that attempts to strip away all context, ignoring the sociopolitical realities in which the learning and the assessment take place, is, by definition reinforcing a barrier for the most marginalized students and inadvertently privileging whiteness.

Professional development of critical consciousness

Within the context of educational testing, ultimately, what Critical Race Theory, Critical Whiteness Theory, and justice-based, antiracist framings demand is the need for a shift in – or development of – the critical consciousness of assessment developers and reviewers. Freire (1973) described critical consciousness as the process of recognizing, or uncovering, systemic inequities persistently perpetuated through the processes, procedures, guidelines, and policies found in our institutions. This uncovering is then followed by action against the oppressive elements that allow these inequities to exist/thrive. Such a shift in the critical consciousness of assessment professionals would require a new approach to the assessment design cycle that decenters whiteness, encourages us to think critically about the ways in which the values and ways of knowing of minoritized students are represented (or not represented) in assessments, and actively rupture processes and procedures that are inherently racist and oppressive.

Ladson-Billings (2000) wrote the issue is "not merely to 'color' the scholarship. It is to challenge the hegemonic structures (and symbols) that keep injustice and inequity in place. The work is not about dismissing the work of European-American scholars. Rather, it is about defining the limits of such scholarship" (p. 217). I make a similar argument with respect to the selection and use of sensitivity reviewers, because packing the review panels with Black, Brown, and Indigenous reviewers is not sufficient. Indeed, as argued earlier in this paper, white supremacy is so deeply rooted into American culture that the presence of diversity alone cannot stomp it out. Nonetheless, the assessment industry has relied heavily on this approach (i.e., racial representation) as the equitable solution to the problems identified in this paper. In her review of the sensitivity practices of several assessment development companies, Golubovich, Grand, Ryan, Schmitt, and Schmitt (2014) found that ethnic minorities and women were more likely than white and male reviewers to attribute their selection as

a reviewer to these demographic characteristics. Indeed, the inclusion of minoritized identities in the review process has been long recommended (see Camilli, 1993; Hood & Parker, 1989), but the presence of a Black body on such a panel does not guarantee the presence of a white supremacy disruptor. Rogers and Mosley (2006) explain that "Tools such as language, symbol systems, nonverbal gestures, art, and media all work to construct and represent whiteness as normalized and privileged. Competing values are seen as deviant. Through our tools for sense making, whiteness is normalized and the associated privileges are made invisible" (p.467). Consequently, whiteness and white supremacist logics are often difficult to recognize – even within communities of color – which allow for its continued perpetuation. To be sure, the ways in which educators/scholars of color have unintentionally served as agents of white supremacist logics has been documented (see, for example, Randall, Poe, Poe, & Slomp, 2021; Zuberi & Bonilla-Silva, 2008). A justice-oriented, antiracist review process requires assessment companies to commit to the ongoing and consistent (Darling-Hammond, Hyler, & Gardner, 2017; Doppelt et al., 2009; Penuel, Gallagher, & Moorthy, 2011; Yoon, Duncan, Lee, Scarloss, & Shapley, 2007) professional development required to raise the critical consciousness of all reviewers (white and minoritized).

Here, I refer to the 2019 10th grade English-Language Arts MCAS exam that received national attention as an example of this point. The students were asked to write a journal entry from the perspective of a character (Ethel from the novel *The Underground Railroad*) who was openly racist. Although the Department of Elementary and Secondary Education (DESE) announced that the question would not be scored and removed from all future exams, several organizations (e.g., the Massachusetts Teachers Association (MTA), the Massachusetts Education Justice Alliance, and the New England Conference of the National Association for the Advancement of Colored People) all demanded that the entire exam (and scores) be pulled due to the harm the question caused some students (Lisinki, 2019). The state education commissioner responded by noting: "The Department has a thorough process for vetting test questions that includes review by educators, review by a committee that looks at possible biases, and field tests of all questions before they are used toward students' scores." In fact, the bias and sensitivity review committee was composed of 14 members – 9 were African-American, Asian, or Hispanic. My point here is twofold: (1) Within an antiracist framework, students would not have been asked to represent the point of view of an obviously racist person. In this case, however, instead of empowering students to – through agency or allyship – address a racial injustice, they were asked to read about an injustice and give reasons for upholding it. (2) The review committee was reportedly quite diverse with respect to race and diversity, but the state of Massachusetts provided no evidence that the reviewers received any professional development to increase their critical consciousness. I maintain that (a) had the review committee been working from a set of antiracist criteria and (b) received training that focused on developing their critical consciousness (and not simply identifying obviously racist or sexist tasks), the outcome may have been different. It is not enough for our assessment processes and tasks to be not-racist, they must be antiracist. Similarly, it is not enough for those involved in the assessment development process to represent racially minoritized identities, they must be critical and committed to identifying and disrupting white supremacist logics.

The task of disrupting this deeply embedded white supremacist hegemony, however, is no easy/small one, particularly when identifying subject matter experts/reviewers for bias and sensitivity review boards. This re-orientation will require the assessment field to engage in intentional partnerships with existing justice-oriented organizations/networks (e.g., Abolitionist Teaching Network) to identify professionals who hold both the requisite content knowledge and a critical perspective. I ask that the profession also consider, at least in the interim as it builds the pool of critical reviewers, relying on the expertise of stakeholders who are not necessarily content experts to interrogate the context (and sub-context) of the assessment experience for evidence of white supremacist logics. Again, such an approach would require the field to engage in partnerships with non-assessments organizations (e.g., Village of Wisdom, Center for Racial Justice & Engaged Youth) to provide the necessary expertise currently lacking in assessment spaces. Such

partnerships, over time, will help to build the capacity of the assessment field to re-orient its work/practices *toward justice*.

Conclusion

It is important to keep in mind that the intent of bias and sensitivity review guidelines "is to remove unnecessary barriers to the success of diverse groups of test takers" (Smarter Balanced, p.24). The inclusion of antiracist content on an assessment may mistakenly be seen as an unnecessary barrier for students who are accustomed to having whiteness both placed at the center and protected in large-scale assessment. I argue, however, that if the barrier exists (and I do not concede that it does) then it is, indeed, a necessary barrier in the same way that difficult words and language (e.g., the inclusion of old/middle English or advanced vocabulary words) on an Advanced Placement Literature exam would be considered a necessary barrier. The shift to a justice-oriented approach to assessment requires us to acknowledge, value, and meaningfully include the sociocultural identities of all students. Such practices go beyond simply highlighting holidays, shading the skin tone of white characters, replacing the names *John* and *Mary* with *Juan* and *Maria*, and describing so-called ethnic foods in reading passages. Meaningful representation will require the inclusion of content that elevates the real contributions of marginalized populations (including their historical and contemporary efforts to resist oppressive systems of injustice), draws on their cultural funds of knowledge (Lee, 1998), and employs their linguistic systems (in the same way/to the same degree as white-centric linguistic systems are employed).

By focusing on the *uncritical* elimination any possible affective/emotional barriers, the assessment industry has, in essence, made the assessment itself a barrier for many students. James Baldwin articulated in his (1963) *A Talk to Teachers*: "If, for example, one managed to change the curriculum in all the schools so that Negroes learned more about themselves and their real contributions to this culture, you would be liberating not only Negroes, you'd be liberating white people who know nothing about their own history." I, of course, argue his words apply when considering the assessment development process. In an effort to *sanitize* our assessments through the use of stringent, Draconian-like bias and sensitivity guidelines, we rob all students of meaningful, impactful assessment experiences in favor of a so-assumed safe assessment experience. In other words, sensitivity guidelines developed from an antiracist lens are liberatory for all students, not just students from marginalized populations.

Acknowledgments

The author would like to thank Ellen Forte (EdCount) and Karla Egan (EdMetric) for their support of this research.

Disclosure statement

No potential conflict of interest was reported by the author.

References

Adichie, C. (2009). The danger of a single story. TED Talk. Transcript retrieved July 31, 2021 from https://www.ted.com/talks/chimamanda_ngozi_adichie_the_danger_of_a_single_story/transcript?language=en.

Alcoff, L. (1988). Cultural feminism versus post-structuralism: The identity crisis in feminist theory. *Journal of Women and Culture in Society*, *13*(3), 405–436. doi:10.1086/494426

American Educational Research Association, American Psychological Association, National Council for Measurement in Education. (2014). *The Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.

Amrein, A., & Berliner, D. (2003). The effects of high-stakes testing on student motivation and learning: A research report. *Association for Supervision and Curriculum Development Educational Leadership*, *60*(5), 32–38.

Au, W. (2011). Teaching under the new Taylorism: High-stakes testing and the standardization of the 21st century curriculum. *Journal of Curriculum Studies*, *43*(1), 25–45. doi:10.1080/00220272.2010.521261

Baldwin, J. (1963). A talk to teachers. (Delivered October 16, 1963, as "The Negro Child – His Self-Image"; originally published in The Saturday Review, December 21, 1963, reprinted in The Price of the Ticket, Collected Non-Fiction 1948-1985, Saint Martins 1985). Retrieved from on July 19, 2021 https://richgibson.com/talktoteachers.htm .

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

Bell, D. A. (1995). Who's afraid of critical race theory? *University of Illinois Law Review*, *4*, 893–910.

Bonilla-Silva, E. (2013). *Racism without Racists: Color blind racism and the persistence of racial inequality in America*. Washington, DC: Rowman and Littlefield Publishers.

Camilli, G. (1993). The case against item bias detection techniques based on internal criteria: Do item bias procedures obscure test fairness issues? In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 397–418). Hillsdale, NJ: Lawrence Erlbaum Associates.

Carastathis, A. (2016). *Intersectionality: Origins, Contestations, Horizons*. Lincoln: University of Nebraska Press. doi:10.2307/j.ctt1fzhfz8

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

Darling-Hammond, L., Hyler, M., & Gardner, M. (2017). Effective Teacher Professional Development. Palo Alto, CA: Learning Policy Institute. Retrieved July 28, 2021 from https://learningpolicyinstitute.org/sites/default/files/product-files/Effective_Teacher_Professional_Development_REPORT.pdf.

Data Recognition Corporation. (2003). *Manual for issues of bias, fairness and sensitivity*. Maple Grove, MI: Author. Retrieved August 1, 2021 from https://education.alaska.gov/tls/Assessments/techreports/spring12_sbascience/sp12sbascience_app3.pdf

Delgado, R. (1995). *Critical Race Theory: The Cutting Edge*. Philadelphia: Temple University Press.

Delgado, R., & Stefancic, J. (2017). *Critical Race Theory: An Introduction* (3rd ed.). New York: New York University Press. doi:10.2307/j.ctt1ggjjn3

Dillard, C. B. (2000). The substance of things hoped for, the evidence of things not seen: Examining an endarkened feminist epistemology in educational research and leadership. *International Journal of Qualitative Studies in Education*, *13*(6), 661–681. doi:10.1080/09518390050211565

Doppelt, Y., Schunn, C. D., Silk, E. M., Mehalik, M. M., Reynolds, B., & Ward, E. (2009). Evaluating the impact of a facilitated learning community approach to professional development on teacher practice and student achievement. *Research in Science and Technological Education*, *27*(3), 339–354. doi:10.1080/02635140903166026

Educational Testing Service. (2002). *ETS standards for quality and fairness*. Princeton, NJ: Publisher.

Educational Testing Service. (2016). *ETS guidelines for fair tests and communications*. Princeton, NJ: Author.

Elliot, N. (2016). A theory of ethics for writing assessment. *Journal of Writing Assessment*, *9*(1), Retrieved from. http://journalofwritingassessment.org/article.php?article=99

Florida Department of Education. (2012). *Bias and sensitivity review: District developed assessments*. Office of race to the top assessments. *Division of Accountability, Research, and Measurement*. https://www.fldoe.org/core/fileparse.php/5423/urlt/bsrdda.pdf

Foote, M. Q., & Bartell, T. G. (2011). Pathways to equity in mathematics education: How life experiences impact researcher positionality. *Educational Studies in Mathematics*, *78*(1), 45–68. doi:10.1007/s10649-011-9309-2

Frankenberg, R. (1993). *White women, race matters: The social construction of whiteness*. Minneapolis, MN: University of Minnesota Press. doi:10.4324/9780203973431

Freire, P. (1973). *Education for critical consciousness*. New York: Continuum.

Gelman, S., & Roberts, S. (2017). How language shapes the cultural inheritance of categories. Proceedings of the National Academy of Sciences of the United States of America. Retrieved from https://www.pnas.org/content/114/30/7900 on July 21, 2021.

Golubovich, J., Grand, J., Ryan, A., & Schmitt, N. (2014). An examination of common sensitivity review practices in test development. *International Journal of Selection and Assessment*, *22*(1), 1–11. doi:10.1111/ijsa.12052

Green, L. (2000). *African American English: A linguistic introduction*. Cambridge, UK: Cambridge Press.

Hamad, R. (2020). *White tears/brown scars: How white feminism betrays women of color*. New York: Catapult. doi:10.2307/jj.1744983

Holmes, A. G. D. (2020). Researcher positionality - a consideration of its influence and place in qualitative research - a new researcher guide. *Shanlax International Journal of Education*, *8*(1), 1–10. doi:10.34293/education.v8i2.1477

Hood, S., & Parker, L. J. (1989). Minority bias review panels and teacher testing for initial certification: A comparison of two states' efforts. *The Journal of Negro Education*, *58*(4), 511–519. doi:10.2307/2295208

Hooks, B. (1984). *From margin to center*. Boston: South End Press.

Hooks, B. (2003). Teaching community: A pedagogy of hope. doi:10.1093/jn/133.2.567S

Hooks, B., & West, C. (2016). *Breaking bread: Insurgent black intellectual life*. New York: Routledge.

Kendi, I. (2019). *How to be an antiracist*. New York: One World.

Ladson-Billings, G. (1998). Just what is critical race theory and what's it doing in a nice field like education? *International Journal of Qualitative Studies in Education*, *11*(1), 7–24. doi:10.1080/095183998236863
Ladson-Billings, G. (2000). Racialized discourses and ethnic epistemologies. In N. Denzin & Y. Lincoln (Eds.), *Handbook of Qualitative Research* (2nd ed., pp. 257–277). Thousand Oaks, CA: Sage Publications.
Lee, C. (1998). Culturally responsive pedagogy and performance-based assessment. *The Journal of Negro Education*, *67* (3), 268–279. doi:10.2307/2668195
Lisinki, C. (2019). "Traumatic MCAS question removed from exam after students complain. WBUR Local Coverage. Retrieved May 11, 2022 from https://www.wbur.org/news/2019/04/04/underground-railroad-mcas-question .
Livingston, G. (2018). *They're Waiting Longer, but U.S. Women Today More Likely to Have Children Than a Decade Ago*. Retrieved July 28, 2021 from https://www.pewresearch.org/social-trends/2018/01/18/theyre-waiting-longer-but-u-s-women-today-more-likely-to-have-children-than-a-decade-ago/
McIntyre, A. (1997). *Making meaning of whiteness: Exploring racial identity with white teachers*. Albany: State University of New York Press.
Mills, C. (1999). *The racial contract*. Ithica: Cornell University Press.
Mills, C. (2007). White ignorance. In S. Sullivan & N. Tuana (Eds.), *Race and Epistemologies of Ignorance* (pp. 13–38). Albany: State University of New York Press.
Mills, C. (2013). Retrieving Rawls for racial justice?: A critique of Tommie Shelby. *Critical Philosophy of Race*, *1*(1), 1–27. doi:10.5325/critphilrace.1.1.0001
Morrison, T. (1992). *Playing in the dark: Whiteness and the literary imagination*. Cambridge, MA: Harvard University Press.
Penuel, W. R., Gallagher, L. P., & Moorthy, S. (2011). Preparing teachers to design sequences of instruction in earth systems science: A comparison of three professional development programs. *American Educational Research Journal*, *48*(4), 996–1025. doi:10.3102/0002831211410864
Qualls, A. (1998). Culturally responsive assessment: Development strategies and validity issues. *The Journal of Negro Education*, *67*(3), 296–301. doi:10.2307/2668197
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. (2022, November). *Assessment reparations?: We got next: A Call to action to the measurement community. 43rd annual Charles H. Thompson Lecture-Colloquium*. Washington DC: Howard University.
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
Rawls, J. (1999). *A Theory of Justice* (Rev. ed). Cambridge, MA: Harvard University Press.
Rhone, A. E. (2006). Issues in education: Preparing minority students for high-stakes tests: who are we cheating? *Childhood Education*, *82*(4), 233–235. doi:10.1080/00094056.2006.10522830
Rickford, J. R. (1999). *African American vernacular English: Features, evolution, educational implications*. Malden, MA: Blackwell.
Rogers, R., & Mosley, M. (2006). Racial literacy in a second-grade classroom: Critical race theory, whiteness studies, and literacy research. *Reading Research Quarterly*, *41*(4), 462–495. doi:10.1598/RRQ.41.4.3
Shade, B. (1982). Afro-American cognitive style: A variable in school success. *Review of Educational Research*, *52*(2), 219–244. doi:10.3102/00346543052002219
Shashkevich, A. (2019). The power of language: How words shape people, culture. *Stanford News*. Retrieved from https://news.stanford.edu/2019/08/22/the-power-of-language-how-words-shape-people-culture/ July 12, 2021.
Smarter Balanced Assessment Consortium. (2021). *Bias and sensitivity guidelines*. Oakland, CA: The Regents of the University of California.
Smith, L. T. (1999). *Decolonizing methodologies: Research and indigenous peoples*. London: Zed Books.
Styres, S. (2019). Decolonizing narratives, storying, and literature. In L. T. Smith, E. Tuck, & K. W. Yang (Eds.), *Indigenous and decolonizing studies in education: Mapping the long view* (pp. 24–37). New York: Routledge. doi:10.4324/9780429505010-2
Thompson, S., Johnstone, C. J., & Thurlow, M. L. (2002). *Universal design applied to large scale assessments* (Synthesis Report 44). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.
Wa Thiong'o, N. (1986). *Decolonizing the mind: The politics of language in African culture*. Nairobi, Kenya: East African Educational Publishers.
Willis, M. (1989). Learning styles of African American children. A review of literature and interventions. *Journal of Black Psychology*, *16*(1), 47–65. doi:10.1177/009579848901600105
Yoon, K. S., Duncan, T., Lee, S. W.-Y., Scarloss, B., & Shapley, K. (2007). *Reviewing the evidence on how teacher professional development affects student achievement* (Issues & Answers Report, REL 2007-No. 033).
Zuberi, T., & Bonilla-Silva, E. (2008). *White Logic, White Methods: Racism and Methodology*. Washington DC: Rowman & Littlefield Publishers.