

A peer reviewed, open-access electronic journal: ISSN 1531-7714

A Practical Approach in Co-Designing Culturally Responsive Mathematics Assessments

Maria Hamdani

Thao Vo, *The Center for Measurement Justice* 

Nisha Acharya Julien, *The Center for Measurement Justice* 

Abstract: This paper aims to address the critical need for practical guidance on developing Culturally Responsive Assessments (CRA) in U.S.-based educational systems. Traditional approaches, often designed with minimal inclusion of diverse cultural backgrounds, can exacerbate the marginalization of already marginalized populations. This paper addresses this exacerbation by presenting a structured approach for CRA development, emphasizing how to incorporate cultural context through sharing assessment development power. We detail a practical process, highlighting three unique culturally responsive procedures: “brainstorming sessions” to understand and gather authentic lived experiences of marginalized groups, blueprint development targeting specific culturally responsive qualities for item creation, and authenticity reviews with underrepresented groups. This approach aims to ensure that assessments are justice-oriented, engaging, and accurately reflect the capabilities of all learners, particularly those who have been historically and systematically excluded.

Keywords: Co-designing assessments, Culturally responsive assessments, Assessment framework

Introduction

Marginalized racial and ethnic communities have faced systemic educational exclusion throughout history, including in assessment processes, from brainstorming and ideation, to item development and writing, to implementation, data analysis, and use (Russell, 2023). Within the current assessment ecosystem, diverse ways of living, thinking, and behaving are often underrepresented (Frierson et al., 2002; Randall, 2025). This lack of representation can perpetuate inequities in how learning, knowledge, and abilities are measured. Therefore, it is essential to make space for diverse norms intentionally, as research demonstrates

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY-4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <https://creativecommons.org/licenses/by/4.0/>

 OPEN ACCESS.

that linguistic, racial, and ethnic diversity strengthens creative thinking, problem-solving, and systems-level approaches to increase innovation (Horowitz, 2019; National Institutes of Health, 2018).

One way to create space with and honor diverse ways of thinking and living is to foster social-emotional characteristics of care, curiosity, and empathy within a society (Noddings, 2012). These characteristics can be intentionally cultivated through U.S. K-12 standardized and classroom practices; however, they are rarely prioritized in assessment design (Walker et al., 2023). Recognizing the importance of these traits, the conceptual paradigm “cultural responsiveness” seeks to honor and center diverse approaches to learning, knowing, and applying knowledge while simultaneously promoting the development of these social-emotional characteristics (Gay, 2018; Hood, 1998; Ladson-Billings, 1995).

Efforts to establish practical strategies for integrating principles of culturally responsive assessments (CRA) into assessment development remain in early stages (Walker et al., 2023). This article shares a practical approach to applying cultural responsiveness in the item writing process by centering culturally diverse students’ lived experiences and contexts in K-12 standardized and classroom mathematics assessment design. The approach presented here has been strengthened through multiple iterations and reflection cycles by involving those most affected by assessments, like students, parents, teachers, and community members. Actively engaging those most impacted in the critique and construction of assessments has enabled the design of assessments that mirror learners’ sociocultural realities, a fundamental aspect of an effective CRA development process (Sul, 2019). While principles of CRAs can be broadly applied, the practical approach in this paper focuses on developing K-12 math assessment items.

How Does Culture Contribute to Assessments?

Culture can be defined as the shared characteristics, beliefs, and values that shape behaviors and practices within a specific group or population (Hood, 1998). The U.S. comprises a range of cultural groups, and recognizing that cultural norms vary across these groups is essential. As in, what is normative for one group may differ from what is normative for another, even when there are perceived similarities in race, ethnicity, and/or socioeconomic status. Implementing a CRA approach can help build systemic improvements to better serve diverse groups, rather than privileging a dominant Eurocentric perspective (Russell, 2023).

Assessments can be viewed as “social artifacts” designed to measure collective knowledge, skills, dispositions, and experiences; however, they often perpetuate the marginalization of diverse cultures (Solano-Flores, 2019). Traditional educational measurement approaches rely on constructs and content that disregard diverse contexts and norms, such as African American Vernacular English (AAVE). Disregarding diverse ways of speaking and language often resulted in harm to those marginalized by erroneously pathologizing language difference, reifying racialized stereotypes, and restricting access to educational opportunities for Black students, as evidenced by critiques of the Scholastic Aptitude Test (Randall, 2021). Restricting opportunities can include assessment scores as a criterion for decision-making at the school, district, and state levels, shaping policies and resource allocation (Moss & Schutz, 2001). Therefore, developing fair and comprehensive assessments for all learners is critical, given the impact associated with individual scores and school collective assessment score averages.

In this article, we define “fairness” as the principle that every individual and group should have an unobstructed opportunity to accurately demonstrate what they have learned in ways that honor their norms, feelings, or stance on the standard being measured (American Educational Research Association et al., 2014). Traditional notions of fairness in the assessment ecosystem are rooted in standardization and uniformity in assessment content, administration, scoring, and interpretation (Sireci & Randall et al., 2023; Sireci, 2020). These standardized approaches are often framed as “color neutral” and are widely practiced in K-12 standardized assessments (Randall, 2021). Although these perceptions of fairness may be well-intentioned, their impact on test-takers, particularly those from diverse cultural backgrounds, does not align with the

definition of fairness. Emerging research that centers on test-takers' experiences demonstrates that standardized testing is not experienced in "color-neutral" ways (Patterson, 2025). For example, cognitive interviews with students interacting with culturally responsive items demonstrated increased feelings of representation and engagement; in contrast, traditional items were perceived as disconnected or discomforting (Patterson, 2025).

A standardized one-size-fits-all approach that often ignores diversity in assessment creation and implementation perpetuates the marginalization of test-takers. This phenomenon, termed "measurement disjuncture," occurs when assessments developed for one culture are implemented in a culturally diverse one that differs from the mainstream (Sul, 2019). Measurement disjuncture affects how marginalized populations access and interpret assessment content, increasing their cognitive load as they attempt to decipher misaligned lived-experience contexts. Missing or misaligned cultural contexts in items can impact test-takers, notably students of color, leading to feelings of disconnection and difficulty engaging in assessments (e.g., Patterson, 2025). Individuals from marginalized groups are further disadvantaged because they must navigate mainstream (i.e., accepted) constructs while simultaneously applying standards-based learning (Randall, 2021). This dynamic can lead to inaccurate and inequitable measurement of standards-based competence, scores, and interpretations, and decision-making, serving as a gatekeeper of opportunities (Russell, 2023).

What are Culturally Responsive Assessments?

The concept of cultural responsiveness in education emerged in the early 1980s in pedagogy and teaching spaces, as educators and scholars recognized a disconnect between culturally diverse students and learning processes (Gay, 2018; González et al., 2006; Ladson-Billings, 1995). As an extension, CRA was introduced by early scholars as a way to connect assessment practices to students' cultural backgrounds and lived experiences (Hood, 1998; Lee, 1998). Progress has been gradual with a growing movement to create CRAs to address concerns about the mismeasurement of students from marginalized backgrounds. This commitment to centering students is evidenced by 35 states that are reinforcing the integration of cultural responsiveness practices and programs (e.g., educator recruitment and hiring reflective of diverse student populations) as part of their Every Student Succeeds Act (ESSA) plans (Schettino et al., 2019).

Assessments serve different roles depending on factors like design and implementation. Traditional assessments of learning often summarize performance to support decisions related to grades, placement, and/or program evaluations (Badrinarayan, 2024). In CRAs, the role is to support academic and cultural growth (Gordon, 2020). In CRAs, a summative assessment can serve as a learning opportunity when introducing students to new cultural content or perspectives (Gordon, 2020). For example, an assessment becomes a "window" of learning through exposure to unfamiliar cultural traditions or lived experiences (Gordon, 2020). Edmund Gordon explains in *Toward Assessment in the Service of Learning* (2020) that assessment should not be limited to ranking or sorting students. Instead, "assessment can and should be an educative experience from which students emerge with new understandings and connections" (Gordon, 2020, p. 74), further supporting principles of cultural responsiveness. CRAs are inherently this "window" for all students to learn new ideas and values from others, given the diverse inclusion of ways of being.

CRAs differ from traditional assessments because they uplift and connect assessments to the experience, values, and needs of all rights-holders, rather than some (Walker et al., 2023). *Rights-holders*, in accordance with the Declaration of Human Rights, are defined as individuals and groups directly affected by assessments and their scores and have traditionally had limited to no decision-making power during design, development, implementation, and/or implications (United Nations, n.d.). *Interest-holders*, such as educators, test developers and publishers, researchers, state education agencies, funders, policymakers, historically hold most, if not all, decision-making authority (Russell, 2023). Interest-holders can control procurement, timelines, funders, and project scope, especially in large-scale assessments.

Distinguishing rights-holders from interest-holders clarifies who holds decision-making power and to what extent. A key way CRAs uplift and strengthen assessment experiences is through their design, development, implementation, and evaluation, by intentionally asking whose cultures are being engaged and by ensuring cultures are genuinely included. Such inclusion requires engaging with various interest-holders early to secure institutional support and resources for meaningful collaboration that shares power in the design process (Walker et al., 2023).

Current CRA Approaches and Frameworks

Theoretical underpinnings of educational practices designed to serve the diverse needs of historically marginalized groups are rooted in culturally responsive pedagogy (Ladson-Billings, 1995; Lee, 1998) and culturally responsive instructional practices (Gay, 2002). Within classroom contexts, these approaches center diverse students' cultural experiences and identities by recognizing culture as a resource for learning rather than simply being celebrated (e.g., funds of knowledge; Paris & Alim, 2014). This perspective underscores shifting assessments to be in alignment further with instructional practices and intended learning outcomes (William, 2011).

In recent years, numerous frameworks, approaches, and principles have been proposed for implementing CRA across assessment contexts. One example is a K-12 education roadmap for designing responsive assessments for linguistically diverse groups, as language affects learners' understanding and responses to assessment items (e.g., Sireci & Faulkner-Bond, 2015). Language considerations in assessment design are important, especially when learners with limited dominant-language proficiency (i.e., White American Vernacular English or WAVE) take an assessment in that language but are unlikely to qualify for accommodations.

A critical extension of CRA is the Justice-Oriented Antiracist Validity (JAV) framework, which focuses on interrogating and dismantling racist logics embedded in assessment design and validation (Randall et al., 2022; Randall et al., 2023). JAV reframes validity as a justice-driven process that requires explicit attention to race, power, and historical inequities at every stage, from construct articulation to score interpretation. By embedding antiracist principles into validity arguments, JAV advocates for assessments that sustain students' cultures and languages, challenge deficit-based assumptions, and evaluate social consequences and technical evidence.

A socioculturally responsive assessment is a framework that conceptualizes learning and assessment as social and cultural artifacts (Bennett, 2025). This approach to CRA emphasizes the need to consider a multiplicity of test-taker social factors (e.g., race, ethnicity, gender identity, socioeconomic status, disability, immigration status, language) that might shape the assessment experience, and to design assessments that better serve these diverse needs (Bennett, 2023).

Randall et al. (2022, 2023) proposed heuristics, or guiding questions that assessment developers can ask when building assessments that center on the cultural experiences, norms, values, and needs of marginalized test-takers. The framework provides guiding questions that ask to what extent marginalized rights-holders are involved during the construct definition stage and whose perspectives and values the intended construct reflects (Randall, 2025; Randall et al., 2022, 2023). Similarly, a justice-based reading assessment framework has suggested key elements for designing CRA, highlighting the need for equal partnership and shared power in decision-making between interest and rights-holders (Forzani et al., 2024).

While discussing what CRA *is*, it is equally imperative to establish what it *is not*. It does not lean into cultural stereotypes about certain groups, nor does the approach lower standards of what diverse groups can do, know, or achieve. Such assessments do not "essentialize" or assume that all members of a particular group (e.g., racial-ethnic, gender, disability, etc.) share the same characteristics, traits, or experiences (Russell, 2024). In a similar vein, CRA does not tokenize culture by superficially including cultural elements in

assessments as a symbolic gesture of responsiveness, without making substantial efforts to include diverse cultural groups. Finally, CRA is not a one-size-fits-all approach; it emphasizes tailoring assessments to the specific needs and context of the rights-holders.

Commonalities in CRAs Frameworks and Approaches

While there are variations in operationalizing a CRA approach, whether from a cultural, sociocultural, or justice-oriented lens, there are three common themes: 1) Cultural context, 2) shared decision-making power, and 3) valuing flexibility. When it comes to surfacing cultural context, as noted earlier, “culture” extends beyond race and ethnicity. Bennett (2023) reminds us that culture encompasses a multitude of sociocultural identities and experiences, such as gender, sexual orientation, disability, and the intersections of these social categories, among others.

The second theme identified is shared decision-making power between dominant and marginalized populations. Shared decision-making intentionally planned and executed by interest-holders stops the process of marginalization of others by directly challenging traditional “top-down” approaches (Forzani et al., 2024). Engaging and integrating input from diverse interest-holders and rights-holders (e.g., students, families, community members, educators) directly creates an approach that includes all perspectives and voices (Forzani et al., 2024; Walker et al., 2023).

Lastly, valuing flexibility in design processes enables test-takers to have autonomy and ownership in identifying ways (e.g., tasks or formats) to demonstrate knowledge of the constructs an assessment intends to measure. Creating student agency in expressing knowledge uplifts students' authentic cultural knowledge reflected in performance (e.g., Bennett, 2023). These common themes illustrate a theoretical guide for a practical approach on how to create a CRA and disrupt traditional development approaches.

How The Practical Approach Proposed Contributes to CRAs

Practical guidance for implementing CRA theoretical frameworks and principles is necessary for K-12 assessments that predominantly reflect dominant White cultural norms and do not include voices from historically marginalized communities (Taylor & Ferrara, 2025). Therefore, the primary purpose of this paper is to demonstrate a practical approach to co-designing the development process for large-scale K-12 math assessment items. In this paper, we provide strategies and examples for the assessment design process to address measurement disjuncture and an equity gap in traditional assessment development: the systematic exclusion of marginalized, diverse cultural and social perspectives (Sul, 2019; Taylor & Ferrara, 2025).

The purpose of this paper is also *not* to expand on cultural competency or existing frameworks, nor to provide a comprehensive literature review, but rather to operationalize the CRA theory and principles into actionable steps for item development. We recommend readers seeking a review of theoretical foundations to read *Culturally and Socially Responsible Assessment: Theory, Research, and Practice* by Evan and Taylor (2022). In CRA development, we reference researchers throughout to guide readers toward additional CRA considerations, such as rights-holder engagement across diverse contexts, including within Indigenous communities (e.g., Englert & Schultz, 2025; Keehne et al., 2025) and classroom-based contexts (e.g., Middleton & Haynes, 2025).

The proposed practical approach presented in this paper has undergone multiple iterations. The first iteration involved creating CRA math items for grades 3-8. In this iteration, we gathered ideas for context from college students of color, these students were trained to write the initial item by an expert assessment developer and then reviewed by assessment experts. A key refinement from that first iteration was that rights-holders wanted more illustrations to support understanding of cultural context or social inequity, and they wanted citations for factual claims within items. In another iteration (Hamdani & Julien, 2025), high

school students were asked to support by sharing ideas via an online board. Within this online board, we posed two questions: “*What is something about your culture that is unique or that you are proud of?*” and “*Tell us about a social inequity you see in your community.*” In asking these questions, we realized that the answers leaned heavily towards food and clothing, making it difficult to address the contextual nuances of cultural experiences, values, and practices. In the third and fourth iterations, we created a 7th- and 8th-grade math CRA and we refined the brainstorming process by explicitly grounding questions in Geneva Gay’s tenets of culturally responsive teaching. This shift generated richer, more usable contexts connected to lived experience and community dynamics. In the fourth iteration, these tenets were adapted into assessment-specific qualities, reflecting the shift from instructional framing to operational assessment design. Building on these iterations, we also expanded our item formats to include options such as audio for text-heavy items and short videos to reduce cognitive load during reading (Julien et al., 2026).

A Practical & Structured Approach In Developing CRA Items

While traditional processes are instrumental in assessment development, as seen in Figure 1, there are four additional phases in our CRA approach, including Phase 0 (Interest-holder Alignment), Phase I (Brainstorming), Phase III (Target Culturally Responsive Qualities), and Phase V (Authenticity Reviews and Independent Expert Culturally Responsive Alignment), in addition to the traditional phases (light cells). The additional phases implement the three common themes previously stated within existing frameworks and are major steps in CRA item development. Each phase of the process is described in the following sections.

Phase 0: Interest-Holder Alignment

Securing alignment with key interest-holders such as state education agencies, test vendors, and funding partners is a critical first step in embedding CRA into assessment development. Large-scale assessment development typically occurs within Request For Proposal (RFP)-driven processes, where timelines, deliverables, and budgets are predetermined. In these cases, rights-holder engagement must be prioritized early on and intentionally built into the project scope. This step involves communicating the purpose and value of CRAs, identifying where co-design fits within existing development timelines, and ensuring that resources are allocated to support engagement, including compensation for rights-holders. Establishing this alignment early creates the conditions for authentic collaboration throughout the development process.

Alongside alignment, collective reflection on positionality should occur as an ongoing exercise for both interest-holders and rights-holders. Positionality refers to understanding how one’s racial, cultural, and social identities—and the power and privilege associated with them—shape perspectives and decisions (Milner, 2007). Milner emphasizes that failing to interrogate these influences can lead to “dangers seen, unseen, and unforeseen,” such as misinterpretation or perpetuation of deficit narratives of assessment constructs and content. Reflection should begin during procurement and proposal development and continue through phases like brainstorming and item writing. Key questions can include: *How do my cultural and racial backgrounds influence what I prioritize in assessment design? Whose perspectives are centered or marginalized in my decisions?* Engaging in this process collectively helps mitigate power imbalances and fosters equity in assessment development.

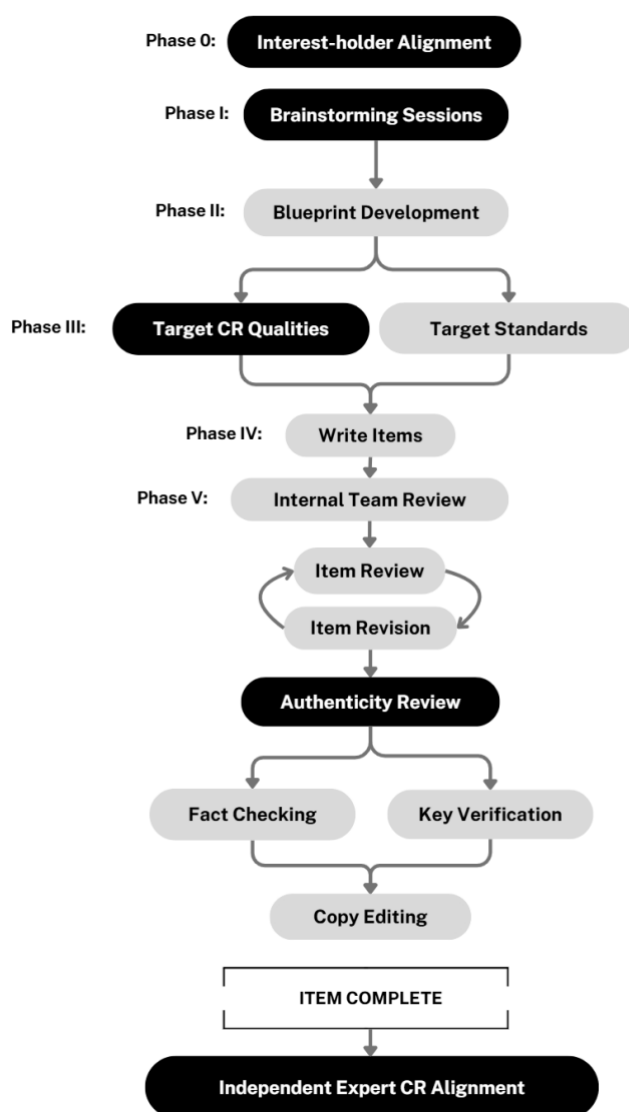
Phase I: Brainstorming Sessions

Developing CRA items begins by building collaborative, trusting spaces with diverse representation of rights-holders, as they are most impacted by assessment outcomes. While students are essential participants in this process, rights-holders can also extend to families, teachers, and other important community members, particularly when assessments target very young learners who may not be able to participate meaningfully in brainstorming sessions. In early childhood or elementary contexts, proxy voices such as parents and caregivers provide insights into the cultural practices, values, and lived experiences that shape

children’s learning environments. Community members (e.g., cultural leaders, Elders, local organizers) further contribute perspectives on sociocultural norms and traditions relevant to the assessment context, ensuring authentic representation even when direct student input is limited.

Brainstorming sessions can take various forms depending on the context. For test-takers who comprehend instructions, open-ended questions that elicit lived experiences and cultural contexts might be more suitable. Younger learners, however, may require engagement with parents and caregivers through interviews or focus groups. Although the practical approach described in this paper did not include direct brainstorming with very young learners, example engagement questions that could be adapted for caregivers might include, “*What cultural traditions, values, or practices are important in you and your community that you would like to see in assessments?*” and “*Are there community events or experiences that could provide meaningful contexts for learning*

Figure 1. A Flow Chart of the Phases of the Culturally Responsive Assessment Item Development Process



Note: Phases II–V require intentional training in CRA practices and deliberate inclusion of rights-holders and marginalized perspectives in item writing, review, and authenticity review processes.

tasks?” Sessions typically include one to three students from the identified rights-holder group to gather rich qualitative data about their experiences, beliefs, and values. It is important to note that while one to three students may be included from a specific marginalized population, these experiences are unique within their own communities rather than a generalized experience for those within that group.

Two strategies operationalize brainstorming effectively. The first is creating a safe and trusting environment, both before and during engagement with rights-holders. Acknowledging and reflecting on one’s positionality (Phase 0, Alignment) helps promote equitable participation and develop active, reflective listening skills (i.e., listening to understand, not respond) to build trust that mitigates power dynamics. When safety and trust are established, brainstorming sessions can build on this foundation to elicit meaningful discussions around the cultural elements students perceive as most meaningful, and that should be reflected in assessment content (e.g., experiences, practices, and narratives). In exchange for their time and insights, we recommend that all participants be compensated to acknowledge the value of their time and for sharing their lived experiences. It is essential to clearly state how personal information will be collected and used. For example, deficit-based content, such as violence, substance use, or insensitive portrayals that perpetuate stereotypes, should *not* be used. Instead, asset-based insights are prioritized and stored in a repository of cultural contexts, which can serve as a resource for item writers to integrate authentic cultural content into item content.

The second strategy is to curate discussion protocols (i.e., interview or focus group questions) for the brainstorming session that align with a specific culturally responsive framework. Our item development example is aligned with Geneva Gay’s (2002) Culturally Responsive Teaching (CRT) qualities. Example prompts include: *“Can you think of a situation where you felt a part of your background was not represented in school? How could that have been different?”* These questions were specifically designed to elicit responses that align with the CRT quality of “Validating” (i.e., when student identity is affirmed through assessments). Another prompt is, *“Are there any social problems that you think could/should be addressed collaboratively in school?”* This question was designed to align with the CRT quality of “Empowering” (i.e., students employing collaborative problem-solving to highlight students’ strengths). In line with common qualitative methods, responses are audio-recorded, transcribed, and summarized into structural CRT qualities (e.g., validating, empowerment, collaboration; see Table 2). Table 1 illustrates an examination of brainstorming questions and paraphrased student cultural context narratives. These examples are intended as a starting point rather than a prescriptive set of questions. The specific prompts used should vary based on the assessment’s purpose, context, and the priorities identified by interest-holders.

Phase II: Blueprint Development

CRA item development begins with a deliberate blueprinting process that defines the assessment’s overall structure (i.e., how the assessment is organized), purpose (i.e., why the assessment is being administered), and scope (i.e., what content or skills will be measured). In this paper, the Blueprint Phase focuses specifically on multiple-choice items, which remain the most widely used format in large-scale assessments. While our emphasis is on developing culturally responsive multiple-choice items, we acknowledge that there are many alternative item types (e.g., constructed-response items, performance tasks) that offer additional opportunities for equity and inclusion. For guidance on designing and scoring such alternatives through a CR lens, see White et al. (2025).

Like the traditional processes, the blueprint serves as a roadmap for fair and equitable item construction. Since our approach included Phase 0 Brainstorming, the blueprint will also embed CRA principles into each specification, including (a) targeted standards remain consistent across groups but are contextualized through culturally relevant scenarios to make them accessible and meaningful, (b) cognitive complexity levels are maintained for rigor but supported with familiar contexts while reducing unnecessary linguistic or cultural barriers, and (c) item formats are selected to minimize bias and maximize accessibility (e.g., avoiding

Table 1. Example Repository of Students’ Brainstormed Cultural Contexts

Brainstorming Question	Students’ Cultural Context
Can you think of a situation where you felt a part of your background was not represented in school? How could that have been different?	<i>A student observed that school photographs and calendars predominantly featured white children. They expressed a desire for these visuals to reflect the student body's true diversity.</i>
Can you think of an experience in school (or any educational setting) when you recognized your lived experience in a lesson? Or assessment? In other words, have you ever had a “this reminds me of…” experience in an educational setting? Describe the lesson/activity and how it relates to your own lived experience.	<i>After a lesson on food deserts and redlining, a student witnessed a striking contrast between adjacent neighborhoods—one affluent and well-maintained, the other visibly dilapidated and neglected. This firsthand observation solidified their understanding of the real-world impact of redlining.</i>
What are some of the lived experiences you wish would “show up” on an assessment?	<i>A student emphasized the need for more positive examples of Black fatherly love to counteract the persistent and untrue stereotype of absent Black fathers.</i>
What are some things about your culture or identity that you wish others understood better?	<i>A student pointed out the urgency of dismantling the harmful stereotype that all Black people are “thugs,” asserting that one’s manner of dress does not equate to criminality.</i>
Are there any social problems that could, or should be, addressed collaboratively in school?	<i>A student highlighted the significance of avoiding microaggressions, providing the example of touching someone's hair without permission.</i>
Can you think of a time when you felt encouraged or supported to take part in a project or activity that made a difference in your community?	<i>A student shared that their school's Black Student Union empowers Black student leaders through conferences, noting that strong teacher support significantly enhances these efforts.</i>

culturally unfamiliar references in distractors, providing visuals where appropriate). Rather than creating different standards for different groups, the blueprint emphasizes alignment with common academic standards while integrating cultural contexts that validate students’ identities and experiences. This ensures fairness through representation and accessibility, rather than lowering assessment expectations or fragmenting standards.

Phase III: Targeting CRA Qualities & Standards

A culturally responsive blueprint incorporates a set of guiding principles that define how cultural knowledge and equity values are embedded throughout the assessment design process. These principles, outlined in Table 2, align with an adaptation of Geneva Gay’s CRT, which has served as a foundation for culturally responsive and antiracist item development in previous empirical studies (see Randall, 2025).

To operationalize these culturally responsive assessment principles, a blueprint must clearly define the target CRA quality, the contextual representation, and the item type, alongside traditional specifications such as standards alignment, Depth of Knowledge (DOK), and difficulty level. Table 3 illustrates an example of assessment specifications that embed CRA directly into the test blueprint. We note that in this example, not

all traditional categories are included. This example illustrates a streamlined operational blueprint that emphasizes specific CRA specifications. In practice, additional operational columns such as workflow columns, item writer and reviewer columns, as well as internal review status columns, and metatags for each item would be present. The sample blueprint demonstrates how culturally grounded contexts, derived from the brainstorming sessions, are paired with relevant academic standards to ensure cultural authenticity and rigor.

Table 2. An Assessment Adaptation of Gay’s (2002) Culturally Responsive Teaching Qualities

Quality	Definition
Validating	Culturally responsive assessment leverages students' cultural knowledge, strengths, and backgrounds to create relevant assessments (or assessment tasks) that bridge the gap between academic concepts and their lived experiences. Validating assessments also challenges stereotypes and promotes social justice.
Comprehensive and Inclusive	Culturally responsive assessment employs cultural resources in the assessment process to maintain students' ethnic identities, community connections, and success ethic.
Multidimensional	Culturally responsive assessment is integrated into the instructional process so that students see the connection between the curriculum, their lived experiences, and the assessment, necessitating the use of a broad spectrum of cultural knowledge and experiences to complete it. Students have agency and are given choice and input (e.g., co-developing a rubric) in evaluating their performance.
Empowering	Culturally responsive assessments and assessment tasks are asset-based, leveraging and highlighting what students know and do well, and encouraging collaborative problem-solving and the acquisition of cultural capital.
Transformative	Culturally responsive assessment fosters students’ cultural consciousness, empowering them as social critics and agents of change.
Emancipatory	Culturally responsive assessments challenge the notion of one absolute scholarly truth (allowing for multiple ways of knowing, understanding, and doing), empowering students to contest and contextualize multiple perspectives. They require critical thinking and active engagement through diverse cultural lenses.
Humanistic	Culturally responsive assessment fosters a deeper understanding of self and others, benefiting both dominant and marginalized populations by providing insights into diverse cultures, histories, and achievements, thereby promoting empathy and interconnectedness across ethnic, racial, and social groups.
Normative and Ethical	Culturally responsive assessments and assessment tasks challenge Eurocentric norms embedded in traditional assessments; do not seek to elevate or protect whiteness.

Note: Qualities derive from Geneva Gay’s principles of culturally responsive teaching (2018) and adapted for culturally responsive assessments by Jennifer Randall (2025).

Table 3. Example of Culturally Responsive Mathematics Assessment Blueprint

Standard	Brainstormed Topic	Targeted CR Quality	Item Type	Intended Representation	Cultural Source Type	Authenticity Review	Geneva Gay Tenet – Final
8.SP.A.1 - Construct and interpret scatter plots for bivariate measurement data	Indian Holi festival - color packets	Comprehensive & Inclusive	Multiple Choice	South Asian; all genders	Authentic	ok - Writer	Comprehensive & Inclusive
8.NS.A.1 - Understand irrational numbers and approximate them using rational numbers	Multi-generational family grocery budgeting	Validating	Multiple Choice	Multigenerational; working-class family	Authentic	ok - Writer	Validating
8.G.B.7 - Apply Pythagorean Theorem to determine unknown side lengths	Black father and son building a garden bed - knowledge-sharing	Empowering	Drag-and-Drop	Black/African American father and son; intergenerational + community	Authentic	ok - Writer	Empowering
8.EE.B.5 - Graph proportional relationships; use unit rate to compare	Mobile health clinic schedule and travel times in the community	Humanistic	Fill-in-the-Blank	Indigenous	Research-Based	ok - (cultural authenticator)	Humanistic
8.EE.C.7 - Solve linear equations in one variable	LGBTQ+ youth center fundraiser	Transformative	Multiple Choice	LGBTQ+ youth; gender-inclusive	Research-Based	ok - (cultural authenticator)	Transformative
8.SP.A.2 - Informally assess the degree of linear relationship between two quantitative variables	Comparing official city air-quality data with community-collected sensor readings	Emancipatory	Multiple Choice	Black/Latinx neighborhood near industrial area	Authentic	ok - (cultural authenticator)	Emancipatory
8.SP.A.3 - Compare and interpret ratios and proportional relationships in real-world contexts	Representation inequities in textbook word-problem examples	Normative & Ethical	Multiple Choice	School community	Research-Based	n/a	Normative & Ethical

In addition to the blueprint, item writers also benefit from having clear guidance on the frequency of different CRA qualities across the item set. The following is a sample percent allocation of the qualities: Comprehensive & Inclusive (~20%); Validating (~15%); Empowering (~15%); Transformative (~10%); Humanistic (~15%); Emancipatory (~10%); Normative & Ethical (~10%); Multidimensional (~5%). Establishing these allocations serves two purposes: it helps item writers understand the intended balance of qualities during development and provides reviewers with a benchmark for evaluating overall coverage.

Phase IV: Item Planning and Writing Process

Once the assessment blueprint is finalized, the development team initiates an iterative item writing process guided by the CRA qualities. Each item writer is assigned a specific set of standards to address and detailed item specifications that align with the blueprint. These specifications include the targeted educational standard, the cognitive complexity level, the designated culturally responsive quality, and the intended demographic focus. Writers must also attend to the overall distribution of difficulty and DoK levels across the assessment. For example, a common guideline is to have a 1:2:1 ratio of easy, medium, and difficult items. In addition, writers follow predetermined allocations of CRA qualities to maintain balanced coverage of principles such as validating, empowering, and transformative practices, ensuring both academic rigor and cultural inclusivity. To support the item development process, writers should also be equipped with a range of foundational resources, including but not limited to: (a) assessment writing training, (b) culturally responsive tenets, qualities, or principles (c) academic standards, a grade-level word selection tool, and (d) a repository of brainstormed cultural contexts generated through early engagement with rights-holders. These materials equip item writers to develop instructionally sound and culturally responsive items. In the context of CRA, terms must align with academic standards while avoiding stereotypes or biases, and instead, offer positive and affirming representations of cultural traits and experiences. Writers should also be attentive to locally defined sensitivities (e.g., inappropriate, insensitive, and/or offensive topics that are deficit-based and reinforcing stereotypes, such as violence or substance abuse).

Throughout the item writing phase, regular check-ins with the item development team should be facilitated to monitor progress and ensure that item coverage reflects the full range of instructional goals and cultural diversity outlined in the blueprint. Guiding questions prompt reflection and support continuous alignment, can include: *How are item specifications distributed equitably across writers? How does the team ensure the comprehensive representation of content and cultural targets? How are item writers' own worldviews being acknowledged and meaningfully reflected in their work?*

The goal of item writing is to create meaningful, contextually relevant scenarios that are aligned with the assessment's purpose and audience. While item writers are encouraged to incorporate scenarios from the brainstormed repository created during earlier stages of development, using these examples is not mandatory for every item. Writers may also draw from their own lived experiences, particularly those connected to marginalized identities, or the lived experiences of others, provided each item undergoes thorough review during later stages of development. When the repository lacks representation of a particular group or socio-political issue, writers are encouraged to conduct research to develop new, culturally grounded contexts. In such cases, sources should be cited within the item to ensure transparency and accountability. Figure 2 demonstrates an item with a supporting citation. Each context must be purposeful, inclusive, and meaningfully connected to students' lived realities, regardless of origin. Scenarios may reflect real-world applications or cultural references that are either familiar to or enriching for the test-taker. Scenarios should remain concise while embedding culturally relevant details.

Where appropriate, writers are encouraged to include visual aids, such as graphs or tables, to enhance comprehension and reduce the need for explanatory text. For example, in a math item involving henna, including an image of henna applied on the palm allows the test-taker to immediately recognize and understand the cultural reference without requiring a lengthy explanation. The use of imagery supports both

Figure 2. Example Culturally Responsive Mathematics Item with a Supporting Citation

With the popularity of social media, there is potential for a variety of perspectives to be shared. However, some creators on one social media platform say they have been targeted for posting Black Lives Matter content. They say videos have been taken down, muted, or hidden from followers. Many Black creators have accused the platform of being anti-Black and not valuing them.

The table shows an example of videos taken down each day for a week.

Day of the Week	1	2	3	4	5	6	7
Number of Videos Taken Down	5	10	15	20	25	30	35

Yuan, J. (2021, February 5). *Months after TikTok apologized to Black creators, many say little has changed.* NBC News

1. What is the constant of proportionality for the relationship of the number of videos taken down to the number of days?

Enter your answer in the box.

Constant of Proportionality =

2. What was the total number of videos taken down over the course of one week?

Enter your answer in the box.

accessibility and cultural relevance, reinforcing the principle that context should be both clear and affirming for all learners. This structured and reflective approach to item development ensures that culturally responsive assessments are aligned to rigorous standards and representative, humanizing, and affirming for the diverse populations they aim to serve.

An exploratory approach to maximizing item writing efficiency is to leverage a large language model (LLM), such as ChatGPT, Latimer, Co-Pilot, or Google Gemini. In this process, writers would begin by entering the grade level of the item, the context or scenario, the text of the targeted standard, the targeted culturally responsive quality, the DOK level, the difficulty level, the item type (e.g., multiple choice, free response, drag and drop...etc.), and any other item specifications desired. For example, a nuanced interpretation of “family” may consider additional complexities, such as multigenerational or elderly individuals, that challenge U.S.-centric notions of nuclear family structures. Assessment writers should also prompt the LLM to ensure that item types include distractors that reflect common student errors and/or misconceptions of the given measured standard.

While AI can streamline item development, its use must be intentional and critically monitored. AI-generated outputs should not be accepted “as is.” Instead, they undergo rigorous review for accuracy, alignment with standards, cultural authenticity, bias, and overall clarity. An example of what a writer might enter into the LLM to produce an initial version of the item, knowing that critical oversight of AI outputs is necessary:


Create a 7th-grade, multiple-choice math assessment item aligned to this standard: 7.RP.A.2c Represent proportional relationships by equations. Use the context of Mina having a henna booth at the school fair to help raise money for the school's international student organization. Make it a DOK2, difficulty level medium, and align it to the Geneva Gay quality of transformative (Culturally responsive assessment facilitates the development of students' cultural consciousness to empower them as social critics and agents of change.) Make sure that the distractors reflect common student errors or misconceptions. Create a picture of Mina in a hijab putting henna on her classmate.

The LLM-generated item (i.e., ChatGPT) serves as a starting point, the reviewed and revised item is shown in Figure 3. When using AI to create images, one key consideration is requiring critical human review after the AI generates a cultural image, since AI algorithms can be biased and reflect and reproduce images based on the data they were trained on (e.g., Alervo et al., 2024; Marinucci, 2023). As a result, the use of AI requires a critical review, and if necessary, revisions to ensure cultural accuracy and sensitivity.

An important consideration in CRA is the role of youth culture as a meaningful context that directly shapes how students engage with assessments. In developing the items referenced in our approach, many

Figure 3. Example Initial Item Generation Using ChatGPT

Mina is running a henna booth at her school's fair to raise money for the International Student Organization. She charges the same amount for each henna design. After 3 hours of work, she has earned \$60 from 10 henna designs.



Which equation best represents the relationship between the number of henna designs, d , and the amount of money Mina earns, m , if the relationship is proportional?

- $m = 10d$
- $m = 6d$
- $m = 6 + d$
- $m = 60d$

item writers were youth, most under 20. Their lived experiences and peer-informed insights added valuable layers of authenticity and relevance to the content. While it is not always feasible or necessary for all CRA items to be written by youth, involving young people in the development process as “co-writers,” “co-reviewers,” and/or “co-collaborators” strengthens the alignment between assessment content and students’ real-world experiences. Including these perspectives helps ensure assessments are not only culturally responsive but also meaningful and engaging for the students they aim to serve.

Phase V: Iterative Item Review and Revision

The item writing review process begins once the item writing is complete. This involves a thorough examination of the item’s content, clarity, and adherence to initial guidelines and specifications. Reviews typically occur internally within the development team and externally with equity experts (e.g., scholars, researchers) to ensure that the items embody the targeted culturally responsive quality (e.g., empowering, validating, multidimensional, etc.). Multiple rounds of discussion, consensus building, and review can be conducted until the item aligns with the CR qualities. It is important to note that an item can align with more than one CR quality, as they are not always mutually exclusive. For example, an item might demonstrate both “Empowering” and “Validating” if it asks students to analyze a social issue relevant to their community (i.e., Empowering) and then propose solutions that draw on their cultural knowledge and values (i.e., Validating).

Along with a concrete understanding of the CRA's qualities, “authenticity reviews” with individuals whose cultural context aligns with the item are necessary to avoid deficit-based narratives and negative stereotyping of a particular marginalized group, and are a critical component of this phase. Authenticity reviews are conducted when the item writer and/or reviewer are not from the culture represented in the item, or when cultural specificity warrants additional confirmation. These reviews delve deeply into the cultural nuances of the item, scrutinizing the chosen contexts, characters, and scenarios to ensure they are portrayed respectfully and accurately.

The final steps (e.g., fact-checking, key verification, and copyediting) follow standard practices in any assessment development. This includes verifying the accuracy of all factual information presented in the item, meticulously confirming the correctness of the answer key, and conducting thorough copyedits to ensure grammatical precision, consistent style, and overall linguistic clarity. To strengthen cultural alignment, we encourage conducting additional CR alignment studies with finalized items and experts in cultural responsiveness, as well as with the target rights-holders (e.g., students, parents, teachers), to ensure that the

finalized items target the intended culturally responsive quality or qualities. Item writers are also encouraged to make any necessary adjustments to the items' content based on the feedback on the alignment studies.

Key Considerations

Implementing this approach requires attention to several critical considerations. First and foremost, self-reflection is an ongoing pursuit required before engaging in item writing. Interest-holders and rights-holders must engage in both individual and collective reflection to deepen their awareness of positionality. Authentic community support is necessary for identifying and shifting limiting beliefs, as we cannot fully recognize our own biases in isolation. Reflection entails acknowledging intersectional identities, biases, and how personal worldviews shape our understanding of others, especially individuals and groups more directly impacted by the assessment. Developers are further removed from this direct impact, but often hold the most decision-making power over what gets measured. As a result, self-reflection is crucial for fostering transparency, which builds trustworthy, authentic, and equitable collaborations.

The second consideration is the commitment to and embracing of collaborative methodologies. Many approaches to collaboration require qualitative and/or action research methods (e.g., participatory action research). Like assessment developers, interest-holders and rights-holders must be prepared to design and employ qualitative methods such as in-depth cognitive interview protocols, focus groups, or think-aloud protocols. These skills are perceived as auxiliary compared to core competencies of instrumentation, generalization, and psychometric modeling (e.g., Ackerman et al., 2023). Examples of qualitative skills include collecting, organizing, analyzing, combining, and identifying patterns in data (e.g., open-ended survey responses, transcripts). After understanding the necessity of qualitative data and the capacity to synthesize and analyze it accurately, consideration should be given to training facilitators in qualitative methods or collaborating with specialized experts, particularly in Phases 0 and I.

Third, developing and refining CRA items requires significant time and resources; thus, a key consideration is the time and resources to train co-designers. One key training could focus on effective collaborative approaches during brainstorming scenarios and on interpreting them to preserve cultural authenticity. The CRA training often includes collaborative item development, such as sharing cultural scenarios, and creating open, inclusive spaces that value all perspectives. These experiences support a deeper understanding of CRA qualities, which are inherently subjective and shaped by individuals' unique backgrounds and lived experiences. For example, what one person may perceive as "Validating" item content may be seen as "Humanistic" by another, depending on their lived experience. Acknowledging and being flexible to this subjectivity is crucial when discussing CRA quality alignment, as it allows a more equitable approach to item development.

Another aspect influenced by time and resources is real-world development constraints, including state buy-in, funding, and grants, to ensure both required standards alignment and the validity of the cultural relevance of item content. While the validity of culturally relevant scenarios is important, it is important to acknowledge that traditional item content is already culturally misaligned for marginalized students (Randall et al., 2023; Sul, 2019). While these constraints have real implications for CR item development, they are important to navigate to deliver on a CRA approach that enables all students to demonstrate academic growth.

Lastly, depending on the assessment's context and the systems it operates within, a key consideration is balancing cultural authenticity with technical requirements. The context, such as whether the assessment is for classroom use, district-level benchmarking, or high-stakes state testing, shapes what is feasible and how cultural elements can be integrated. For example, large-scale standardized assessments often impose strict constraints on item length, readability, and psychometric properties, thereby limiting the depth of cultural

scenarios. In contrast, classroom-based assessments may allow for richer narratives and more flexible formats. Additional considerations should include (a) appropriate grade-level vocabulary, (b) cognitive and reading loads, and (c) links between the cultural scenarios and the mathematical concepts. One way to support the inclusion of both context and guidelines is by incorporating images that illustrate key points. Research shows that images in assessment items serve as important semiotic resources that help examinees interpret and make sense of item content, a finding that supports the inclusion of images to visually solidify student understanding of cultural contexts (Solano-Flores, 2019).

Implications

This practical approach to co-designing a math CRA with both interest-holders and rights-holders reveals several key implications. First, despite requiring substantial resources, including time, funding, and managerial oversight, co-developing assessments ensures that diverse communities actively contribute to the design of the assessments. Recruiting, training, and compensating rights-holder co-designers requires planning and adequate funding before conducting brainstorming sessions.

Another key implication is the importance of building diverse cultural representations. Brainstorming sessions with rights-holders (i.e., students) help build a repository of culturally relevant scenarios that item writers can reference when writing culturally relevant items. This can also be achieved through LLMs and AI image generators that incorporate human oversight. Language models and AI image generators offer efficiency in content creation; therefore, AI is best suited for ideation, with necessary additional human oversight to ensure accuracy and authenticity while protecting privacy. The resource-intensive nature of this approach is a strength, as it triangulates data from multiple sources rather than relying on a single source.

Lastly, it is important to recognize that there are traditions of assessment development grounded in culturally responsive or justice-oriented frameworks that are currently being written, tested, and practiced today (e.g., Badrinarayan et al., 2025; Forzani et al., 2024). Traditional assessment practices have benefited from over a century of iterative refinement and institutional support (Araneda, 2025). As a relatively new endeavor, embedding the cultural responsiveness that centers historically marginalized populations into the development process requires greater time, resources, and intention. Our practical approach illuminates assessment development areas where extensive effort is considered key, including the brainstorming phase, targeting CR qualities, authenticity review, and fact-checking. We remind readers that, despite introducing greater nuance, it is still critical to balance with traditional test specifications (e.g., standards, difficulty level, depth of knowledge).

Conclusions

In this article, we illustrate a collaborative item-writing process that seeks to foster an inclusive assessment environment, redistributing the initial power held by interest-holders to actively involve rights-holders. Our approach highlights the importance of sustained investment in development training, qualitative data analysis, and collaborative environments. Building capacity among interest-holders and rights-holders creates opportunities for further refinement and transferability across sociocultural contexts, thereby ensuring that assessments remain relevant as society evolves.

Future research can measure student engagement and achievement using both traditionally created and CRA assessments, implementing this practical approach in the U.S. and internationally. The initial conceptualization and insights of this approach are rooted in the development of CRAs within the U.S.-centric contexts. Thus, there is an opportunity for future research to explore the transferability and adaptability of this approach in international settings, where sociocultural contexts uniquely shape

assessment, conception, usage, and refinement. Extensions of this work can modify and illuminate universal principles of cultural responsiveness for diverse cultural groups worldwide. This ongoing refinement and international application will be crucial in pursuing more

Received: 6/17/2025. **Accepted:** 3/15/2026. **Published:** 5/1/2026.

Citation: Hamdani, M., Vo, T., & Julien, N. A. (2026). A practical approach in co-designing culturally responsive mathematics assessments. *Practical Assessment, Research, & Evaluation*, 30(2)(10). Available online: <https://doi.org/10.7275/pare.3258>

Corresponding Author: Maria Hamdani. Email: hamdani.maria@gmail.com

References

- Araneda, S. (2025, April). *100 years of criticisms against standardized testing* [Conference presentation]. NCME 2025 Annual Meeting.
- Alvero, A. J., Lee, J., Regla-Vargas, A., Kizilcec, R. F., Joachims, T., & Antonio, A. L. (2024). Large language models, social demography, and hegemony: comparing authorship in human and synthetic text. *Journal of Big Data*, 11(1), 138. <https://doi.org/10.1186/s40537-024-00986-7>
- Badrinarayan, A. (2024). Instructionally relevant assessment systems: What is the role of performance assessments? Learning Policy Institute. <https://doi.org/10.54300/411.469>
- Englert, K., & Schultz, P. (2025). Reconsidering theories of action using cultural and community validity to transform assessment development, implementation, and use. In *Culturally responsive assessment in classrooms and large-scale contexts* (pp. 157–173). Routledge. <https://doi.org/10.4324/9781003392217-12>
- Forzani, E., Corrigan, J., Slomp, D., & Randall, J. (2024). Prioritizing equitable social outcomes with and for diverse readers: A conceptual framework for the development and use of justice-based reading assessment. *Educational Psychologist*, 1-24. <https://doi.org/10.1080/00461520.2024.2418400>
- Frierson, H., Hood, S., & Hughes, G. (2002). *A guide to conducting culturally responsive evaluation*. In *The 2002 user-friendly handbook for project evaluation* (pp. 63-73). National Science Foundation.
- Gay, G. (2018). *Culturally responsive teaching: Theory, research, and practice*. Teachers College Press.
- Gary, A. & Holmes, D. (2020). Researcher Positionality—A Consideration of Its Influence and Place in Qualitative Research—A New Researcher Guide. *Shanlax International Journal of Education*, 8(4), 1–10. <https://doi.org/10.34293/education.v8i4.3232>
- González, N., Moll, L. C., & Amanti, C. (Eds.). (2006). *Funds of knowledge: Theorizing practices in households, communities, and classrooms*. Routledge.
- Hamdani, M. & Julien, N. (2025, April). *Re-humanizing Assessment: A Justice-Oriented Framework for Culturally Responsiveness*. [Panel Session]. National Council on Measurement in Education Annual Meeting, Denver, CO.
- Horowitz, J. M. (2019, May 8). Americans See Advantages and Challenges in Country's Growing Racial and Ethnic Diversity. *Pew Research Center's Social & Demographic Trends Project*. <https://www.pewresearch.org/social-trends/2019/05/08/americans-see-advantages-and-challenges-in-countrys-growing-racial-and-ethnic-diversity/>
- Hood, S. (1998). Culturally responsive performance-based assessment: Conceptual and psychometric considerations. *The Journal of Negro Education*, 67(3), 187–196. <https://doi.org/10.2307/2668188>

- Julien, N., Hamdani, M., Cruz, S. (2026, April). *Centering Rights-Holders From Co-Design To Identifying Impact in K-12 Assessments* [Panel Session]. National Council on Measurement in Education Annual Meeting, Los Angeles, CA.
- Keehne, C. N. K., Kelling, I. K., & Awo-Chun, K. (2025). Manifesting the mana of culturally responsive assessments. In *Culturally responsive assessment in classrooms and large-scale contexts* (pp. 265–283). Routledge. <https://doi.org/10.4324/9781003392217-19>
- Lee, C. D. (1998). Culturally responsive pedagogy and performance-based assessment. *Journal of Negro Education*, 268-279.
- Ladson-Billings, G. (1995). Toward a theory of culturally relevant pedagogy. *American Educational Research Journal*, 32(3), 465-491.
- Marinucci, L., Mazzuca, C., & Gangemi, A. (2023). Exposing implicit biases and stereotypes in human and artificial intelligence: State of the art and challenges with a focus on gender. *AI & Society*, 38, 747–761. <https://doi.org/10.1007/s00146-022-01474-3>
- Middleton, K. V., & Haynes, J. A. (2025). Implications of culturally responsive education for classroom assessment practices. In *Culturally responsive assessment in classrooms and large-scale contexts* (pp. 245–263). Routledge. <https://doi.org/10.4324/9781003392217-18>
- Milner, R. (2007). Race, Culture, and Researcher Positionality: Working through Dangers Seen, Unseen, and Unforeseen. *Educational Researcher*, 36(7), 388–400.
- Moss, P. A., & Schutz, A. (2001). Educational Standards, Assessment, and the Search for Consensus. *American Educational Research Journal*, 38(1), 37–70. <https://doi.org/10.3102/00028312038001037>
- Noddings, N. (2012). *The caring relation in teaching*. Oxford Review of Education, 38(6), 771-781
- Paris, D., & Alim, S. (2014). What are we seeking to sustain through culturally sustaining pedagogy? A loving critique forward. *Harvard Educational Review*, 84(1), 85–100. [doi:10.17763/haer.84.1.9821873k2ht16m77](https://doi.org/10.17763/haer.84.1.9821873k2ht16m77)
- Poe, M., Oliveri, M. E., & Elliot, N. (2023). The Standards Will Never Be Enough: A Racial Justice Extension. *Applied Measurement in Education*, 36(3), 193–215. <https://doi.org/10.1080/08957347.2023.2214656>
- Randall, J. (2025). Seeking Justice-Oriented Assessment: Exploring the Impact of Culturally Responsive and Antiracist Content. *Educational Assessment*, 1–30. <https://doi.org/10.1080/10627197.2025.2488452>
- 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. <https://doi.org/10.1002/jaal.1142>
- Randall, J. (2023). It ain't near 'bout fair: Re-envisioning the bias and sensitivity review process from a justice-oriented antiracist perspective. *Educational Assessment*, 28(2), 68-82. <https://doi.org/10.1080/10627197.2023.2223924>
- Randall, J. (2021). “Color-neutral” is not a thing: Redefining construct definition and representation through a justice-oriented critical antiracist lens. *Educational Measurement: Issues and Practice*, 40(4), 82–90. [doi:10.1111/emip.12429](https://doi.org/10.1111/emip.12429)
- 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](https://doi.org/10.1080/10627197.2022.2042682)

- Randall, J., Poe, M., Oliveri, M. E., & Slomp, D. (2023). Justice-Oriented, Antiracist Validation: Continuing to Disrupt White Supremacy in Assessment Practices. *Educational Assessment*, 1–20. <https://doi.org/10.1080/10627197.2023.2285047>
- Russell, M. (2023). *Systemic racism and educational measurement: Confronting injustice in testing, assessment, and beyond*. Routledge.
- Saporito, M. (2020). Factors Contributing To The Overrepresentation Of Minorities In Special Education. *The Interdisciplinary Journal of Advances in Research in Education*. <https://doi.org/10.55138/ab104284fms>
- Schettino, I., Radvany, K., & Wells, M.S. (2019, Sept. 23). Culturally responsive education under ESSA: A state-by-state snapshot. *Phi Delta Kappan*, 101 (2), 27-30.
- Solano-Flores, G. (2019). Examining cultural responsiveness in large-scale assessment: The matrix of evidence for validity argumentation. *Frontiers in Education*, 4, 43. <https://doi.org/10.3389/feduc.2019.00043>
- Sul, D. A. (2019, March). Reclaiming educational autonomy and minimizing measurement disjuncture through a culturally specific assessment development process [Conference Presentation]. *Culturally Responsive Evaluation and Assessment (CREA) Conference*, Chicago, IL.
- Taylor, C. S., & Ferrara, S. (2025). A measurement argument for culturally responsive assessment. In *Culturally responsive assessment in classrooms and large-scale contexts* (pp. 135–153). Routledge. <https://doi.org/10.4324/9781003392217-11>
- Walker, M. E., Olivera-Aguilar, M., Lehman, B., Laitusis, C., Guzman-Orth, D., & Gholson, M. (2023). Culturally responsive assessment: Provisional principles. *ETS Research Report Series*, 2023(1), 1-24. <https://doi.org/10.1002/ets2.12374>
- White, L., Nesbitt, J., Roeters-Solano, H., Quesen, S., Lottridge, S., & Lochbaum, K. (2025). Antiracist approaches to scoring large-scale assessments. In *Culturally responsive assessment in classrooms and large-scale contexts* (pp. 203–221). Routledge. <https://doi.org/10.4324/9781003392217-14>
- William, D. (2011). What is assessment for learning? *Studies in educational evaluation*, 37(1), 3-14. <https://doi.org/10.1016/j.stueduc.2011.03.001>
- United Nations. (n.d.) *Universal declaration of human rights*. <https://www.un.org/en/about-us/universal-declaration-of-human-rights>
- U.S. Department of Education. (n.d.). *Assistance to states for the education of children with disabilities: Child find* (34 C.F.R. § 300.111). Individuals with Disabilities Education Act. <https://sites.ed.gov/idea/regs/b/b/300.111>