


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

Exploring Cognitive and Emotional Processing of Culturally Responsive and Antiracist Test Items

Chris Patterson, *University of Iowa (Current affiliation: Knox College)* 

Abstract: Culturally responsive and antiracist (CRA) test items have been discussed as a way to increase validity of test scores for students of color, accounting for the ways culture and cognition intersect while encouraging the field of educational measurement to be inclusive and antiracist. Given these premises, CRA items discuss culture, race, and antiracism, which some are concerned may cause cognitive or emotional disturbance in test takers. The current study explored these concerns through a reflexive thematic analysis of 20 cognitive interviews where college students interacted with CRA items. Results of this study show promise for CRA items via minimal negative emotional disturbance and actionable direction on how to avoid cognitive disturbance.

Keywords: Culturally-responsive assessment, Cognitive interviews

Introduction

Multiple-choice assessments have been criticized as biased against people of color when created through the lens of the Standards of Educational and Psychological Measurement (Standards; American Educational Research Association [AERA] et al., 2014). Critiques of using the Standards and other popular research frameworks to create multiple-choice items include centering whiteness in defining constructs (Randall, 2021; Randall et al., 2022), lack of examples that apply to the lives of test takers of color (National Council for Measurement in Education [NCME], 2022), and the lack of acknowledgement of culture-cognition relationships while validating results (Hood et al., 2005; Hall, 1992; Randall, 2021; Randall et al., 2022). Culturally responsive (CR) and antiracist assessments address these concerns through creating frameworks for item development (NCME, 2022; Patterson, 2025; Randall, 2022) and validation (Randall et al., 2022) that integrate cultural responsiveness and antiracism into their processes. CR assessments hold items that aim to represent the cultures of test takers, while incorporating micro interventions that allow test takers to learn more about their and others' cultures (NCME, 2022). Antiracist assessments differ slightly, having

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.

goals of challenging internally held notions of racism through items that explicitly discuss manifestations of racism in global or local contexts (Randall, 2021).

As discussions on cultural responsiveness and antiracism have recently mainstreamed in the field of educational measurement, there are questions as to how test takers will react to CR and antiracist items. As the philosophy and structure of such items have been highly discussed, little research exists on how test takers engage with CR and antiracist items. While some believe that test takers who interact with CR and antiracist items will be more accurately scored, others believe the contexts woven into these items may interrupt or distract test takers' cognitive focus ("cognitive barriers") and induce emotional reactions that weaken the validity of results. As a result, more research needs to be done to understand test takers' response processes on CR and antiracist items. The following sections will challenge typical item development practices through discussing a) the need for contextual information in multiple-choice items, and b) the advantages of including CR and antiracist contexts into items. After a brief discussion on stated concerns with the inclusion of inclusive contexts, the current study is presented.

Culture, Cognition, and Item Response

When taking a test, the first step in answering any item is interpreting what the item is asking the test taker to do (Sudman et al., 1996). Guidelines for best practices in item construction, such as the Standards and research on item construction (e.g., Haladyna & Rodriguez, 2013), emphasize clarity to ensure items are understood as intended. These recommendations typically include omitting contexts, stories, or examples to keep items short and accessible (Haladyna & Rodriguez, 2013; Zieky, 2016). Implicit in this guidance is the assumption that decontextualized items are more accessible and therefore heighten validity arguments. However, research on cognition and information processing through the lenses of race and culture challenges this assumption.

Early scholarship following racial integration of schools in the United States hypothesized that observed achievement differences reflected cultural misalignment between classroom expectations and students' lives experiences, not inherent differences in ability (Au & Jordan, 1981; Brown-Jeffy & Cooper, 2011). Many racially and culturally marginalized communities emphasize relationality and contextual reasoning, like consideration of communal factors in decision making (Gay, 2018; Spring, 1995). Students socialized into these communities often connect contextual information to facts, making meaning of events through both the event and its contextual factors (Hall, 1992). In contrast, educational norms in the U.S. privilege forms of information processing that are framed as neutral or objective. Importantly, this orientation should not be understood as a natural or preferential (or even unique to white students), but rather a product of socialization into whiteness. Whiteness in educational measurement is often positioned as socially neutral, rendering its cultural assumptions invisible while other forms of meaning making are marked as subjective or inappropriate (Dixon-Román, 2020). As a result, contextualized reasoning may be seen as extraneous, as it departs from dominant ideas of objectivity in information processing.

From this perspective, the issue is not whether certain groups of students "learn best" with or without context. Rather, assessment items designed to appear context-free may rely on culturally dominant assumptions that align more with those already centered in the educational system—white students. Prior scholarship has suggested that test takers who process information more contextually may be disadvantaged by items created from guidelines emphasizing decontextualization, due to misalignment between item design and culturally patterned item response processes (Ibarra & Cohen, 1999; Santelices & Wilson, 2010).

As a result, tests created with the guidance of having less context in multiple-choice items may possess weakened validity arguments when item response processes, are considered in relation to culture and cognition (Randall, 2021). Incorporating contextual information into multiple-choice items may actually support cognitive processing of multiple-choice items; a recent study (see Sinharay, 2025) incorporating

cultural responsiveness into multiple-choice items noted a reduction in score gaps without evidence of construct distortion, suggesting contextual alignment with test takers' lives can strengthen validity by improving item response process and not altering the construct being measured.

Newer validity frameworks like Randall et al.'s (2022) antiracist, justice-oriented validity framework emphasize that test items should positively activate the cognition pathways of all test takers. To do this, multiple-choice items should include contextual information. The Standards allow for the use of contexts or examples, but caution that they should be broadly understandable and perceived as "socio-culturally neutral." Yet, so-called neutral contexts frequently reflect dominant cultural experiences and may be foreign or boring by some test takers. Boring and foreign contexts can be harder to cognitively process (Bennett, 2023; Walker et al., 2023), resulting in biased test results.

The use of real-life, applicable contexts is well established in culturally relevant pedagogy (CRP; Ladson-Billings, 1995; 2014) and culturally responsive teaching (CRT; Gay, 2018), where instruction draws on students' everyday lives and cultural communities. When used, CRP and CRT facilitate higher outcomes achievement in all students due to increased student engagement (Aronson & Laughter, 2016). Applied to testing, engagement with item contexts that reflect students' experiences may enhance accuracy and interpretability of test results. CR and antiracist item designs take this into consideration, calling for contexts that explicitly relate to test takers' local communities (NCME, 2022; Patterson, 2025, Randall, 2021). Further, test takers at the K-12 (NCME, 2022) and postsecondary levels (Patterson, 2025) have shown interest in the context shown in CR and antiracist items.

Racial Identity and Emotional Response

Including relatable contexts into multiple-choice items may allow test takers to understand their own and others' communities (NCME, 2022; Patterson, 2025). Further, CR and antiracist items place these contexts as direct reflections of test takers and society. According to popular item-writing guidelines, these inclusive items should be avoided due to their divisive nature; they may induce emotional triggers within test takers, contributing to construct-irrelevant variance via item response processes. According to the *Standards*, this argument would be used to negate the use of CR and antiracist items on tests, as construct-irrelevant variance weakens validity arguments. However, answering different questions may lead us to think otherwise. Instead of assuming inclusive item topics affect everyone to a serious degree, we should ask: *Who* is affected when answering CR and antiracist items? Further, what level of affect is present when processing these items? A glimpse of these answers lies in knowing how socialization into race forms one's reactions to race-based stimuli.

Research in racial identity development frequently notes how white people do not deeply understand ourselves¹ as racial beings (Abaied & Perry, 2021). In educational testing, "best practices" of validation processes were created using research and ideas that centered white students, which was then projected onto all students using a sweeping argument of objectivity (Dixon-Román, 2020). These best practices also include the argument that all test takers cognitively and emotionally process test items similarly. A prime example of this is found within the *Standards*, where offensive or disturbing test content poses as a threat to validity (AERA et al., 2014). When this guideline is operationalized, CR and antiracist items are likely to be flagged by panels due to panel members' perceptions that inclusive topics are naturally offensive, divisive, or disturbing, even if the topic is not seen that way by test takers (Patterson, 2025). However, topics of race, racism, and culture may be processed differently by different racial and cultural groups in the United States.

¹ When talking about white people and whiteness, I use identifiers that signify my connection to whiteness. It is important for white people to acknowledge our connections to whiteness (Gillborn, 2006), including whiteness in educational measurement.

When emotionally processing and responding to topics related to race and racism, there is typically a stark division between white people and people of color. When white people are exposed to a stimulus related to racism, we are likely to feel challenged by the information, creating cognitive dissonance (Watt, 2015). (Sub)consciously, then, white people are likely to respond defensively in ways that protect us from having to reckon with our involvement in systematic oppression (Watt, 2015; Watt et al., 2021). This process stems from white peoples' socialization into thinking of ourselves as non-racial beings (Spanierman & Cabrera, 2014), meaning the resulting dissonance comes from being told "you as a white person are a racial being that contributes to racism." In testing, this sort of dissonance may lead a white test taker to divert energy from answering an item to calming their emotional state before continuing with their test. This is likely what literature refers to as an emotional barrier to construct-irrelevant variance (Zieky, 2016).

When observing the response of people of color to discussions about race and racism, there may be a different sort of emotional response. Given people of color (POC) are socialized to think of themselves as racialized and "othered" beings through their experiences in living in and through racism (Molla, 2024), one-time exposure to topics of race and racism may not induce the same emotional dissonance or defensive instinct. However, repeated exposure to certain discussions can invoke negative responses through either mentally reliving the topic of discussion (e.g., racial violence) or knowing someone who was affected. Called linked fate (Gee et al., 2012; Monk Jr., 2020), indirect exposure can induce stress, anger, or other negative emotions in POC. However, linked fate has not been studied in educational testing. On the other hand, in certain structured spaces like culturally responsive classrooms, POC feel empowered to participate and engage in class (Cholewa et al., 2014), lowering negative emotional disturbance. Therefore, there may be minimal negative emotional reactions, along with more positive ones, if CR and antiracist tests are used and structured well.

Through this lens, literature claiming that CR and antiracist topics should not be included in tests due to emotional disturbance may be (un)intentionally centering white test takers' emotional well-being. In ignoring how POC may process race-based stimuli, not including CR and antiracist items on tests may be upholding whiteness through assuming that everyone responds in the same way white people do, setting white values as the norm. Therefore, the omission of CR or antiracist items may perpetuate further oppression toward POC, especially in high stakes testing environments.

In total, CR and antiracist items fulfill two well-cited needs of tests for stronger validity arguments: (1) these items have contexts, which allow test takers to be tested at a higher cognitive stage than just defining something; and (2) the contexts are relatable to test takers' lives and communities, where the representation and discussion of certain topics may lead to heightened test engagement. Together, CR and antiracist items have the potential to more accurately measure test takers' abilities. These items also provide micro-interventions that can subtly teach test takers about the world around them (Patterson, 2025, Randall, 2021). It may seem like CR and antiracist assessments are the obvious future for tests through the lens of validity. However, parts of the educational measurement community hold concerns toward such items, which should be empirically explored.

Concerns and Purpose of Study

Although CR and antiracist multiple-choice items are argued to increase the quality of validity arguments, there are concerns from the broader educational measurement community about implications for various test takers. These concerns include the potential to emotionally and cognitively disrupt test takers in various ways, explained further below. As many of these concerns have not been discussed empirically, literature from other areas of research, such as sociology, are used to further discussion here.

The primary concern with CR and antiracist items is, as discussed briefly above, their potential to disrupt test takers through introducing discomfort in reading about culture, race, and racism. This concern is mostly

pointed at white test takers, given that white people have historically reacted negatively toward intentional racial inclusion in other areas of life, for example in news media (e.g., Levy & Myers, 2020). However, repeated exposure to such items may lead POC to also react in a negative way, as a past study has found that POC want to be authentically represented in tests (Patterson, 2025). The lack of authentic representation (e.g., reductive stereotypes) may induce emotional disturbance, lowering the quality of validity arguments.

Although most concerns toward CR and antiracist items are emotional, there are some concerns in how test takers may cognitively process these items. Introducing inclusive contexts involves writing about cultural tenets, racism, and antiracism. Test takers who are unfamiliar with the topics may feel that knowing about them would allow the test taker to answer the item correctly, a classic example of construct-irrelevant variance (Zieky, 2016). However, using critical frameworks such as culturally relevant pedagogy to inform item development should result in contexts that are relevant to test takers' lives and communities, even when those contexts are not uniformly familiar to all individual students. Further, the addition of CR and antiracist contexts may lengthen the item stem or response options to the point where test takers become lost while reading the item, resulting in exerting more cognitive effort toward figuring out what the item is asking, and less effort toward solving the item.

With the rapid rise in advocating for CR and antiracist multiple-choice items, some are concerned about how such items might operate. However, these concerns have not been extensively empirically explored. The current study aims to explore these cognitive and emotional concerns through answering two research questions (RQs):

RQ1. What cognitive barriers, if any, do test takers experience when answering CR and antiracist multiple-choice items?

RQ2. What emotions, if any, do test takers experience when answering CR and antiracist multiple-choice items?

Method

Sample

The sample consists of 20 students enrolled in an R2 university in the Mid-Atlantic region of the United States, recruited via maximum variation sampling until saturation of codes and themes occurred. Participants self-enrolled in the study via a campus website and were incentivized through their own course instructors. The sample varied in racial and gender identities and were between 18-21 years of age. Information on participants' identities, including aliases used for anonymity, is in Table 1.

Items

The items used in this study were constructed by Patterson (2025), which consist of 10 diversity-infused (DI) and 10 sociopolitical consciousness (SPC) items, revised from a test of information literacy (Clarke & Radcliff, 2018). DI items are CR items focused on building cultural competence within test takers, showcasing diverse cultural practices. SPC items are a type of antiracist item which discusses the manifestations of antiracism on local levels to raise test takers' sociopolitical consciousness (Patterson, 2025).

Data Collection

Data collection occurred via semi structured cognitive interviews (Leighton, 2017) in Fall 2023. Participants were randomly sorted to engage with 1 of 2 interview types: all DI or all SPC items. In each interview, participants were first introduced to the study and signed a consent form. After, participants were given verbal and written instruction on how to participate in a cognitive interview. The author, as interviewer,

Table 1. Sample characteristics and items interacted with

Name	Race	Gender	Type of Items Interacted with
Adam	White	Man	Diversity-Infused
Alaina	White	Woman	Sociopolitical Consciousness
Alice	Asian American	Woman	Sociopolitical Consciousness
Ally	White/Caucasian	Woman	Diversity-Infused
Audrey	Latina	Woman	Sociopolitical Consciousness
Bella	White	Woman	Sociopolitical Consciousness
Blue	Caucasian	Woman	Diversity-Infused
Calvin	White	Cis Man	Sociopolitical Consciousness
Crystal	Black & White	Trans Man	Diversity-Infused
Ella	White	Woman	Sociopolitical Consciousness
Gigi	African American & Mexican	Woman	Diversity-Infused
Jane	White	Woman	Sociopolitical Consciousness
John	African American	Man	Diversity-Infused
Kelly	Hispanic	Woman	Diversity-Infused
Mary	Latina	Woman	Sociopolitical Consciousness
Mila	Latina & Middle Eastern	Woman	Diversity-Infused
Nyx	Black	Woman	Sociopolitical Consciousness
Phillip	White	Man	Diversity-Infused
Skyla	Asian American	Woman	Sociopolitical Consciousness
Susan	White	Woman	Diversity-Infused

Note: Race and gender identities are self-reported

then showed the participant, by example, how to answer a question via the think aloud method. The participant then practiced the think aloud method on a different practice question. Data collection then started once the participant felt comfortable, where the author led the participant in answering 10 items. After answering each item, the author asked probing questions as necessary to gather further information on cognitive and emotional reactions to answering the item. After all 10 items were worked through, the author then asked the participant questions about their feelings toward the 10 items as a set, gathering more information about cognitive and emotional reactions. All interviews were recorded via Zoom, transcribed by a third-party private company, and analyzed via NVivo version 13 (Lumivero, 2020). Further specifics on the interview process are noted in the supplemental materials.

Data Analysis

Reflexive thematic analysis (RTA) is a method that holds flexibility in its theoretical orientation, taking many different forms to answer different qualitative research questions (Braun & Clarke, 2019). This method relies on dedicated reflexivity of a single data analyst, consistently reflecting on their assumptions, data, and positioning (Braun & Clarke, 2020). As a result, assumptions must be stated prior to analysis to ground both the researcher and reader in understanding results. Byrne (2022) describes each of these assumptions in depth.

In this study, I took a constructionist epistemological approach, noting the need to weave in meaningful narratives, regardless of the frequency in which they appeared, into results (Byrne, 2022; Braun & Clarke,

2006). Next, I assumed an experiential orientation to data, where my interpretation of data focused more on how participants as a collective were processing cognitively and emotionally, and not trying to interpret results through a critical theory lens (Byrne, 2022). Next, a balanced semantic-latent approach was used, where codes were interpreted both at “face-value” (semantic) while some were double-coded for a deeper meaning (latent). Finally, this specific analysis utilized both an inductive and deductive process, using both open and a priori coding to interpret data.

The process of data analysis followed a popular process outlined by DeCuir-Gunby et al. (2011). First, previous literature was surveyed on cognitive and emotional processing of items to gather a priori codes. After data had been collected, RTA was employed through a flexible coding process, where participants’ narratives could fall into a previously created code, or fall into a new code completely (an open code). After the first read-through, I interrogated the codebook by adding, subtracting, or combining codes, journaled, then coded all transcripts again using the updated codebook. This cyclic process continued until codes were applied in ways where codes were believed to be representative of the narratives that fell underneath each code. After, codes were then arranged into themes, where codes with conceptual connections were grouped under the same theme, and codes that differed from one another were placed under different themes. This process was also cyclic, where I also interrogated thematic structure and rearranged codes as necessary, where themes accurately represented the narratives that comprised them. During this process, data were horizontally weighed, meaning no one participants’ narrative was seen as more important than others’ narratives, and the number of times a narrative occurs is irrelevant; it is more important to notice that experiences are being had.

In using both a priori and open coding processes, original codes were generated after reading through literature on manifestations of cognitive barriers and emotional reactions in testing. After sifting literature, 4 codes were generated for cognitive barriers (item length, too much information, too little information, and jargon), and 3 were generated for emotional reactions (offended, disturbed, bored). Note that due to the coding process, not all of these codes appear in results.

Researcher Positioning

Race-based studies in educational measurement necessitate researchers’ acknowledgement of how any study is an extension of their values (Dixon-Román, 2016), stemmed from their socialization into race, gender, and other intersecting systems. Therefore, an identity statement does not serve the purpose of objectivity but rather gives the reader insight into how my identities and knowledge thereof have affected the setup, execution, and conclusions of the current study.

I am a white, queer, neurodivergent man who has spent most of my career in academia. My relationship with race in educational measurement poses that my way of thinking and existing is the norm and has been rarely questioned in the history of educational measurement. As such, I acknowledge this study is an example of me using my privilege to enter conversation, while attempting to use that space to question and de-center white people and whiteness in educational testing.

Rigor and Trustworthiness of Results

Due to the flexible nature and interpretive spirit of RTA, actions must be performed to ensure rigor and trustworthiness of data collection and analysis, as well as presentation of results (Krefting, 1990; Levitt et al., 2017). Credibility was ensured through member check-ins both after initial transcription and in early drafts of results. Peer-examination was done via multiple researchers reading the codebook, some coded transcripts, and results. An audit trail was also kept through record-keeping of codes and explanations of the creation of new codes, deletion of old codes, and the combining of certain codes. Finally, in the spirit of RTA, reflexivity must consistently be performed to ensure that the assumptions of analysis are kept while

recognizing and shelving personal bias toward the study. This was done through consistent journaling during and after recording interviews, as well as during analysis and reporting.

Results

Although analysis of both research questions occurred simultaneously, results are presented separately in this section. RQ1 (cognitive barriers) results will be shown first, followed by results for RQ2 (emotional barriers). All items (DI and SPC) and racial identities are grouped within initial results, but sensitivity analyses by item type and racial group were added to aid in interpreting experiences between different facets of the experiment. Figure 1 shows two thematic maps of results – one for each research question.

RQ1: Possible Cognitive Barriers

One theme emerged in these results, aptly named “Possible cognitive barriers,” captures parts of an item that can interrupt information processing. Four codes comprise this theme, including notable language, item length, information overload, and need for further information.

Notable Language

Typically, test developers worry about hyper-specific jargon related to the construct being measured and language that may exceed the reading or listening comprehension of the test taker (given the vocabulary is not part of the construct being measured; Zieky, 2016). For some participants, item protagonists’ names were hard to say out loud. Bella externalized her resistance to pronounce a Chinese name out loud (“Zhang Wei”).

Bella: Should I—skip over [this name], or...

Interviewer: Go ahead and do the best you can

Bella: I don't want to, like, mess it up—

Of important note, Zhang Wei was one of few names where participants hesitated or skipped entirely when reading. Other instances of notable language included words related to race, culture, or racism. Some of the words participants stumbled over included “condolence ceremony,” (Ally) and “colonialism,” (Bella and Audrey).

Many students could pronounce words correctly, however a small group had difficulty defining certain words or phrases, including Adam, saying “I don’t know where or who the Manahoac people are,” and “I’m not exactly sure what Juneteenth is. Something for Black freedom?” Typically, an inability to define words in an item can lower the likelihood of item comprehension. However, Ally noted her ability to confidently answer an item while encountering new and unfamiliar words:

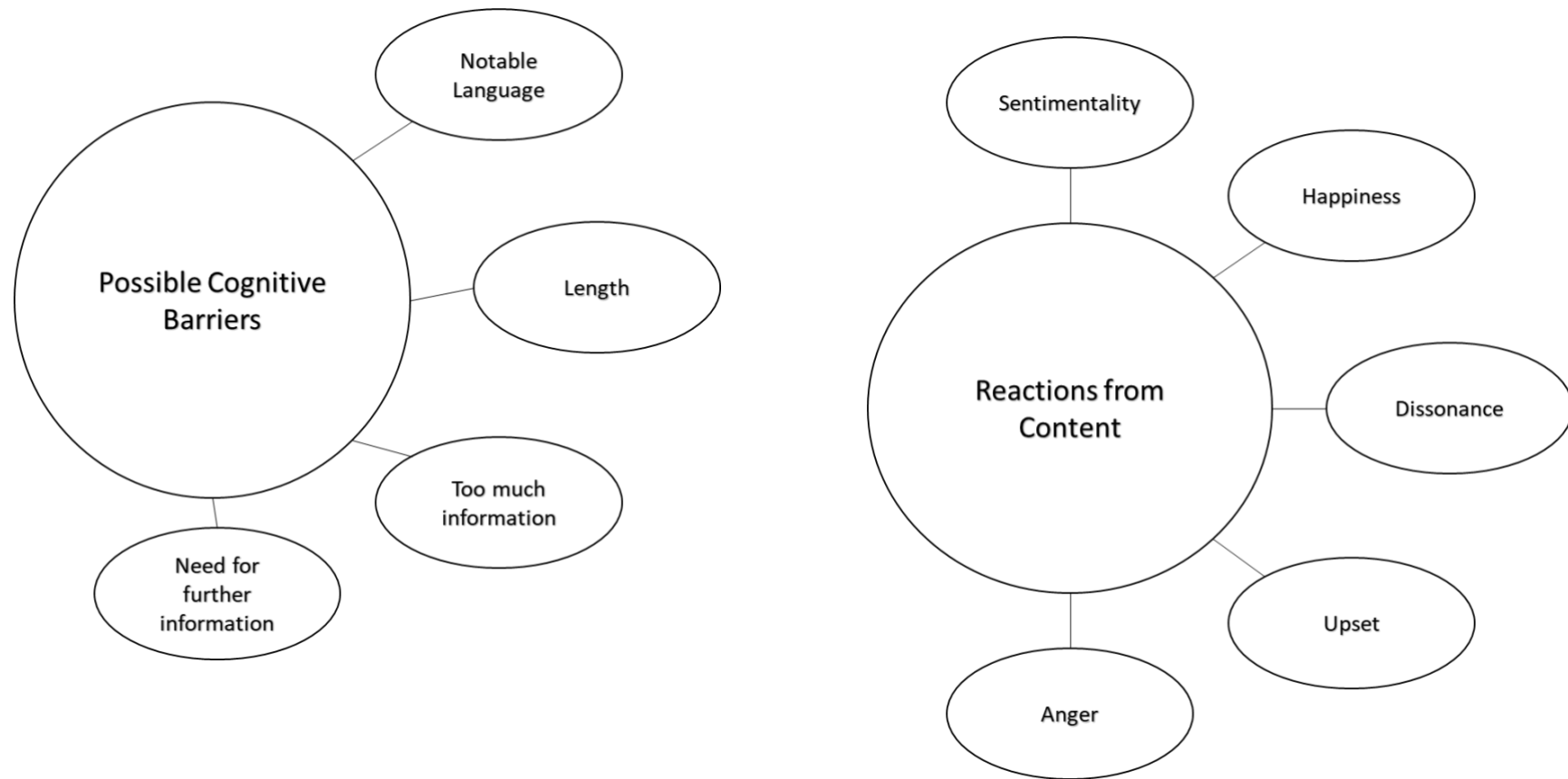
There were some words in there that I'm not familiar with. But again, I didn't think that that really fit or had to do with what the possible answers were. So I was able to use info that I didn't know to be able to come up with my final decision.

Item Length

It is reasonable to think long sentences or long item stems, regardless of reading comprehension ability, can lead to cognitive exhaustion, an obstacle in understanding and processing an item (Zieky, 2016). As item stems become longer, test-takers must spend more time trying to figure out the focal point of the sentence of item. As a result, more cognitive effort is expended on what the item is asking instead of trying to answer the item.

Comments from participants pertaining to the length of sentences and items were highly saturated across both DI and SPC items. Gigi noted “Some of the other questions have, like, a long paragraph, but I don’t

Figure 1. Thematic Maps of Results for Research Questions 1 (possible cognitive barriers) and 2 (emotional reactions



read the whole thing. I read a little at the top and a little it at the bottom and call it a day.” Many participants noted the length of a specific item that incorporated academic citations into the response options. In the response option that showcased a peer-reviewed journal citation, Audrey said “The answers are hard to read because I don’t think nobody’s going to read all these names and words.”

Although long sentences and items can come from poor sentence and passage structure—a narrative well-reflected from participants – long items can also manifest from introducing too much information.

Information Overload

The context noted in all CR and antiracist items are used to tell a story that the construct is applied to. However, introducing too much information leads to longer item stems and response options, which can make test takers spend too much time determining what information is relevant to answering the item.

Some participants felt that actions performed by an item’s protagonist were irrelevant and confusing. Calvin provided comments on this idea:

Right, because I could have just read one sentence that said, “She found an article from the NYT with several university studies that investigate why...” I don’t even need to read past “why”. So, I feel like it’s not unnecessary information, but for this question it is because you could have said anything. You could have been like, just “Rachel”, not even last name and talking about statistics or, I don’t know, bus routes, and you could use the same answers. So, it’s not really relevant and it doesn’t make me think differently, so I just don’t see why it’s there.

Others, including Calvin felt mentioning racial identities or contexts contributed to confusion, as they did not see the relevance of mentioning such contexts. Calvin noted this on a different item they interacted with:

Firstly, I’ve been kind of thinking about it – it says “a predominantly Black community” so my first impression is I don’t really see why that was mentioned, because none of the answers related back to race at all and that was kind of a random thing to say, because it’s like talking about a pollution problem in Harrisonburg and saying “the home of JMU.” Like, it doesn’t need to be said and no part of the answers relate back to it.

Overall, information overload showed the importance of connecting construct and techniques in item design, determining how to present CR and antiracist contexts in ways that still seem relevant to test takers.

Need for Further Information

On the opposite end of information overload, some participants felt some items were difficult to process because of how little information was present. One concern in item development is that if test takers perceive the item to have too little information, they may try to fill in that missing information themselves. Skyla noted about one item, “[It] just said “capstone project.” It’s not like—a project is a very broad term on what it could be, so just finding the right source [is hard].” Adam’s narrative aligned with Skyla’s, saying “Having some background knowledge on [these topics] could be important to answering these questions right.” Finally, others like Mila felt they needed to know more about the item’s protagonist to answer the item correctly:

Something that made [this item] hard was maybe not having more information about Noah’s culture.... We don’t know where Noah comes from or where his family comes from or if they have any idea about the Juneteenth flag.

Shorter items, or items that introduce incomplete context, may induce a sense of “wondering” within test takers. This wondering may lead to further confusion, as test takers feel like they must fill in knowledge gaps on their own, and then answer the item(s) presented to them.

Sensitivity Analyses

DI vs. SPC Items. There were no structural thematic differences between DI and SPC items; narratives from both DI items and SPC items made up each code. While most codes had equal contribution by DI and

SPC narratives, those who interacted with SPC items contributed relatively more to *information overload* than those who interacted with DI items.

Racial Identity. A racial sensitivity analysis revealed that the structure of the theme is the same when comparing white students to students of color. However, saturation of each code differs by group contribution. Notably, white students contributed more to the *information overload* and *notable language* codes than students of color. In the same light, all but one narrative in the *need for further information* code was from white participants.

Differences did not just exist in the saturation levels across racial groups, but in what the narratives were describing. White participants' narratives of item length were often related to relevance of racial contexts in items. According to white participants, longer items introduced information overload, but narratives from students of color did not discuss the relationship between item length and information overload. A second notable area of difference was contributions to the *need for further information* code. White students contributed the most to this code, noting they needed further information about the protagonist, the actions taken by the protagonist, or other logistics to make a more informed decision, where participants of color did not always feel they needed more information.

RQ2: Emotional Reactions

Emotional reactions can occur when the content of an item presents topics that can be uncomfortable for test-takers, like political issues or topics that relate to one's social identities (Zieky, 2016). Through coding, one theme, *reactions from content*, emerged with five codes: sentimentality, happiness, upset, anger, and dissonance. Each code is explained further below.

Sentimentality

Sentimentality was an open code used to describe feelings of tenderness and nostalgia invoked from item content. Crystal described accounts of sentimentality through relating his experiences as a biracial person with the protagonist of a DI item:

I like this question because I'm part African American and soul food is a big part of my culture. And I don't know – it made me think about that. I don't know. She's trying to find stuff about her own culture and I tried to find some of my own culture. I find it a pretty cool and relatable question. I like it.

Gigi also expressed sentimentality reading through another DI item, discussing her mother's passing when processing the topic of familial recipe sharing:

My mom passed away, so there's some things that like...I'm not. I can never eat [certain foods] again because it's not the same, like when she'd make it so that and when there's something I like she made like really well like I want to learn how to make, but I just didn't get the chance to.

Sentimentality often led to a different emotion, pending what the participant was being sentimental about. Crystal expressed happiness after, while Gigi felt upset.

Happiness

Upon processing certain topics in items, some participants felt positive emotions related to happiness. Crystal's happiness manifested from feeling represented in items, especially after he talked about relating to an item protagonist:

I love these questions. I like that they are very culturally appropriate and really aware and representative. I like that... They made me feel represented in testing, which is something that I didn't think that I wanted until I got it. [laughs] We're all represented in media and stuff like that. Okay, cool. Seeing my face on TV, that's all right. I've seen that before. But seeing it in tests is something I didn't know I needed until I got it. That's kind of really cool because I didn't really notice

that every single test up until this point has been like Sarah wants to do this, this, this, this, this. And I don't know, in my brain, I kind of conjured up this generic white situation between white people things. And I don't know, this one seemed more personal.

Others, like Susan, felt happiness through interest in the topic being discussed. Susan noted “This test kept everything more interesting. I actually think, like, if you took a longer test and it had like a variety of stuff, it would keep you engaged longer.”

Upset

Becoming upset wasn't always something that was explicitly stated. While talking about her mother's passing from a conversation about an item, Gigi's sentimentality (see above) turned into physical upset, when she started crying and apologized for it: “Sorry, I didn't mean—do you have a tissue? I didn't realize this would happen, I'm sorry.” However, some participants started feeling upset. Alaina noted the set of SPC items she interacted with were invoking negative emotions:

I think all of the topics we've been through were upsetting. Like, the one—it was about Black communities—like, the housing one. I mean, not upsetting per se, but like this one, where people been killed or went missing—I mean, that's always like, part of life, which sucks. I think it's terrible.

Anger

Rather than feeling generally upset, Mary and Crystal had more extreme negative emotions, like anger. Mary noticed herself moving from being upset to anger while discussing an SPC item on missing and murdered Indigenous women:

The emotion that I'm feeling now is anger because, again, it's not something that's talked about. It's because it's a group of minorities, and like, at least in this country, unless it's a white person it's not really looked into. Like, I remember that there was the whole case of... I forgot her name, but it was a white woman gone missing and all over the media her picture was being shown and all this stuff, but when it's a minority it's not really talked about or gone global, which is upsetting because yes, finding that woman was important, but so is the rest of the people that have gone missing and killed.

Crystal's anger was fueled by seeing a perceived sacred topic, Juneteenth, being shared on a public test:

I don't want to get into it, but I just don't like the fact that [Juneteenth] is [a national holiday] now. It used to be just a Black people thing and now it's the whole country's thing. I don't know. I don't like it. So tying back to this, it's more like, 'Okay, why is this a question?' I don't like this question. I don't know.

Dissonance

Very few participants experienced emotional dissonance. In an exemplar narrative, Crystal felt dissonance about feeling represented and wanting to see more questions relating to his racial and cultural identities, but felt like certain topics should not be shared on a test:

Yeah, it gives mixed messages – not mixed messages, but mixed feelings. I was reading the question and I was like, Okay, blah, blah, blah, what's the meaning – Juneteenth? Why is that here? I kind of don't like it. It's not that I don't like the holiday or like the meaning of the day. That's cool and all. I love that. We have a nice big cookout every year for it. But in a question? I feel like it's kind of the fact that it's in a question on a test. If I saw this on a test, this would make me a little bit more upset. Because it's already a holiday and now you guys are putting it in tests. I understand that this is an academic thing but I don't know, I feel like this one just shouldn't be touched. That's why. It kind of strikes a nerve a little bit. Why is this here? I like it. But at the same time, I don't like – not everything should be everywhere, in tests and things. I understand if you guys want to be – what's it called – inclusive and stuff, but I don't need to put Juneteenth in there. Leave us out of it.

Sensitivity Analyses

DI vs. SPC Items. Comparing narratives by item type shows more frequent and stronger emotional reactions in SPC items than in DI items. Although students across both groups experienced positive emotions, more students who interacted with SPC items noted negative reactions such as being upset and angry. However, it is important to note that these differences are driven by only a few, highly expressive participants. For example, Crystal was the only participant interacting with DI items that expressed anger and dissonance, while only two participants (Mary and Nyx) expressed anger or dissonance when interacting with SPC items.

The source of emotional reactions also differed between DI and SPC item narratives, specifically in negative emotions. Those who interacted with DI items noted their negative reactions stemmed from identity-based experiences, like Crystal. However, those who interacted with SPC items noted their negative reactions stemmed from their passion or discomfort around discussions of race, racism, and racial equity.

Racial Identity. A racial sensitivity analysis showed the type and strength of emotional reactions expressed differed between white participants and participants of color. Therefore, the thematic structure may not hold perfectly between the two groups, where differences in reactions are driven primarily by one or two participants.

For example, happiness was experienced by participants of each racial identity group, but the source of happiness differed. White participants generally described her happiness through interest and excitement toward seeing inclusive content in the items. However, participants of color expressed their happiness via noting feeling represented in items. Negative emotions more greatly differed between groups. White participants primarily noted how upset or uncomfortable they were by describing either their empathy toward marginalized communities or discomfort in discussing race and racism within SPC items. Participants of color primarily noted anger, although some noted feeling upset. This difference, though, is primarily driven by Crystal and Nyx, as no other participant of color noted such extreme or even negative emotions.

Discussion

CR and antiracist assessments are framed as assessments designed to ensure stronger validity of results across test takers, especially those of marginalized racial and cultural identities (Bennett, 2023; Randall, 2021). Specifically, these assessments combat the perpetuation of whiteness through introducing items that are contextually related to test-takers' lived experiences, heightening validity arguments while introducing micro interventions to combat internally held notions of racism (Patterson, 2025; Randall, 2021; Randall et al., 2022).

The current study explored concerns regarding test takers' cognitive and emotional processing of CR items through interacting with a specific subset CR and antiracist items (DI and SPC items, respectively). Results showed mixed alignment between the advantages and perceived shortcomings of CR and antiracist assessments. In terms of cognitive barriers, participants noted seeing items that were too long, had either too much or too little information, and sometimes included difficult names or terms. This is not different than what current literature (e.g., Haladyna & Rodriguez, 2013; Zieky, 2016) already note—the themes from this study are the most common forms of cognitive barriers in item comprehension. There were instances of parallelism between the *Standards* and this study in emotional reactions as well: some topics, especially the topics in the SPC items, triggered certain emotions that could be considered extreme enough to distract test takers from answering items in a real test environment.

There were times when cognitive barriers intersected with emotional disturbance. For example, while one participant's narrative was wondering about what Juneteenth was in an item, another participant's narrative showed anger against the inclusion of the topic of Juneteenth itself. While some may see an item

as a cognitive barrier, others may already be able to comprehend it and have an emotional reaction to it. In the context of CR and antiracist item construction, this reinforces the need to educate test takers on race-based topics (NCME, 2022), while ensuring the topic being discussed is not a sacred cultural one (Patterson, 2025).

The sensitivity analyses done for each theme, by race and by item type, may shed light on how certain people processed certain items. The SPC items elicited more narratives on information overload and upset, leading to closer examination of how the SPC items can be shortened to include only the most relevant information when discussing antiracism. Further, the majority of white students noted being upset toward at least one SPC item vs. very few POC. This may be a manifestation of the white defensiveness from a race-based stimulus (Watt, 2015) or could be just upset about the general existence of racism. In either case, this may mean that antiracist assessments are more difficult for white students, as they would experience emotional disturbance.

On the other hand, a higher proportion of participants who were not white experienced more positive emotions. Although positive emotions in assessment have been studied (e.g., Lehman et al., 2017; Pekrun et al., 2004), the source of happiness during test taking has been rarely discussed. The existence of happiness in this study primarily came from feeling a connection to the items' contexts in some way, like *Crystal*.

Limitations and Transferability

Limitations in this study are few but allow opportunities for further research. First, the constructivist approach to this study means that conclusions note general possible cognitive barriers and emotional reactions to DI and SPC items. However, to further research in CR and antiracist item development and item processing, a critical approach is needed. Specifically, using critical race theory (Delgado & Stefancic, 2017) and critical whiteness theory (Matias et al., 2014) as frameworks to interpret data may give researchers more information on how socialization into culture and cognition may be related to socialization into racial social structures, a factor in validating any assessment. The other large limitation of this study is in participant recruitment. Although self-selection is common in educational studies, those who responded to participate in the study may have higher cultural competence or multicultural exposure than typical test takers. Future studies could include random sampling from a more culturally diverse population. The final limitation is that emotions were not studied in context with test scores. Given the setup of this study, future studies should examine the impact of emotions on CR and antiracist test scores when interacting with multiple-choice items.

Transferability of results should focus on college-aged students and adult learners interacting with items that reflect mostly localized examples of culture, race, and racism. Extrapolation of results to other populations should consider the cognitive and emotional development of the population being studied, including how to frame discussions of culture, race, and racism with that population. The topics presented in this study, or how the topics are presented in the items used in this study, may be too complex for K-12 students compared to college students and adult learners. Further, students in other parts of the United States and the world may also react differently.

Implications and Future Directions

As CR and antiracist assessments gain further momentum in both research and practice, it is important to know how these items affect test takers' emotional and cognitive processes. Given the results of this study, both practitioners and researchers should work together to learn more about various facets of test construction and the testing experience.

First, learning the appropriate topics to introduce in items, as well as how to introduce them, should be studied more. Some antiracist and CR assessment frameworks now emphasize working with the test-taking

population to understand what topics are comfortable to discuss, and how to discuss them (Randall et al., 2022; Walker et al., 2019). While consulting with the proper populations, researchers and practitioners should take care to not reduce cultural experiences or manifestations of antiracism to stereotypes or generalized assumptions to make items sound more “interesting” to other test takers. Racial capitalism, the practice of using the experiences of racially marginalized to generate financial or social gain (Ralph & Singhal, 2019), can manifest in this way. Future research and practice should address where the boundary lies between wanting authenticity in items while profiting off of those authentic experiences.

Next, as emotional reactions were not all negative, future work should address the relationship between emotional reactivity and engagement with CR and antiracist items. Although negative emotions and their effects in testing have been researched, positive emotions have been less so, especially in CR and antiracist items and tests overall. Some participants expressed strong positive emotions while interacting with items; if negative emotions lead to negative test engagement and weakened validity arguments, the opposite could be true with positive emotions. Future research could explore this relationship, as well as establishing what topics may induce such positive emotions.

Finally, this research is posed as a reminder that our socializations into race are what primarily dictate reactions to stimuli (Spanierman & Cabrera, 2014; Watt, 2015). As a result, it can be argued that socializations into racial identities influence the reactions we have to CR and antiracist test items. To truly achieve racial justice, researchers and practitioners should strive to understand how socialization into race affects one’s reactions to CR and antiracist items. In total, if racial justice is seen as an end goal, then we must work with others in education to prevent socialization into oppressive racial systems where talking about race and racism is taboo.

Received: 4/16/2025. **Accepted:** 2/22/2026. **Published:** 2/24/2026.

Citation: Patterson, C. (2026). Exploring cognitive and emotional processing of culturally responsive and antiracist test items. *Practical Assessment, Research, & Evaluation, 30*(2)(7). Available online: <https://doi.org/10.7275/pare.3055>

Corresponding Author: Chris Patterson, Knox College. Email: crpatterson@knox.edu

References

- Abaied, J.L., & Perry, S.P. (2021). Socialization of racial ideology by white parents. *Cultural Diversity and Ethnic Minority Psychology, 27*(3), 431-440. <https://psycnet.apa.org/doi/10.1037/cdp0000454>
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *The standards for educational and psychological testing*. American Educational Research Association.
- Aronson, B., & Laughter, J. (2016). The theory and practice of culturally relevant education: A synthesis of research across content areas. *Review of Educational Research, 86*(1), 163-206. <https://doi.org/10.3102/0034654315582066>
- Au, K., & Jordan, C. (1981). Teaching reading to Hawaiian children: Finding a culturally appropriate solution. In H. Trucba, G. Guthrie, & K. Au (Eds.), *Culture and the bilingual classroom: Studies in classroom ethnography* (pp. 139-152). Newbury House.

- Bennett, R. (2023). Toward a theory of socioculturally responsible assessment. *Educational Assessment, 28*(2), 83-104. <https://doi.org/10.1080/10627197.2023.2202312>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77-101. <https://doi.org/10.1191/1478088706qp063oa>
- Braun, V., & Clarke, V. (2019). Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise, and Health, 11*(4), 589-597. <https://doi.org/10.1080/2159676X.2019.1628806>
- Braun, V., & Clarke, V. (2020). Can I use TA? Should I use TA? Should I *not* use TA? Comparing reflexive thematic analysis and other pattern-based qualitative analytic approaches. *Counseling and Psychotherapy Research, 21*, 37-47. <https://doi.org/10.1002/capr.12360>
- Brown-Jeffy, S., & Cooper, J.E. (2011). Toward a conceptual framework of culturally relevant pedagogy: An overview of conceptual and theoretical literature. *Teacher Education Quarterly, 38*(1), 65-84. <https://files.eric.ed.gov/fulltext/EJ914924.pdf>
- Byrne, D. (2022). A worked example of Braun and Clarke's approach to reflexive thematic analysis. *Quality & Quantity, 56*, 1391-1412. <https://doi.org/10.1007/s11135-021-01182-y>
- Cabrera, N.L. (2018). *White guys on campus: Racism, white immunity, and the myth of "post-racial" higher education*. Rutgers University Press.
- Cholewa, B., Goodman, R.D., West-Olatunji, C., & Amatea, E. (2014). A qualitative examination of the impact of culturally responsive educational practices on the psychological well-being of students of color. *The Urban Review, 46*, 574-596. <https://doi.org/10.1007/s11256-014-0272-y>
- Chung, A. H., & Slater, M. D. (2013). Reducing stigma and out-group distinctions through perspective-taking in narratives. *Journal of Communication, 63*, 894-911. <https://doi.org/10.1111/jcom.12050>
- Clarke, K.E., & Radcliff, C. (2018). Information literacy assessment for instruction and improvement and demonstration of library value: Comparing locally-grown and commercially created tests. *Libraries, 161*. https://commons.lib.jmu.edu/letfspubs/161?utm_source=commons.lib.jmu.edu%2Fletfspubs%2F161&utm_medium=PDF&utm_campaign=PDFCoverPages
- Coleman, J.S., Campbell, E.Q., Hobson, C.J., McPartland, J., Mood, A.M., Weinfeld, F.D., & York, R.L. (1966). *Equality of educational opportunity*. United States Government Printing Office.
- DeCuir-Gunby, J.T., Marshall, P.L., & McCulloch, A.W. (2011). Developing and using a codebook for the analysis of interview data: An example from a professional development research project. *Field Methods, 23*(2), 136-155.
- Dixon-Román, E. (2016). Diffractive possibilities: Cultural studies and quantification. *Transforming Anthropologies, 24*(2), 157-167. <https://doi.org/10.1111/traa.12074>
- Román, E. (2020). A haunting logic of psychometrics: Toward the speculative and indeterminacy of blackness in measurement. *Educational Measurement: Issues & Practice, 39*(3), 94-96. <https://doi.org/10.1111/emip.12375>
- Gaddis, S.M. (2017). How black are Lakisha and Jamal? Racial perceptions from names used in audit correspondence studies. *Sociological Science, 4*, 469-489. https://www.sociologicalscience.com/download/vol-4/september/SocSci_v4_469to489.pdf
- Gay, G. (2018). *Culturally responsive teaching: Theory, research, and practice* (3rd ed.). Teachers College Press.

- Gee, G.C., Walsemann, K.M., & Brondolo, E. (2012). A life course perspective on how racism may be related to health inequities. *American Journal of Public Health, 102*, 967-974. <https://ajph.aphapublications.org/doi/full/10.2105/AJPH.2012.300666>
- Haladyna, T.M., & Rodriguez, M.C. (2013). *Developing and validating test items*. Routledge.
- Hall, E.T. (1992). *An anthropology of everyday life* (2nd ed.). Anchor.
- Hathcoat, J.D., Meixner, C., & Nicholas, M.C. (2018). Ontology and epistemology. In P. Liamputtong (Ed.), *Handbook of research methods in health social sciences*, (pp. 1-19). Springer.
- Hood, S., Hopson, R., & Frierson, H. (2005). *The role of culture and cultural context: A mandate for inclusion, the discovery of truth and understanding in evaluative theory and practice*. Information Age Publishing.
- Ibarra, R.A., & Cohen, A.S. (1999). Multicontextuality: A hidden dimension in testing and assessment. In *Proceedings from new directions in assessment for higher education: Fairness, access, multiculturalism, and equity (FAME) conference* (pp. 16-30).
- Ladson-Billings, G. (1995). Toward a theory of culturally relevant pedagogy. *American Educational Research Journal, 32*(3), 465-491. <https://doi.org/10.3102/00028312032003465>
- Ladson-Billings, G. (2014). Culturally relevant pedagogy 2.0: A.k.a. the remix. *Harvard Educational Review, 84*(1), 74-135. <https://doi.org/10.17763/haer.84.1.p2rj131485484751>
- Lehman, B., Hebert, D., Jackson, G. T., & Grace, L. (2017). Affect and experience: Case studies in games and test-taking. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (pp. 917-924). ACM.
- Leighton, J.P. (2017). *Using think-aloud interviews and cognitive labs in educational research*. Oxford University Press.
- Levy, M., & Myers, D. (2020). Racial projections in perspective: Public reactions to narratives about rising diversity. *Perspectives on Politics, 19*(4), 1147-1164. <https://doi.org/10.1017/S1537592720003679>
- Lumivero. (2020). *NVivo* (Version 13, 2020 R1). <https://www.lumivero.com>
- Lyons, S., Johnson, M., & Hinds, B.F. (2021). *A call to action: Confronting inequity in assessment*. Lyons Assessment Consulting.
- Matias, C.E., Viesca, K., Garrison-Wade, D.F., Tandon, M., & Galindo, R. (2014). “What is critical whiteness doing in OUR nice field like critical race theory?” Applying CRT and CWS to understand the white imaginations of white teacher candidates. *Equity and Excellence in Education, 47*(3), 289-304. <https://doi.org/10.1080/10665684.2014.933692>
- Molla, T. (2024). Racial othering: Structural roots and anti-racist actions from below. *SN Social Sciences, 4*. <https://doi.org/10.1007/s43545-023-00806-4>
- Monk Jr., E.P. (2020). Linked fate and mental health among African Americans. *Social Science and Medicine, 266*, 1-8. <https://doi.org/10.1016/j.socscimed.2020.113340>
- National Council on Measurement in Education. (2022, January 19). *Cultural relevance versus construct relevance: How do we create culturally responsive assessments?* [YouTube]. <https://www.youtube.com/watch?v=bB1FtEJLqAE>.
- Patterson, C.R. (2025). Observing test takers’ perceptions of culturally responsive items. *Educational Assessment, 30*(3), 183-196. <https://doi.org/10.1080/10627197.2025.2487115>

- Pekrun, R., Goetz, T., Perry, R. P., Kramer, K., Hochstadt, M., & Molfenter, S. (2004). Beyond test anxiety: Development and validation of the test emotions questionnaire (TEQ). *Anxiety, Stress, & Coping, 17*(3), 287–316. <https://doi.org/10.1080/10615800412331303847>
- Ralph, M., & Singhal, M. (2019). Racial capitalism. *Theory and Society, 48*, 851/881. <https://doi.org/10.1007/s11186-019-09>
- 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.
- 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. <https://doi.org/10.1080/10627197.2022.2042682>
- Santelices, M.V., & Wilson, M. (2010). Unfair treatment? The case of Freedle, the SAT, and the standardization approach to differential item functioning. *Harvard Educational Review, 80*(1), 106-133. <https://doi.org/10.17763/haer.80.1.j94675w001329270>
- Seemiller, C., & Grace, M. (2016). *Generation z goes to college*. Wiley.
- Sinharay, S., Johnson, M. S., Bennett, R. E., Lopez, R., Guzman-Orth, D., Sparks, J.R., & Pillarisetti, S. (2025). Culturally responsive assessments in the context of the National Assessment of Educational Progress: An initial look. *Educational Measurement: Issues and Practice*. <http://doi.org/10.1111/emip.70008>.
- Solano-Flores, G., & Nelson-Barber, S. (2001). On the cultural validity of science assessments. *Journal of Research in Science Teaching, 38*(5), 553-573. <https://doi.org/10.1002/tea.1018>
- Spanierman, L.B., & Cabrera, N.L. (2014). The emotions of white racism and antiracism. In V. Watson, D. Howard-Wagner, & L. Spanierman (Eds.), *Unveiling whiteness in the twenty-first century: Global manifestations, transdisciplinary interventions* (pp. 9-28). Lexington Books.
- Spring, J. (1995). *The intersection of cultures: Multicultural education in the United States*. McGraw-Hill.
- Sudman, S., Bradburn, N., & Schwarz, N. (1996). *Thinking about answers: The application of cognitive processes to survey methodology*. San Francisco: Jossey-Bass.
- Walker, M.E., Olivera-Aguilar, M., Lehman, B., Laitusis, C., Guzman-Orth, D., & Gholson, M. (2023). *Culturally responsive assessment: Provisional principles*. ETS. <https://onlinelibrary.wiley.com/doi/pdf/10.1002/ets2.12374>
- Watt, S.K. (Ed.). (2015). *Designing transformative multicultural initiatives: Theoretical foundations, practical applications and facilitator considerations*. Stylus
- Watt, S.K., Mueller, J.A., Parker, E.T., Neel, R., Pasquesi, K., Kilgo, C.A., Mollet, A.L., Mahatmya, D., & Multicultural Initiatives Consortium. (2021). Measuring privileged identity in educational environments: Development and validation of the privileged identity exploration scale. *Frontiers in Education, 271*. <https://doi.org/10.3389/feduc.2021.620827>
- Zieky, M.J. (2016). Fairness in test design and development. In N.J. Dorans & L.L. Cook (Eds.), *Fairness in educational assessment and measurement* (pp. 9-31). Routledge.

Supplemental Materials:
**Exploring Cognitive and Emotional Processing of Culturally Responsive and Antiracist Test
Items**

Table of Contents

- I. Items used in each cognitive interview
- II. Cognitive Interview Process

Section I.

All Diversity-Infused (DI) and Sociopolitical Consciousness (SPC) Items Used in the Study

Note: Each row represents a DI-SPC pair derived from the same original item.

Original items not shown due to copyright.

Items are published here with permission from the author of the items.

Sociopolitical Consciousness Item	Diversity-Infused Item
<p>Aaliya Khan wants to do her senior capstone project on biased policing against the African American community in the United States. She is hoping to use this research to lobby for police reform on campus. Which of the following sources would be useful to help this student <i>get started</i>?</p>	<p>Aaliya Khan is starting her senior capstone project on wearing the Hijab in the 21st century. She is hoping this research will inform how she navigates her campus and workplace. Which of the following sources would be useful to help this student <i>get started</i>?</p>
<p>A: A peer reviewed journal article, like: Spencer, K.B. (2016). Implicit Bias and Policing. <i>Social and Personality Psychology Compass</i>, 10(1), 50-63.</p>	<p>A: A peer reviewed journal article, like: Haddad, Y.Y. (2007). The post-9/11 hijab as icon. <i>Sociology of Religion</i>, 68(3), 253-267.</p>
<p>B: An entry from a subject encyclopedia, like: Darity, William A. (ed.). Policing, biased. In <i>International Encyclopedia of the Social Sciences</i> (Vol 6., 2nd Ed.). Gale Publishing</p>	<p>B: An entry from a subject encyclopedia, like: King-Irani, L. (ed.). (2018, June 8). Hijab. <i>Encyclopedia.com</i>.</p>
<p>C: A scholarly book, like: Weitzer, R., & Tuch, S. (2006). <i>Race and Policing in America: Conflict and Reform</i>. Cambridge University Press.</p>	<p>C: A scholarly book, like: Pasha-Zaidi, N., & Pasha, S. (eds.). (2017). <i>Mirror on the veil: A collection of personal essays on hijab and veiling</i>. Critical, Cultural and Communications Press.</p>
<p>D: A chapter from a scholarly book, like: Ridgeway, G., & MacDonald, J. Methods for Assessing Racially Biased Policing. In <i>Race, Ethnicity, and Policing</i>, edited by Stephen K. Rice & Michael D. White, pp. 180-204. New York, NY., 2010.</p>	<p>D: A chapter from a scholarly book, like: Mir-Hosseini, Z. (2011). Hijab and choice: Between politics and theology. In M. Kamrava (ed.), <i>Innovation in islam</i> (pp. 190-212). University of California Press.</p>

After hearing about land acknowledgements, Carson Kay is working on a presentation about the land that Western European settlers have taken from Indigenous peoples. He wants to spread awareness of the effects of colonialism by showing the sizes of Indigenous homelands from the year 1600 to now. He should

- A: talk about the differences between the sizes of Indigenous homelands in 1600 and today.
- B: quote from a source that talks about how land has been taken from Indigenous peoples.
- C: show pictures depicting the size of Indigenous homelands in 1600 and today.
- D: play an audio recording of an expert on Indigenous history.

Ballroom culture is an underground queer, Black, and Latino community that has influenced pop culture fashion design. Carson Kay is working on a presentation to show this influence. He wants to show the similarities between fashion styles in Ballroom culture and pop culture fashion designers. He should

- A: talk about the similarities between Ballroom and pop culture fashion designs.
- B: quote from a source that details how Ballroom culture has influenced current pop culture fashion.
- C: show pictures of fashion garments in Ballroom culture and current pop culture fashion.
- D: play an audio recording of an expert on fashion.

As a part of Isabella Cortez's internship, she was assigned to research why a larger proportion of Black and Indigenous people have died from COVID-19 compared to White people. When she searches the term "COVID-19 deaths" in a database, Isabella gets over 3000 results. How can she narrow the search results?

- A: Add a search term that makes the search more specific (e.g., COVID-19 deaths and race)
- B: Limit results to the most recent to ensure currency
- C: Ask her supervisor to change the topic to one that returns fewer results from her search
- D: Follow the work of an expert on race and COVID-19

Isabella Cortez wants to write her anthropology paper on Indigenous death rituals. When she searches the term "death rituals" in a database, she gets over 3000 results. How can Isabella narrow the search results?

- A: Add a search term that addresses a specific death ritual (e.g., scaffold burial, condolence ceremony)
 - B: Limit results to the most recent to ensure currency
 - C: Change her topic to one that returns fewer results from her search
 - D: Follow the work of an expert on Indigenous death rituals
-

Shanice Richards wants to raise awareness of housing segregation in Harrisonburg. To inform her actions, she needs to research the history of housing segregation in the United States. Which of the following sources would be most useful for her to use?

- A: A website by the Shenandoah Valley Black Heritage Project that features housing segregation
- B: A website created by a JMU professor that studies history
- C: A chapter from *The Color of Law*—a book written by scholars in race and housing
- D: A brochure from a realtor on houses in Harrisonburg

Shanice Richards is writing her final research paper on the origins of soul food to learn more about her own culture. Which of the following sources would be most useful for her to use?

- A: A website that features recipes of different soul food dishes
- B: A website created by a soul food chef
- C: A chapter from *Hog & Hominy*—a book on soul food origins written by a food history scholar
- D: A brochure menu from a local soul food restaurant

Jamie Stubbs is working on a paper for her environmental science class on how to help the people of Reserve, Louisiana, a predominantly Black community that suffers from the effects of untreated chemical waste. Which of the following sources would be most appropriate for her to include?

- A: a Wikipedia entry titled, *Cancer Alley*
- B: a book that claims that Reserve, Louisiana is not a real place
- C: a peer-reviewed article written by scientists who study the effects of untreated chemical waste
- D: an opinion piece written by an elected official seeking re-election in the Reserve, Louisiana area

Jamie Stubbs is researching controlled forest burning, an agricultural practice of Indigenous peoples. Which of the following sources would be most appropriate for her to include?

- A: a Wikipedia entry titled, *Controlled burn*
 - B: a book that claims that controlled burning is a conspiracy
 - C: a peer-reviewed article written by scientists who study controlled burning
 - D: an opinion piece written by an elected official who lives near Indigenous peoples
-

Mia Ong is giving an advocacy speech in her communications class on supporting the Stop Asian Hate movement. Due to her own experiences as an Asian woman, she is confident that she knows a lot. Due to her life experience, she did not feel the need to include any sources that supported her argument. In meeting with her professor, her professor tells her to revise her speech. Why would Mia's professor do that?

- A: Mia's professor is dismissing her experience
- B: Mia did not support her experience with additional experts' evidence
- C: Mia's professor is punishing her for not doing enough research
- D: Mia's experience doesn't belong in an academic presentation

Sergio Rodriguez is giving an informative presentation on the creation stories told by the Manahoac peoples, which teach Manahoac children about life lessons and spirituality. He is confident that he knows a lot about this, as his grandmother is a member of the Manahoac nation. Due to his knowledge, he did not feel the need to include any sources that supported his experience.

His professor gave him a lower grade than expected. Why would Sergio's professor do that?

- A: Sergio's professor is dismissing his experience
- B: Sergio did not support his experience with additional experts' evidence
- C: Sergio's professor is punishing him for not doing enough research
- D: Sergio's experience doesn't belong in an academic presentation

Through a practicum experience, Zhang Wei has been working with the Lucy Simms Center to help support low-income families in the northeast neighborhood of Harrisonburg, which is a primarily Black neighborhood. Zhang was invited to present his work to city council. He came to the meeting and showed a video about the history of the Lucy Simms Center, which took up the majority of his speaking time. What could have Zhang done to best improve the experience?

- A: Provide a link so that his audience can view it later
- B: Nothing, showing the entire video was appropriate
- C: Create a presentation that doesn't include the video
- D: Use part of the video to help him show an example of the importance of the Center

Deja Taylor is giving a class presentation on how family recipes are passed down in Black culture. As most family meals are shared through spoken word and not a recipe, she made a video that showed her mother and grandmother cooking together. The video took up most of Deja's presentation time. What could Deja have done to best improve her presentation?

- A: Provide a link so that her audience can view it later
 - B: Nothing, showing the entire video was appropriate
 - C: Create a presentation that doesn't include the video
 - D: Use part of the video to help her show an example of how food recipes are passed down in her culture
-

The student government association wants to build a display honoring Indigenous women that have been killed or went missing in the past year. In their work, the association is researching why Indigenous women have gone missing and been killed at higher rates than other groups. They are having a hard time finding sources for this topic, despite trying a few different strategies. What can they do to move their work forward?

A: Ask for help from a librarian

B: Ask other students on campus to help them brainstorm

C: Ask for help from their campus advisor

D: Ask a trusted mentor to talk through ideas with them

Khaled Johnson has been assigned to add to a public Wikipedia page for a class. To help inform others on how they can protest to support the Black Lives Matter movement, he chooses to add to a page on *Protest Tactics*. He is having a hard time finding sources to help add additional content to the page. What should Khaled do?

A: Ask to change his topic to one he can easily locate sources

B: Talk to a librarian about where he might look

C: Narrow his topic further to find more specific sources

D: Use the sources he can easily find

Noah Reese is creating a presentation on the meaning of the shapes and colors on the Juneteenth flag. He finds that the arc stands for hope and a promising future, and the star stands for freedom for Black Americans. He is having a hard time finding sources that tell the meaning of the colors of the Juneteenth flag. What can Noah do to move his work forward?

A: Ask for help from a librarian

B: Ask his roommate to help him brainstorm

C: Ask his professor if he can change his topic

D: Ask his mom to talk through his ideas with him

Khaled Johnson is giving a presentation on the popular misconceptions of voodooism. He chooses to address how media has led people to believe that voodoo dolls are used to harm others. He wants to show people that voodoo dolls are actually used for communicating with deities. However, he is having a hard time finding sources to help add additional content to the page. What should Khaled do?

A: Ask to change his topic to one he can easily locate sources

B: Talk to a librarian about where he might look

C: Narrow his topic further to find more specific sources

D: Use the sources he can easily find

Rachel O'Brien will give a speech in her communications class to advocate for voting access in the United States. She has found an article from *The New York Times* reviewing several university studies that investigate why People of Color have access to fewer voting booths than White people. How should Rachel use this source?

- A: Use *The New York Times* article since it summarizes the work for her
- B: Use the research studies referenced in the article because they are much more in-depth
- C: Use both the news article and the university studies to introduce and provide detail
- D: Use other sources as the article already has done this work and she needs to be original

Rachel O'Brien is giving a speech in her class about meat preparation and consumption in India. She has found an article from the *New York Times* reviewing several university studies that investigate the differences in how meat is prepared, consumed, or avoided in different regions across India. How should Rachel use this source?

- A: Use *The New York Times* article since it summarizes the work for her
 - B: Use the research studies referenced in the article because they are much more in-depth
 - C: Use both the news article and the university studies to introduce and provide detail
 - D: Use other sources as the article already has done this work and she needs to be original
-

Section II. Cognitive Interview Process

Part I: Introduction

Once the participant walked into the room, they were introduced to the study, its purpose, and consent was obtained. The participant was told they could leave at any time or could have their responses taken out of the study without penalty. Once the participant signed the IRB consent form, the interview began.

Part II: Instructions and Practice

The participant was given information as to the structure of the cognitive interview, noting they were to read aloud the item stem and response options, and verbalize their thought process while answering items. After answering questions about the process, the interviewer modeled the process by answering a question, then asked the participant to practice on another question. After answering any further questions about the process and gaining further consent, the interview began.

Part III: Answering Item-Level Questions

The participant turned to the first item in the item packet in front of them, read the item stem and response options out loud, then processed and answered the item aloud. After the participant answered the item, they were asked questions about their cognitive and emotional processes when answering the item, as well as other thoughts about the content of the item itself. After questions were answered, which included the use of small probes to gain further information, the participant repeated the process for the next item. This process continued for 10 items.

Part IV: Answering Test-Level Questions

After all 10 items were processed and discussed, the interviewer asked the participant questions about the set of items they interacted with. Questions included if the participant wanted to comment on any of the items, the inclusion of culturally responsive or antiracist content into items, and how they felt about seeing these items on a real test.

Part V: Ending the Interview

Once the participant and interviewer felt that the participant had contributed their full thoughts, the interviewer ended the interview. During this part, the interviewer told the participant that they could email the interviewer at any time to have their results taken out of the study. The participant was also told they would be contacted once the interview was transcribed, where they could provide further commentary on any item or part of the interview.

Part VI: Post-Interview Process

All interviews were transcribed via a third-party company. After transcription, participants were emailed a written copy of their specific interviews and given a week to provide further commentary on their responses to items, as well as another opportunity to leave the study. After the week deadline, the newest version of transcriptions was used in analysis.

A more in-depth protocol can be found at this link (posted with permission from the author):
<https://www.tandfonline.com/doi/suppl/10.1080/10627197.2025.2487115>