IMPACT OF A JUSTICE-ORIENTED ASSESSMENT ON STUDENT EXPERIENCES

Susan Lyons, Fiona Hinds, Hannah Denker, and Sanford Student

June 2022

Lyons

ASSESSMENT CONSULTING

IMPACT OF A JUSTICE-ORIENTED ASSESSMENT ON STUDENT EXPERIENCES

Susan Lyons, Fiona Hinds,
Hannah Denker, and Sanford Student

ACKNOWLEDGEMENTS

[1] Teachers are intentionally unnamed to help protect the identities of
the participating students.

TABLE OF CONTENTS

TABLE OF CONTENTS

ABSTRACT

While curriculum and instruction have been rapidly evolving to embrace the central roles of culture and identity in learning, the field of educational assessment has remained largely stagnant in this regard. The goal of this research is to understand how the district-wide standardized, multiple-choice assessments at Boston Public Schools (BPS) might be able to better support the learning of all students by recognizing and engaging with the backgrounds and identities of BPS' diverse student body. We planned an experimental, mixed-methods study design that relies on assessment and survey data to gather information on student test taking performance and experiences. The plan also included follow-up focus groups to provide additional, qualitative insight into the quantitative findings. Due to difficulty recruiting educators during a surge of the COVID-19 Omicron variant, we were unable to reach our target sample size and instead implemented a fully qualitative study with 13 students. This paper focuses on the process and outcomes associated with developing the culturally responsive, justice-oriented ELA assessment with engagement from a small but dedicated group of Boston Public Schools educators. Survey and focus group results reveal that students were able to identify with the people in the reading passages and able to make connections between the reading passages and their own lives. The process of assessment development employed for this study represents a departure from current industry practice in standardized assessment development and therefore has the potential to inform more far-ranging assessment innovations within the district and beyond, including impact for large-scale assessment programs.

INTRODUCTION

The Boston Public Schools (BPS) 2020-2025 Strategic Vision states that,

> "At BPS, every child in every classroom is entitled to an equitable, world-class, high-quality education. Each child should have the same unfettered access to every conceivable resource to unlock the greatness within them. For this to happen, we must eliminate the structural and institutional obstacles, especially those exacerbated by race, language, special learning needs, socioeconomic status, and other factors. This requires a commitment to systemic change in the way we allocate funding, provide access to information, deliver instruction, and make resources available to meet students' needs."

While curriculum and instruction have been rapidly evolving to embrace the central roles of culture and identity in learning as recognized by the strategic vision, the field of educational assessment has remained largely stagnant in this regard. The goal of this research is to understand how the district-wide standardized, multiple-choice assessments at Boston Public Schools might be able to better support the learning of all students by recognizing and engaging with the backgrounds and identities of BPS' diverse student body. BPS has already undertaken efforts to update the passages used in its grade 6-12 reading Benchmark assessments to better reflect its cultural, racial, and socioeconomic diversity. The current research study explores the feasibility of more deeply connecting to the specific cultures and identities of the students served in this district by collaborating with teachers to curate even more relevant passages for use in the Benchmark assessments. The intention is for the passages to go beyond diversity and align with the district's goals to develop an anti-racist assessment system that values all students, attends to the particular sociopolitical realities in the community, and empowers students as agents of social justice.

An anti-racist assessment system is one that is intentionally designed to disrupt racism and racist structures. Anti-racism in assessment attends to both the *content and design* of assessments, and the *interpretation and use* of assessments to create a more just future for all. The current study is intended to help the district take a step in this direction by attending to just the first of the four key components for an anti-racism assessment system: the content.

> An anti-racist assessment system is one that is intentionally designed to disrupt racism and racist structures. Anti-racism in assessment attends to both the *content and design* of assessments, and the *interpretation and use* of assessments to create a more just future for all.

Historically, item writing for large-scale assessments has attempted to guard against introducing bias through insensitive or culturally specific language. However, the idea that all cultural context can be removed from an assessment is a false premise. When we believe we can "de-contextualize" our assessment items we are simultaneously adopting outdated assumptions about the nature of competence and developing items centered in the dominant culture (Resnick & Resnick, 1992; Randall, 2021). Developing items that seek to connect to and affirm the cultures of students of color and foster their identities as agents of social justice is a way to center anti-racism within large scale test design.

The BPS Strategic Vision states its theory of action as follows: "IF we give every student what they need, earn the trust and true partnership of families, community members, and stakeholders through authentic engagement and shared leadership, deliver excellent service to students and families, and provide educators and staff with professional development and clear expectations... THEN we will become a high-performing, nation-leading district that closes gaps and improves life outcomes for each student." The content of assessments is a crucial component of supporting students and educators in facilitating learning. If items are to support the learning of all students–especially as it pertains to BPS' stated commitment to closing historical opportunity gaps–then BPS needs to administer items that have been shown to meaningfully engage students of all backgrounds with the intended content and to avoid reifying historical opportunity gaps by reinforcing White-centered notions of competence.

> Developing items that seek to connect to and affirm the cultures of students of color and foster their identities as agents of social justice is a way to center anti-racism within large scale test design.

In the dominant paradigm for standardized assessment item-writing, items that contain culturally specific language or potentially sensitive or political content (such as those passages dealing with racism) have been excluded from standardized assessment. Because of this historical practice, items that engage with students' cultures and backgrounds are understudied. In a standardized assessment setting, items are written toward an emic conceptualization of the subject being assessed, with a focus on the common resources that students bring to bear to answer the item. However, when an individual student responds to the item, this event must be understood from an etic perspective that recognizes that every student's background, both in terms of learning and in terms of culture and identity. These individual factors and experiences provide them with a unique set of resources that may or may not align with those that item writers expected students to use to answer the item (Mislevy, 2018). This means that even when items are written in line with current learning science, students will inevitably respond to items in ways unanticipated by item developers.

This study seeks to better understand how the context put forth within the assessment items interacts with students' test taking experiences. If we find that culture does in fact play a role in how students experience and respond to the assessment, this would be evidence that our current models for standardized assessment are unable to effectively meet standards related to standardization in administration conditions, and ultimately, score comparability. The findings of this study have implications for the design of standardized assessments adopted by the district, as well as more broadly in the field of educational assessment.

THEORETICAL FRAMEWORK

Our study draws on the framework for culturally sustaining assessment put forth by Lyons, Johnson and Hinds (2021) where students are valued, engaged, and empowered while participating in an assessment experience. This study attends to all three dimensions of this framework.

We hypothesize that students of color are likely to feel valued and affirmed within an assessment experience when the stimuli and texts are drawn from the literatures and histories of marginalized people and groups, and feature protagonists that resemble the experiences of the students taking the assessment. An assessment can value the students its intended to serve by drawing on the cultural wealth that students bring with them into the classroom, for example the skills and assets associated with speaking more than one language or language variant, being connected to a large familial or community networks, and resisting dominant negative messages about one's own prospects (Yosso, 2005). In this way, culturally sustaining assessment items are designed to recognize the brilliance of Black, brown, and indigenous children—interrupting the marginalization and criminalization of children in the classroom (Stuart-Wells, 2019).

We believe that by connecting directly to students' interests and experiences from their own lives, students are more likely to be engaged in the assessment experience. Traditional notions of bias and sensitivity try to avoid many of the topics that students are likely to find most relevant, and therefore, most engaging to read about. Recently, bias and sensitivity guidelines have become even more restrictive (e.g., removing all references to gender, removing references to migrating animals) in an effort to avoid triggering or offending students (e.g., those with non-binary genders or those who have immigrated)[2]. Traditional orientations to item writing only serve to further marginalize students within the assessment rather than leveraging what we know about how students learn and seeking to connect to and engage minoritized students within the content of the assessment (e.g., by featuring non-binary protagonists or highlighting the experiences of those who have immigrated).

> Traditional orientations to item writing only serve to further marginalize students within the assessment rather than leveraging what we know about how students learn and seeking to connect to and engage minoritized students within the content of the assessment (e.g., by featuring non-binary protagonists or highlighting the experiences of those who have immigrated).

Lastly, and potentially most powerfully within the context of Boston Public Schools, we seek to empower students within the assessment experience. We can empower students through the assessment experience by focusing on the development of the student as a cultural, situated being who has power to confront structural systems of power and oppression (Behizadeh & Pang, 2015; Penuel & Watkins, 2019). Culturally sustaining assessments empower students as agents of change in their lives and communities, to advocate for and advance social justice.

---

[2] See Randall & Rashid (2022) for more criticism of our current approaches to bias and sensitivity reviews.

HYPOTHESIS AND RESEARCH QUESTIONS

The primary research questions for this study are:

1) How can we leverage educator knowledge of their students to develop assessment items that attend to the cultures, identities, and lived experiences of the students in the Boston Public Schools?

2) What impacts do culturally responsive, justice-oriented items have on students' test-taking experiences?

Our hypothesis is that by drawing reading passages from the community (e.g., local museums, student newspapers, prominent graduates) and closely engaging with dedicated teachers to make final selections of passages, we will be able to develop assessments that are truly responsive to the cultures, identities, and lived experiences of the students served in the Boston Public Schools. Secondly, we expect that as a result of engaging with these assessments, students will feel valued and due to the engaging nature and familiarity of the context, students will be able to more easily demonstrate their knowledge and skills (i.e., assessment performance will improve).

CULTURALLY RESPONSIVE ASSESSMENT DEVELOPMENT PROCESS

Passage Selection

Passage finding was the first, and arguably most important, step for developing the culturally responsive assessment for this study. Given our limited time with the volunteer teachers, the researchers engaged in a preliminary passage finding exploration prior to meeting with the BPS teachers so that we could elicit the educators' feedback on our example passages as part of our initial interaction with them. The assessment form we were adapting into a culturally responsive, justice-oriented form was one that contained a set of paired passages around a theme. We curated passages that addressed themes relevant to the everyday lives of students, topics that we thought would be of interest, and written by or from the perspective of young people. Additionally, we prioritized texts that were relevant to the local Boston context. For example, one of the passages we selected was an excerpt from a Boston Globe article that describes a day in the life of a Massachusetts high school student during the latest Omicron surge of the pandemic. This article explored topics including anxiety over illness and schoolwork, disruption due to teacher and student absences, confusion about conflicting messages on masks, and frustrations with inconsistencies in COVID-19 testing protocols. Table 1 provides relevant information about the control and culturally responsive passages that were curated prior to engaging with the teachers from Boston Public Schools.

Table 1. Characteristics of Initial Set of Control and Culturally Responsive Paired Passages

Form	Theme	Excerpt Title	Text Source	Text Author	Excerpt Length	Form Total Word Count	Text Lexile Range
Control	Coral Reefs	Redacted to protect IP			514	901	1010-1200
					387		1010-1200
Culturally Responsive Option 1	Pandemic	From bell to bell, a lesson in fortitude	Boston Globe	Jenna Russell	924	1177	1010-1200
		Someday soon… wearable tech to detect illness	ScienceNews for Students	Silke Schmidt	253		1010-1200
Culturally Responsive Option 2	Climate Change	Unpopular	Speech at the U.N. Climate Action Summit	Greta Thunberg	428	862	1010-1200
		Young Climate Activists	Boston Globe	Zoe Greenberg	434		1210-1400

During our initial meeting with the volunteer Boston Public School teachers, we provided an overview of the planned research study and facilitated a discussion about how this work fits within the larger goal of the district to move toward an anti-racist assessment system. We did this by introducing the framework from Evans (2021) that places the relationship between content and culture on a continuum (see Figure 1), acknowledging that these definitions are necessary but insufficient pre-cursors for achieving BPS' vision for an anti-racist assessment system.

Culturally "_____" Education

SENSITIVE
· Aware that cultural differences and similarities between people exist without assigning them a value.

RELEVANT
· Students' heritage and community cultural practices are resources to honor and explore.
· Teachers take students' everyday lived cultural experiences and make the appropriate linkages to intended learning; teachers are cultural translators and bridge builders.

RESPONSIVE
· Schools adapt to students.
· Teachers see their students as sources of knowledge, full of concepts, and bearers of information.
· Students have many assets that can be more responsively leveraged in order to give students the opportunity to bring their own cultural references and fluencies into learning spaces.

SUSTAINING
· Students' heritage and community cultural practices are resources to honor, explore, *and extend.*
· Youth are producers of cultural as well as consumers.

ANTI-RACIST ASSESSMENT SYSTEM
A system of assessments that is intentionally designed to disrupt racism and racist structures. Anti-racism in assessment attends to both the content and design of the assessments and the interpretation and use of assessments to create a more just future for all.
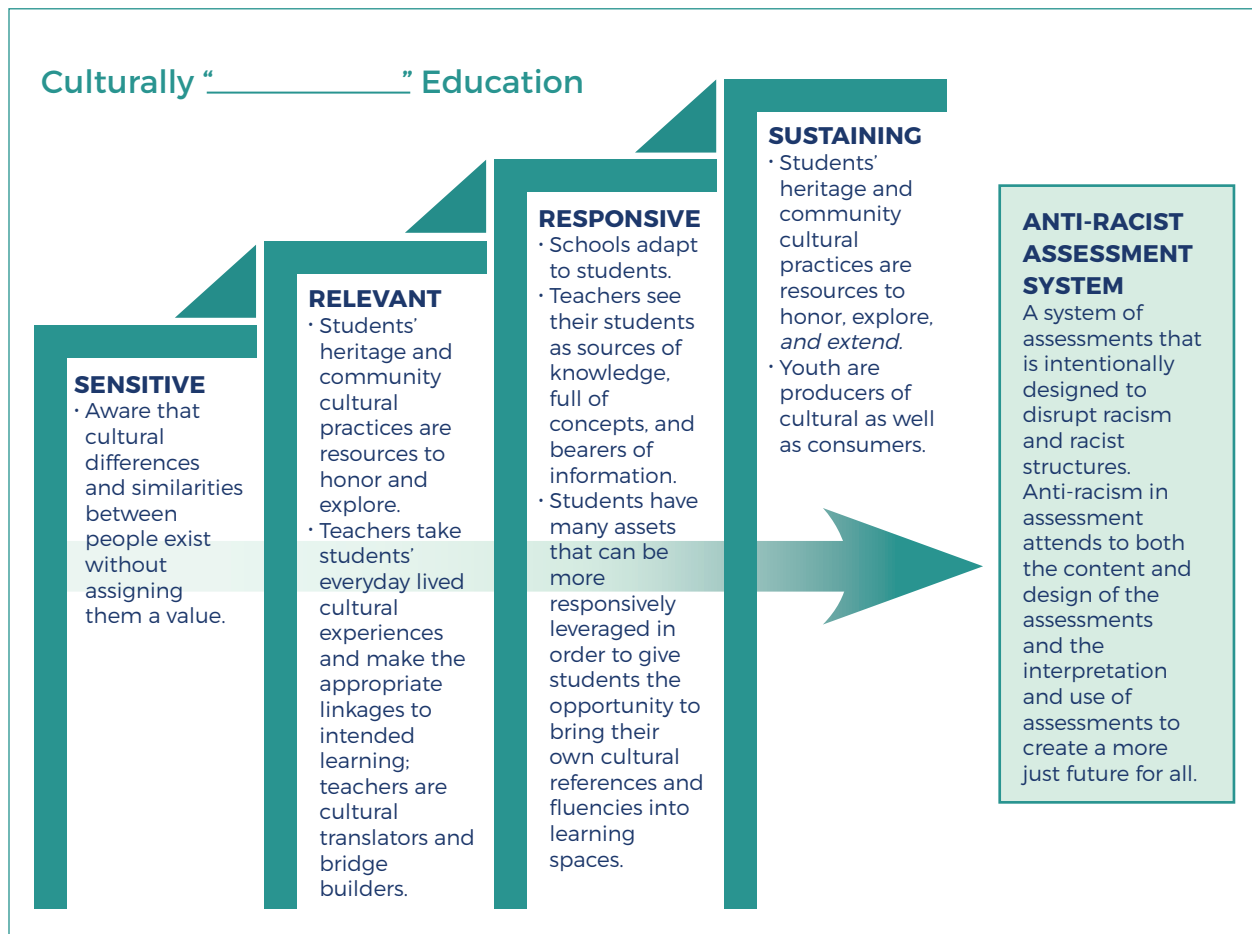
Figure 1. Relationship of Evans' (2021) continuum for culturally relevant education to Lyons et al. (2022) definition of an anti-racist assessment system.

We then directed the teachers' attention to the two options for culturally responsive passages, asking them to keep in mind the following four questions as they read each passage:

1) How likely are your students to find these passages engaging and relevant to their interests?

2) Are your students likely able to relate to the people or contexts presented in the passages?

3) Do the situations and/or experiences presented in the passages reflect your students' lived realities and provide connections to their real lives?

4) Would your students likely see aspects of themselves represented in the passages?

The participating teachers provided incredible depth of insight into the students' interests and perspectives. At every opportunity, they pushed us to move more directly in the direction of addressing issues of racism and social justice. For example, the teachers acknowledged the relevance and familiarity that their students would likely see in the text that excerpts a speech delivered by

> The participating teachers provided incredible depth of insight into the students' interests and perspectives. At every opportunity, they pushed us to move more directly in the direction of addressing issues of racism and social justice.

world-renowned youth climate activist, Greta Thunberg. However, they noted that their students, who are predominantly students of color and a significant portion of which are immigrants from developing countries, might be more interested in thinking about the racial dynamics at play that led to Greta Thunberg being a recognizable face of youth activism over countless other worthy youth activists of color. This question seems particularly relevant given that people of color and those living in developing nations are disproportionately impacted by the negative effects of climate change. Additionally, the teachers recommended we pursue articles that highlight the well-known youth leaders within the Boston Public Schools culture as potential sources of text for these passages.

After meeting with the teachers, the researchers identified two additional options for culturally responsive forms that responded to the teachers' reactions to our initial attempts. The characteristics of the new options are outlined in Table 2.

Table 2. Characteristics of Additional Set of Culturally Responsive Paired Passages

Form	Theme	Excerpt Title	Text Source	Text Author	Excerpt Length	Form Total Word Count	Text Lexile Range
Culturally Responsive Option 3	Generational Shifts in Activism	Speak your truth	Boston Globe	Jenna Russell	715	989	1210-1400
		Gen Z and activism	Bustle	Margaux MacColl	274		1210-1400
Culturally Responsive Option 4	Youth Activists of Color	Climate change is racial injustice	NPR Illinois	Elissa Nadworny & Sequoia Carrillo	465	897	1010-1200
		Meet a young activist of color	Vox	Nylah Burton	432		1210-1400

After an asynchronous review process the teachers ultimately selected the passages associated with Culturally Responsive Option 4: Youth Activists of Color. After seeking permission from the copywrite holders to use the selected excerpts for research purposes, the next step of the assessment development process was item writing.

Item Writing

For the purposes of this research, we had two primary goals for item development: 1) ensure that the item construction is as parallel as possible to the control form, and 2) where possible, lean into an anti-racist orientation to the work. Often, these two notions were at odds with one another and we had to favor decisions that would maintain the viability of the experimental design for making meaningful comparisons in student performance. To do this well, items were developed with a number of considerations in mind to maintain comparability with the control form, including: the assessed standard, the nature of the cognitive task, item difficulty, and the rationale for distractor construction.

An example of how form comparability and anti-racism could be at odds with one another is item six on in the eight-item form. The original item from the control form reads: "Which sentences from '[Redacted

Passage Title Here]' BEST shows that people must work together to save coral reefs?" The parallel item that we initially drafted on our culturally responsive form was "Which sentence from 'Climate Change in Racial Injustice' BEST shows that environmental racism is an important topic for youth of color to speak about." We believe that this item helps move the content of the assessment towards anti-racism as it has students considering both the injustice behind environmental racism and the power that youth, and youth of color in particular, have to make a change in their communities to disrupt existing inequities. However, when we were drafting the item distractors, we were unable to find parallel examples in the text that mirrored the distractor rationales from the control text. To maintain the comparability of our control and test forms to the maximum extent possible, we had to modify our item stem, which resulted in removing the anti-racist leanings in our item. The final item stem reads, "Which sentence from 'Climate Change is Racial Injustice' BEST shows that climate destruction is an important topic for youth to speak about?". While this change is disappointing, it also provides hope that without the limitations of form comparability for research purposes, anti-racist items can feasibly be developed within a standards-aligned, multiple-choice format. This provides an existence proof for a concept that has been greatly doubted and debated within the field of standardized assessment. The final item stems for all eight items in our culturally responsive assessment form are provided in Table 3.

> "Which sentence from 'Climate Change in Racial Injustice' BEST shows that environmental racism is an important topic for youth of color to speak about." We believe that this item helps move the content of the assessment towards anti-racism as it has students considering both the injustice behind environmental racism and the power that youth, and youth of color in particular, have to make a change in their communities to disrupt existing inequities.

Table 3. Item Stems for the Culturally Responsive Assessment Form

Item Number	Item Type	Item Stem/Prompt
1	Selected Response	What is the main idea of "Climate Change Is Racial Injustice"?
2	Selected Response	According to "Meet a Young Activist of Color," how do youth activists MOST help the environment?
3	Selected Response	Which sentence from "Meet a Young Activist of Color" BEST supports the idea that young people of color can make a difference in their communities?
4	Selected Response	Read this sentence from paragraph 4 of "Climate Change is Racial Injustice". *"The episode begins with the idea that, 'climate change is racial injustice,' and focuses on the idea that pollution and the environment impact communities differently."* What does the word <u>injustice</u> mean?

Table 3: continued

5	Selected Response	How do the authors of the texts interpret youth activism in different ways?
6	Selected Response	Which sentence from "Climate Change is Racial Injustice" BEST shows that climate destruction is an important topic for youth to speak about?
7	Selected Response	Read this excerpt from the passage "Climate Change is Racial Injustice." *The students cited in their reporting research that shows climate change, including rising temperatures in cities and poor air quality, disproportionately affects non-white neighborhoods. Black communities face dangerously high levels of pollution and are more likely to live near landfills and industrial plants that pollute water and air and harm quality of life, according to a 2018 report from the Environmental Protection Agency.* Which statement BEST summarizes the evidence in the excerpt?
8	Essay	Write an informative essay that explains why talking about climate destruction is important and why young climate activists of color are crucial to this conversation. Use information from both texts to develop your response. In your essay, be sure to: • Clearly introduce your topic and ideas; • Use evidence from the text to support your thinking; • Include a concluding statement to summarize your response.

METHODS

Planned Quantitative Methods

Due to the devastating impact of the COVID-19 Omicron variant on the students, faculty, and staff of Boston Public Schools in early 2022, we were unable to successfully recruit a sufficient sample size of teachers and students to participate in our study. This means that we had to abandon the mixed-methods experimental design we had originally planned, and instead rely only on a qualitative study with 13 participating students. The originally planned quantitative methods are provided below to provide insight into how we had hoped to answer the intended research questions.

Throughout the year in BPS classrooms, students take Benchmarks for Learning ("Benchmark") assessments. These assessments are designed to align with groups of standards, and teachers have flexibility to administer them at the time at which their students have recently encountered the material that a given assessment covers. This research planned to leverage the current Benchmark assessments as a control for comparing the performance and experiences of students taking the newly developed culturally responsive assessments. For those eighth-grade students randomly assigned to the treatment group, students would have seen the newly developed culturally responsive assessment in lieu of their regularly scheduled standardized assessment. Because the assessments are designed to cover the same standards as the matched Benchmark assessment, the students and teachers would have still received the same educational benefit from the experimental form as the control form.

Students in both the treatment and control conditions would have responded to a short, eight-item survey at the conclusion of the assessment that will ask a series of Likert scale agreement items that gauge the degree to which the assessment experience was reflective and affirming of their cultures, identities, and lived experiences. An example is: "The experiences described in the reading passages were relatable." See Appendix A for the full survey. Survey responses would have been analyzed for reliability using indices of reliability (α and Guttman's λ_6) and for dimensionality using exploratory factor analysis, then aggregated into student scale scores. A four-way analysis of variance (ANOVA) will be used to test the hypothesis that each of these three variables–research condition, test performance, race/ethnicity, and English learner status–influences students' responses to the survey as quantified by their survey scale score. Additionally, ANOVA tests the hypotheses that there are meaningful interactions between the variables–that, for example, students in the treatment condition who scored highly on the assessment were likely to respond positively to the survey above and beyond what would be predicted from each of those variables separately.

Lastly, we would have run an analysis of variance to understand the effect of the culturally responsive assessment on student performance as compared to performance on the typical Benchmark assessment. This analysis will provide insight in to the second half of our second research question that relates to the impact of culturally relevant assessments on student performance. Administrative data will be required to support this analysis including the following elements:

- Student race/ethnicity
- English learner status
- Current and prior scores on the Benchmark for learning assessment from AY 2019-2020 and AY 2020-2021.

For this analysis, we planned to run a similar four-way ANOVA to that described above. In this second analysis, the factors in the ANOVA will be treatment condition, race/ethnicity, English learner status, and prior score on the Benchmark for Learning assessment; the outcome variable is the student's score on the current Benchmark for Learning. The four-way ANOVA again tests the hypotheses that each of these factors and their interactions are meaningful predictors of variance in the outcome.

Implemented Qualitative Methods

Due to our limited sample size of 13 students, we had all participating students engage in the treatment condition (i.e., the culturally responsive, justice-oriented assessment form), and respond to the 8-item Likert scale survey about their perceptions of the assessment. Five of the 13 students also participated in focus groups to help us gain insight into the specific ways in which the newly developed assessment did or did not affect student test-taking experiences. In these focus groups, students were asked to discuss how the reading passages compared to their experiences with other assessments, if they felt valued while reading the passages, and whether they felt that their culture and identity were represented in the passages. The focus groups were semi-structured in order to target our primary research question while retaining the flexibility to support rich conversation about students' unique experiences with the assessment. See Appendix B for the Focus Group Protocol. Focus group discussions were conducted via Zoom during non-academic time, as coordinated with the school leader and teachers.

RESULTS

Survey Results

Figure 2 provides a summary of the 13 student responses to the 8-item Likert scale survey immediately following their engagement in the culturally responsive assessment form. While some items show that the assessment was somewhat successful at meeting its goals related to students identifying with the people in the passages and creating connections between the passages and the students' own lives, the other items reveal more mixed results. Unfortunately, because we had to forego the experimental design with random assignment to a control condition, these results provide no insight into the degree to which the students responses differ from those we might expect from a standard assessment form. Luckily, the focus group data, described in the following sections, provide a better window into the student perceptions of this assessment relative to their more typical Benchmark exams.
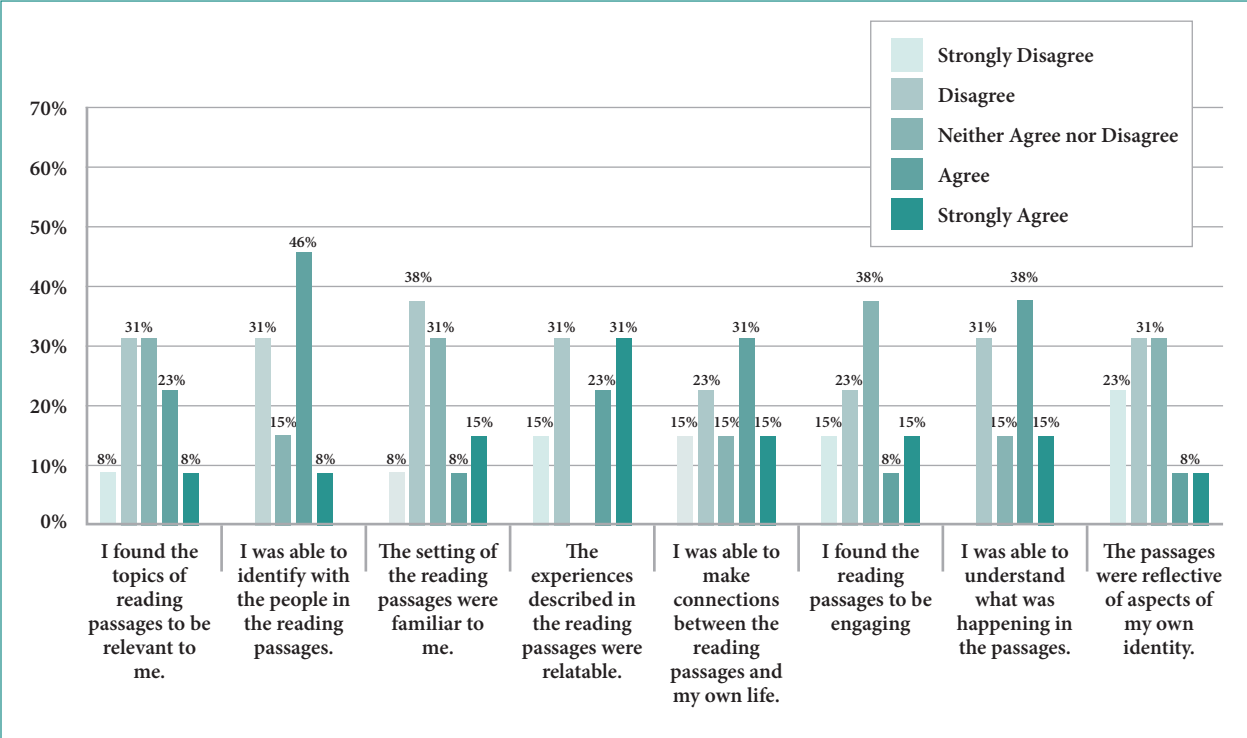

Figure 2. Student Post-Assessment Survey Responses

Focus Group Data Processing

During April 2022, our team, assisted by classroom teachers, conducted two focus groups via Zoom. Altogether, our focus groups included five, eighth-grade students. All five students identified as people of color with four of the students self-identifying as Black. Two research team members led the focus groups using a semi-structured interview protocol developed for this project. An additional team member scribed student responses during each of the 30-minute sessions. The interview protocol encouraged students to recall the topics of the culturally relevant assessment and then reflect on how they differed from other assessment passages they had read in the past. Students were also asked to identify if features of the culturally relevant assessment passages made the passages engaging to read, relevant to their lives, and relatable to students' held identities. Data from both focus groups was combined for the qualitative analysis.

The focus group transcripts were processed in Excel to produce 41 unique excerpts across the five students. Each excerpt was composed of an utterance – a full thought from one student. The cleaned sample was imported into a qualitative analysis software (Dedoose) for coding. Each individual excerpt was coded for the following main themes: Engagement/Interest; Real-world Relevance; Identity; and Other. These main themes were developed from the focus-group interview protocol as students were asked questions related to these areas. A sub-theme, Connection to Lives/Experiences, was derived from the main themes of Real-world Relevance as we wanted to note whether students had experienced these issues in their own lives as opposed to issues that they believed were important but had not yet experienced for themselves. Additionally, two sub-themes were developed from the Identity main theme, once again because they were present in the focus group interview protocol: Relate to Characters and Relate to Contexts. The latter sub-theme was later dropped as it overlapped almost entirely with the Connection to Lives/Experiences sub-theme. Finally, the last main theme, Other, was divided into four sub-themes that emerged from the data: Easy, Confusing, Additional Passage Topics, and Essay. We used the frequency of codes in addition to actual text from the excerpts to draw conclusions about these students' experiences with the culturally relevant assessment.

Focus Group Results

Overall, students reported that they found the culturally relevant assessment passages interesting and described the topics as relevant to their lives. It appears that the Engagement/Interest code and the Real-world Relevant code were closely tied together as excerpts were often coded under both main themes. Thus, students believe that texts that they can connect to their lives *are* engaging/interesting. Specifically, students repeatedly referenced shared identities (between themselves and the subjects of the passages) as a key aspect that made the new passages relatable, interesting, and different from other assessment passages they had read in the past.

In their responses, students found the subjects' Black identities and young ages to be different from what they usually saw in assessment passages. These identities seemed to resonate with the four Black students who saw themselves in the passage subjects. When asked what stood out to them as they reflected on the culturally relevant assessment passages, one student said, "…they were interesting because it kind of made me think of myself. It kinda reminded me of myself 'cuz I'm Black." Several students mentioned that they were able to relate to the characters in the passages because they were "just like me." In response to a question about how the new assessment was different from other ones they had taken, another student replied, "We get to see the perspective of young children and their perspective of what's happening." Seeing themselves in the passage helped students connect the passages to their own lives and increase their engagement.

> When asked what stood out to them as they reflected on the culturally relevant assessment passages, one student said, "…they were interesting because it kind of made me think of myself. It kinda reminded me of myself 'cuz I'm Black." Several students mentioned that they were able to relate to the characters in the passages because they were "just like me."

Most students found the passages more engaging than other assessment passages they had previously read because the culturally relevant passages were about topics that were more relatable. One student

specifically mentioned that the passage about podcasts caught their attention, and students related to descriptions of contexts in which discrimination against Black people is present. When students were asked if they were able to relate to the places in the passages, one student responded "Yes. They were talking about the community and how in other places they don't like [how] they get treated different 'cuz the color of their skin." Moreover, a few students mentioned one of the contexts in the passages (i.e., aging school buildings) reminded them of their schools. This suggests that students can relate to passages that describe communities like theirs and high-interest topics.

Students referenced how the culturally relevant passages compared to other English Language Arts (ELA) content they had previously encountered in school. First, one student mentioned that the assessment texts were like texts they read in school. Although, as previously stated, the new assessment texts seemed to be different from texts they read in other assessments. This suggests that the culturally relevant assessments may support what students are reading and learning in class better than the more traditional assessments they have typically taken. When asked if they thought the pictures that were a part of the culturally relevant assessment added to the interest of the passages, all students responded yes.

Finally, the research team hypothesized that designing an assessment around culturally relevant passages could make the questions easier for students to answer relative to other assessments. One student mentioned that the questions on this assessment were easier because the passages were interesting; however, this was not the consensus. Three students mentioned several barriers they encountered while taking the culturally relevant assessment. First, one student shared that they found the passages to be confusing and needed to re-read them for understanding. Another student mentioned that they would prefer the passages to be shorter, and a different student stated that they just do not like reading so they were not going to find assessment passages engaging no matter the topic. Additionally, several students mentioned that the essay prompt was difficult or not engaging. Towards the end of the focus group, the research team asked students to identify additional topics they would like to see in assessment passages. The topics students mentioned included technology, gaming, Black culture, communities, and happiness.

IMPLICATIONS FOR PRACTICE

Though this study incurred unpredicted disruptions due to the COVID-19 pandemic, we believe that our work in culturally responsive assessment development with the educators within the Boston Public Schools has important implications for research and practice. The first major implication of the work is the existence proof that items with anti-racist messages can be developed using standard item formats such as selected response and essay items and do impact student experiences. Students in the focus groups reported both higher interest and engagement in the assessment because of its connection to who they are as youth of color. The students reported experiencing culturally responsive instruction, but this concept seemed to be a new experience for assessment and represents a potential area of growth for BPS to move more fully towards implementing its strategic plan.

Secondly, the assessment development process used for this study underscored the central importance of consulting with community members such as educators, students, or families to gain a strong sense of the topics and areas of interest that are relevant, engaging, and reflective of the lives of the students who will be taking the assessment. Without this information, assessment developers run the risk of making false assumptions about students, essentializing their experiences or cultures, or missing important aspects of

their identities. Therefore, Boston Public Schools and others purchasing assessment programs should not request pre-identified culturally relevant assessments from their vendors, but instead expect to engage their communities and educators in the assessment development process alongside vendors who wish to be responsive to their students and contexts. With these findings, we look forward to continuing to advance research and practice related to culturally responsive assessment within Boston Public Schools and beyond.

REFERENCES

Behizadeh, N., & Pang, M. E. (2016). Awaiting a new wave: The status of state writing assessment in the United States. *Assessing Writing, 29,* 25–41.

Evans, C. (2021, November 3). *Culturally sensitive, relevant, responsive, and sustaining assessment* [Blog post]. Centerline. Retrieved from: https://www.nciea.org/blog/culturally-responsive/culturally-sensitive-relevant-responsive-and-sustaining-assessment

Lee, C. D. (1998). Culturally responsive pedagogy and performance-based assessment. *Journal of Negro Education*, 268-279.

Lyons, S., Johnson, M., & Hinds, F. (2021). *A call to action: Confronting Inequity in Assessment.* Lyons Assessment Consulting.

Mislevy, R. J. (2018). *Sociocognitive foundations of educational measurement.* Routledge.

Moll, L. C., Amanti, C., Neff, D., & Gonzalez, N. (1992). Funds of knowledge for teaching: Using a qualitative approach to connect homes and classrooms. *Theory into practice, 31*(2), 132-141.

National Academies of Sciences, Engineering, and Medicine. (2018). *How people learn II: Learners, contexts, and cultures.* National Academies Press.

Penuel, W. R., & Watkins, D. A. (2019). Assessment to promote equity and epistemic justice: A use-case of a research-practice partnership in science education. *The ANNALS of the American Academy of Political and Social Science, 683*(1), 201–216.

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., & Rashid, M. (2022, April). *Re-envisioning the bias and sensitivity review process from a culturally sustaining and antiracist perspective.* Paper presented at the 2022 Annual Meeting of the American Educational Research Association, San Diego, CA.

Resnick, L. B., & Resnick, D. P. (1992). Assessing the thinking curriculum: New tools for educational reform. In *Changing assessments* (pp. 37-75). Springer, Dordrecht.

Stuart-Wells, A. (2019, April). *An inconvenient truth about the new Jim Crow of education.* Presidential Address at the Annual Meeting of the American Educational Research Association, Toronto, Canada.

Yosso, T. J. (2005). Whose culture has capital? A critical race theory discussion of community cultural wealth. *Race ethnicity and education, 8*(1), 69-91.

APPENDIX A:
STUDENT SURVEY

Please select the degree to which you agree with each of the following statements about the Benchmark assessment you just took.

1.	I found the topics of the reading passages to be relevant to me.	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
2.	I was able to identify with the people in the reading passages.	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
3.	The settings of the reading passages were familiar to me.	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
4.	The experiences described in the reading passages were relatable.	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
5.	I was able to make connections between the reading passages and my own life.	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
6.	I found the reading passages to be engaging.	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
7.	I was able to understand what was happening in the passages.	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
8.	The passages were reflective of aspects of my own identity.	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree

APPENDIX B:
FOCUS GROUP PROTOCOL

This focus group protocol will be used to guide semi-structured small group interviews via Zoom with students shortly after they have taken the treatment condition of the Benchmark assessment.

1. Do you remember the reading passages that were included in your most recent Benchmark assessment? Can you tell me about them?

2. Did you find these passages to be engaging? Why or why not?

3. Were you able to relate to the characters in the passages? How so?

4. What about the passages was familiar to you? Please explain.

5. What about the passages was confusing or new?

6. Were these passages different than the passages you normally read on the Benchmark assessments? How so?

7. Do you think those differences made it easier to connect with what was happening in the text? Is that helpful for answering the assessment questions about the text?

8. How did the images provided with the passages impact your experience, if at all?

9. What could we do to develop assessments that are even more relevant to you? Any ideas?

Lyons

ASSESSMENT
CONSULTING

susan@lyonsassessment.com
www.lyonsassessmentconsulting.com