**Reframing Research and Assessment Practices: Advancing an Antiracist and Anti-ableist Research Agenda**

Angela Johnson

American Institutes for Research

Elizabeth Barker

CAST

Marcos Viveros Cespedes

Oregon State University

Corresponding Author:

Angela Johnson

[angelasun@alumni.stanford.edu](mailto:angelasun@alumni.stanford.edu)

1400 Crystal Drive, 10th Floor

Arlington, VA 22202-3289

**Abstract**

Educators and researchers today strive to build informed policies and practices on data and evidence, especially on academic achievement scores. When assessment scores are inaccurate for specific student populations or when scores are inappropriately used, even data-driven decisions will be misinformed. To maximize the impact of the research-practice-policy collaborative, every stage of the assessment and research process needs to be critically interrogated. In this paper, we highlight the need to reframe assessment and research for multilingual learners, students with disabilities, and multilingual students with disabilities. We outline a framework that integrates three critical perspectives (QuantCrit, DisCrit, and critical multiculturalism) and discuss how this framework can be applied to assessment creation and research.

**Acknowledgements**

The authors are grateful for helpful feedback from Kim Baker, Ayesha Hashim, Nate Jensen, Teresa Krastel, Megan Kuhfeld, Karyn Lewis, and Erik Ruzek.

**Reframing Research and Assessment Practices: Advancing an Antiracist and Anti-ableist Research Agenda**

Educators and researchers today strive to build informed policies and practices on data and evidence, especially on academic achievement scores. However, when assessment scores are inaccurate for specific student populations or when scores are inappropriately used, even data-driven decisions will be misinformed. To maximize the impact of the research-practice-policy collaborative, every stage of the assessment and research process needs to be critically interrogated. In this paper, we highlight the need to reframe assessment and research for multilingual learners (MLL)[[1]](#footnote-1), students with disabilities (SWD)[[2]](#footnote-2), and multilingual students with disabilities. Historically, research on MLL and SWD tended to show large test score gaps against monolingual students without disabilities (e.g., Carnoy & Garcia, 2017; Gilmour et al., 2019) and position MLL and SWD as at-risk student populations. However, the validity of assessment scores for MLL and SWD used for research and reporting is seldom examined. While practically all technical documents for standardized assessments provide information about validity and reliability for monolingual students and students without disabilities, few if any separately report such evidence for MLL and SWD or describe the inclusion of MLL and SWD in the test creation and norming process. That test scores accurately reflect the true knowledge, skills, and abilities for MLL and SWD is often an untested assumption.

The validity of test scores for MLL and SWD is critical because test scores drive many high-stakes decisions, such as eligibility for services and access to curriculum (Wu et al., 2021). Perceived test score disparities between MLL and SWD and their monolingual peers without disabilities shape opportunity gaps that persist from kindergarten to higher education. MLL and SWD are severely under-identified for gifted and talented programs (e.g., King, 2022; Peters et al., 2019), take fewer core content and advanced courses in high school (e.g., Johnson, 2019), and pursue postsecondary education at much lower rates (e.g., Kanno & Cromley, 2013; Campanile et al., 2022). Addressing assessment score validity is the first and most important step to ensuring that students receive appropriate supports and access academic materials.

In addition to using uninterrogated assessment scores, traditional approaches to research are problematic in that they are complicit in or can contribute to minoritizing and marginalizing students such as MLL and SWD. Even among works conducted by researchers with the best intentions, some have been framed in ways that assume that the problem lies with the students, their families, or their communities rather than with greater social or structural issues (Smith, 2021). From complete exclusion to within-school segregation, the history of the US school system has been a history of oppression for MLL and SWD. By centering on affluent, White, monolingual students without disabilities, traditional research approaches treat MLL and SWD as peripheral, at-risk, and needing to be fixed (Patel, 2015). This framing reinforces the illusion of meritocracy within an educational system that normalizes the achievement of the dominant group and sets up failure for marginalized student populations.

Until we concretely name systemic oppression as a problem, we will be limited in our capacity to effectively advance a more just research agenda that tackles multiple disparities in education. Racism and ableism are woven together in education research and practices through policies, curricula, and interactions that create and maintain narrow expectations of what it is to be “normal;” this sustains White and non-disabled standards as the most desirable and discriminates against populations perceived as different (Annamma & Handy, 2019; DeMatthews, 2020). There is an urgent need for research and assessment that consider how racism and ableism are produced, reproduced, and enacted within the education system.

In this paper, we outline a framework that integrates three critical perspectives (QuantCrit, DisCrit, and critical multiculturalism) and discuss how this framework can be applied to assessment creation and research. First, we describe our critical framework and positionalities. Then, we summarize the historical and current policies and practices, including assessment and research, that place MLL and SWD at the intersection of multiple axes of oppression within the education system. Finally, we provide recommendations grounded in critical theories for assessment design and administration, data reporting and interpretation, and research practices.

**Critical Framework**

We propose reframing research and assessment grounded in the tenets of QuantCrit, DisCrit, and critical multiculturalism. In doing so, we hope to re-envision methodologies to effectively advance equity and disrupt oppressive systems during the research and assessment processes and beyond. We chose these three theories because of their joint application to our own research and our positionalities as researchers and persons. We are a team diverse along many dimensions, and we are keenly aware of how these backgrounds, experiences, and identities shape our own work.

**QuantCrit**

QuantCrit, or quantitative critical race theory, encourages researchers to use numbers for social justice and equity (Gillborn et al., 2018). Its first principles focus on the centrality of racism in society and the inability for quantitative data and statistical inquiries to speak for themselves or render neutral or natural representations of race and other social categories (Gillborn et al., 2018). This framework recommends that researchers reflect on their own positionalities, frame research from an asset-based perspective for understanding marginalized students and communities, and make purposeful decisions throughout the research process to make sure categorizations are informed and the perspectives of the communities of interest are represented (Castillo & Gillborn, 2022). Researchers are urged to consider the difference between interpretation and result and critically examine ways of measuring outcomes and representing data. In this paper, we apply QuantCrit in two ways: (1) to affirm the need for asset-based framing, and (2) to interrogate categorizations of students employed in previous research. As we detail in the next section, traditional assessment and research practices tend to center measurement and data analysis on White, English-monolingual, and non-disabled student populations, resulting in a deficit-framing of MLL and SWD academic achievement (e.g., Carnoy & Garcia, 2017). Drawing on QuantCrit principles, we propose assessment and research practices that center on SWD and MLL.

**DisCrit**

DisCrit, or disability critical race theory, argues that racism and ableism are always present and working in interconnected ways, particularly in educational contexts (Annamma et al., 2013). Influenced by works on intersectionality (Crenshaw, 1991), this framework values multidimensional (rather than singular) identities and prioritizes voices of marginalized populations. DisCrit emphasizes that race and ability are socially constructed categories that have material and psychological impacts on individuals. We draw on DisCrit in our analysis of research and assessment approaches and their consequences. Specifically, we argue that racism and ableism collude in shaping opportunity disparities in the system. This results in a disproportionate representation of students of color, including MLL, in certain disability categories. Research has found racially and linguistically minoritized students to be both over- and under-identified for disabilities, depending on the age and disability category. For example, Black, male students tend to be over-identified for emotional, behavioral, and intellectual disabilities or specific learning disabilities, and MLL are often over-identified for speech or language impairment (e.g., Dever et al., 2016; Vougarides et al., 2017). While over-identification places students at risk of exclusion from opportunities afforded by the general education context, under-identification results in a denial of necessary services and supports. In light of these risks, we highlight the need to critically interrogate the role of racism and ableism in assessment for identifying SWD. Further, we argue that research and data reporting need to reflect students’ multidimensional identities and needs, such as the intersection of eligibilities for language services and special education.

**Critical Multiculturalism**

Rooted in intersectionality, critical race theory, and critical pedagogy perspectives, critical multiculturalism seeks to addresses issues of racism, ableism, sexism, classism, and other systems of oppression (May, 1999). Unlike surface-level multiculturalism, which tends to focus on celebrations and can be fraught with stereotypes, this approach argues that “focusing on one axis of oppression offers only a partial analysis” (May & Sleeter, 2010, p. 10). Examining racism or ableism alone is insufficient. Rather, we need to address the system that marginalizes people based on their multiple identities and people’s experiences of intersectionality. Multilingual students with a disability face unique challenges in the school system. For example, although it is common practice to provide only language support or only disability support to students, this “one-at-a-time” approach fails to address students’ intersection of needs (Cioè-Peña, 2020; Park, 2014). To conduct an analysis that recognizes the multidimensionality of identities and interconnection between systems, we use critical multiculturalism perspectives to supplement principles from QuantCrit and DisCrit.

**Positionality Statements**

As researchers, we are in a position of power and privilege from which we can disrupt systemic oppression and reframe research for the benefit of all students. Positionality represents the researcher’s world view in relation to the social and political context of the research. Positionality affects every phase of the research endeavor, from the way the research problem is initially constructed and designed to how the findings are disseminated (Coghlan & Brydon-Miller, 2014). Therefore, we operate from this perspective, recognizing our positionalities and biases to critically reflect on the nuances and implications that our work can have on others.

Johnson is a first-generation immigrant and a former English Learner (EL). As a student who received and then exited EL services in US public schools, she attended both fully-segregated and English-only instructional environments. Her research and perspectives are also informed and shaped by a decade of training and practice as an ESL instructor in higher education and an English proficiency test developer. Johnson acknowledges that her views are influenced by privileges associated with her US citizenship and non-disabled status.

Barker is a person with a disability. Diagnosed with a specific learning disability in both reading and writing, she was eventually recognized as dyslexic and placed on an Individualized Education Plan. As a child of the 1980s, she experienced a range of policy and landmark law changes, such as the shift from Public Law 94-142 and the reauthorization of the Individuals with Disabilities Education Act (IDEA), which had a large impact on her ability to achieve an advanced degree. These experiences have influenced her teaching, her research, and the perspective she has on education and assessment today.

Viveros Cespedes is a first-generation college student, immigrant, and Hispanic whose first language is different than English. Upon moving to the US in 2014, he began to realize that his Hispanic roots made him a part of a historically marginalized community in the US. Moreover, Viveros Cespedes also identifies as gay, which has made him a part of another historically oppressed community. His intersecting identities and his experiences as a scholar have encouraged him to focus his teaching and research on disrupting systemic oppression through the lens of critical theory approaches and critical multiculturalism.

**MLL and SWD at the Intersection**

Studies over the past two decades have broadened our understanding of the mechanisms by which schools systematically marginalize and minoritize MLL, SWD, and multilingual students with disabilities (e.g., Brayboy et al., 2007; Harper, 2012; Solórzano et al., 2002) based on race, income, disability, language, and immigration status. Operationalizing critical multiculturalism, we need to consider how students’ intersecting identities affect their access and outcomes in the education system.

According to 2014–15 data from the U. S. Department of Education, MLL currently developing English proficiency[[3]](#footnote-3) are disproportionately from racially minoritized groups. Hispanic students constitute 75% of MLL and only 25% of the overall student population, and Asian students 11% and 5%, respectively. MLL also comprise a larger fraction of migrant (39%) and homeless (15%) students and students attending Public Title I Schoolwide Programs (14%) than their overall representation (10%). Thus, MLL status overlaps almost entirely with racial minoritization and correlates heavily with poverty, homelessness, and other socioeconomic disadvantages. As language is not a protected class, racism, classicism, and ableism can marginalize students by operating through language policies and practices (Kibbee, 2016).

The social construction of disability is similarly interconnected with race, language, and socioeconomic strata. Students of color, MLL, and children from low-income families are disproportionately represented in special education in several ways (Cruz et al., 2018; Morgan et al., 2017; Thorius, 2019). First, their placement in special education often far exceeds those of their White peers (Thorius & Stephenson, 2012; Thorius, 2019). Black and Brown students, students experiencing poverty, and MLL are also overrepresented in emotional and behavioral disabilities, specific learning disabilities, cognitive disabilities, and speech and language impairments (e.g., Devers et al., 2016; Voulgarides et al., 2017). Potential factors contributing to racial and linguistic disproportionality in special education include but are not limited to (a) assessments and identification procedures that fail to distinguish disabilities from developing English language proficiency (Umansky et al., 2017), (b) implicit bias against marginalized groups, and (c) deficit ideology at the systemic level (Lewis & Diamond, 2015).

Exclusionary education policies that marginalize MLL and SWD have very real material and psychological implications for students. On one hand, underrepresentation, a situation in which students are errantly not provided with essential services, can be harmful (Voulgarides & Aylward, 2022). On the other hand, if a student is misdiagnosed with a disability, they may be placed in inappropriate educational settings, receive instruction that do not match their education needs, and experience emotional harm (Agran et al., 2020; Raj, 2016). Students of color are more likely to be incorrectly identified and, subsequently, may be placed in restrictive and often segregated environments, and denied access to general education (Dever et al., 2016). The restrictive placements resulting from identification in certain disability categories could be emotionally damaging or have implications for future opportunities, such as college entrance. SWD, especially students with emotional disturbance, are overrepresented in the correctional system (Rutherford et al., 2002). Thus, over-identification may also contribute to the school-to-prison pipeline (Kim et al., 2010).

MLL and SWD oppression comes not only from structural inequalities within the education system but also from erasure or devaluation in Western, positivist research traditions and methodologies (Kauffman et al., 2018). These research approaches often position White, middle- and high-income students and their communities as the norm and hold all other populations accountable for failing to attain the same academic or socioeconomic standards. As a result, “multilingual” and “disabled” have become synonymous with problematic, at risk, and in need of intervention (Patel, 2015). Instead of addressing the social dysfunctions that positioned MLL and SWD on the lower strata of achievement, some research focused on short-term interventions that provide experiences that mimic those of affluent, White, monolingual students without disabilities; in doing so, research becomes complicit in societal and institutional efforts to deny historical formations of inequality and oppression (Smith, 2021). Contrary to this approach, research that champions longer-term interventions that integrate specific supports for MLL and SWD merits acknowledgement. For example, a body of literature is developing on the impacts of dual language education programs, which provide academic content instruction in English and a partner language for six or more years (e.g., Steele et al., 2017). For students with learning disabilities or dyslexia, scholars continue to extend research on brain activation underlying reading fluency (e.g., Ozernov-Palchik et al., 2022) and the science of reading (e.g., Georgiou et al., 2021; Gough & Tunmer, 1986).

Despite empiricist and positivist claims, research is not a distant, neutral, objective, and purely academic endeavor. Instead, research is a product of culture and sociopolitics embedded in a system built on coloniality, violence, and oppression (Patel, 2015; Smith, 2021). As such, research has actively participated in the reproduction of power dynamics and systemic racism, even without intent or realization by the researchers. In the sections that follow, we outline the ways in which traditional approaches to testing, reporting, and research contribute to the exclusion, unfair assessment, and deficit-framing of SWD and MLL within the current education system.

**A History of Exclusion in Research and Assessment**

Section 504 of the 1973 Rehabilitation Act was the first disability civil rights law enacted in the US that had an impact on assessments (Jonson & Geisinger, 2022). Section 504 prohibits discrimination against people with disabilities in programs that receive federal funding (deBettencourt, 2002) and acknowledges that SWD need access to assessments. However, prior to the 2000s, separate reporting by SWD status was uncommon in research, particularly in studies that focused on the effects of high-stakes assessments. Mentions of SWD who are developing English proficiency was even rarer. Researching achievement outcomes, discriminatory patterns, and test fairness with respect to SWD has been a recent endeavor. The reason for the lack of research was that very few states had included SWD in their assessment data until the late 1990s (McGrew et al., 1993). In 1997, amendments were made to IDEA to ensure that SWD were included in accountability programs. This reauthorization included the requirement for SWD to have access to the same high standards as their non-disabled peers (Wolfe & Harriott, 1998).

Currently, IDEA requires participation of all students, including SWD, in all state and district administered assessments, and to provide accommodations when appropriate. Additionally, the Elementary and Secondary Education Act (ESEA) confirms the participation of SWD and requires the participation of MLL in state administered assessments. In 2015, with the reauthorization of ESEA, IDEA mandates reporting the number of students using accommodations, including SWD and MLL developing English proficiency, for both academic content and English language proficiency assessments (Lazarus & Thurlow, 2016).

Prior research that focused on MLL tended to report whether students currently require language support, separating students eligible to receive services from students with fluent or native English proficiency, with no further disaggregation within the population of MLL developing English (e.g., Carnoy & Garcia, 2017). This is because accountability systems prior to the Every Student Succeeds Act (ESSA, 2015) did not distinguish between subgroups of MLL. This is problematic for several reasons. Students who currently require language services are a diverse group with various linguistic and academic strengths and needs. Newcomer immigrant students differ substantially from US-born students who continue to need services after several years (Johnson, 2019). Thus, reporting needs to reflect the heterogeneity within the students receiving language services to accurately capture their diversity. This is consistent with the tenets of both QuantCrit and DisCrit. Comparing the academic achievement of students developing English proficiency to fluent or native English users can also result in inflated score differences (Kieffer & Thompson, 2018). Students exit language services on a regular basis after attaining high levels of English proficiency (and, in some states, academic achievement). By definition, students who remain eligible for services are students with low test scores; therefore, test score disparities between the two groups will always be large (Abedi, 2004; Saunders & Marcelletti, 2013). Focusing on these gaps alone diminishes the linguistic and academic progress made by MLL as a group. As required by ESSA (2015), schools must now report multilingual students’ progress not only on academic achievement but also English proficiency, and schools must report separately for subgroups, such as MLL with disabilities.

**Setting up “Achievement Gaps”: Inappropriate Assessment Practices and Data Use**

The goal of educational assessment is to collect valid data on students’ achievement and progress and enable data-informed decisions for improving the learning outcomes and experiences of all students. Commonly perceived as neutral, assessment plays a role in reproducing educational and social inequities. Assessments and assessment systems are cultural products that “contain practices and artifacts that reflect culturally-determined world views, sets of knowledge and skills valued, and ways of representing information, building arguments, and asking and responding to questions” (Solano-Flores, 2019, p.1). Assessing students from diverse cultural (e.g., linguistic, socioeconomic, racial/ethnic, disability) backgrounds with the same instrument is likely to result in scores that reflect cultural differences and penalize marginalized student populations, producing questions and challenges related to score interpretation, fairness, and validity (AERA, APA, & NCME, 2014; Solano-Flores, 2019).

A major criticism of standardized achievement tests has been the exclusion of MLL and SWD from the norming group for these tests (e.g., Abedi, 2004). Few technical documentations specify to what extent and how MLL and SWD participated in field testing, if their teachers, families, and communities were consulted in test development, and if MLL and SWD were included when score benchmarks were established. Thus, whether or not test scores for MLL and SWD should be compared to monolingual students without disabilities remains unclear. For example, Ortiz and Wong (2022) pointed out, “Until and unless the differences in language development that exist among ELs even in their heritage language are stratified and controlled for in the creation of a normative sample, the resulting test scores cannot be presumed to be fair or valid… testing with ELs will remain—as it has always been to this point—nothing more than an illusion that gives the appearance but not the reality of fairness” (p. 315, 328).

Traditionally, test developers prioritized standardization to reduce variation that would advantage some students and disadvantage others (Mislevy et al., 2013; Sireci, 2020). The idea was that presenting the same surface features (e.g., language and appearance of the test materials) and the same test procedures (e.g., instructions, time allowed) would maximize test “fairness.” Psychometricians worked to address threats to validity, such as bias, all while centering the standard assessment on monolingual students without disabilities (Jonson & Geisinger, 2022). In their technical reports, few standardized tests provide evidence about validity specifically for MLL and SWD. Many tests are administered in English and often without accommodations due to concerns about validity, namely, that accommodations would offer SWD and MLL unfair leverage over their peers. For example, text-to-speech (TTS) has been shown to assist dyslexic students in demonstrating text comprehension abilities without affecting the performance of students without dyslexia (Koak-Babuder et al., 2019); yet state policies vary on its use (Lazarus et al., 2021). However, this perspective fails to consider threats to score validity when accommodations are necessary but denied.

Recent scholarship pointed out that the traditional definition and operationalization of test fairness is misinformed (AERA, APA, & NCME, 2014; Jonson & Geisinger, 2022). One flawed assumption typically made in the traditional approach is that the population is homogeneous (Sireci, 2020). As Solano-Flores (2019) stated, “Underlying the notion of standardization in tests is the implicit assumption that all individuals being tested with the same instrument share the same set of cultural experiences and are equally familiar with the features of items” (p. 1). In reality, equivalent surface conditions, such as the language of test administration, do not yield equivalent evidence for knowledge and skills for a diverse student population (Mislevy et al., 2013). Assessments constructed and normed for monolingual students and students without disabilities have been shown to have low reliability and validity for MLL and SWD (e.g., Abedi et al., 2000; Abedi et al., 2003; Kettler & Dembitzer, 2022).

IDEA requires that SWD be included in state assessment programs with appropriate accommodations. However, under the No Child Left Behind Act, whether tests were taken with accommodations was not systematically documented, and whether the resulting scores were accurate measures of SWD achievement is unknown due to a lack of research (Geisinger, 2007; Thurlow & Larson, 2011). Because of the previous lack of SWD data on accommodations, many state assessment accommodation policies were built from theories or opinions rather than empirical data. For example, paper/pencil braille was considered an acceptable accommodation (which does not alter the construct being measured) whereas read-aloud was considered a modification (a change to the construct being assessed). The lack of appropriate accommodations can hinder students’ ability to understand test items and demonstrate their skills, creating construct-irrelevant variance (Abedi et al., 2003; Abedi, 2004; Jonson & Geisinger, 2022; Messick, 1995; Thurlow, 2022). For MLL and SWD, English tests taken without appropriate accommodations can result in scores that do not reflect students’ true skill level or ability, but instead may reflect lower bounds or underestimates.

**Data Use**

When researchers analyze test score data without interrogating the assessment itself, their findings perpetuate the inequities embedded in the assessment system. Historically, data analysis, like data collection, also conformed to placing monolingual students without disabilities at the center. Using underestimates of MLL and SWD achievement to compare to the average scores of their monolingual, non-disabled peers, traditional research tended to generate stark “achievement gaps” that are one standard deviation or larger (e.g., Carnoy & Garcia, 2017; Gilmour et al., 2019). Given how the assessments set up MLL and SWD to fail, these results are hardly surprising but misinformed. In other words, MLL and SWD may be penalized twice—first by the assessment process during which they cannot demonstrate their full knowledge and skills, and then by the research process during which they are treated as “other” and “abnormal” by the imposed standards of the dominant, monolingual, non-disabled majority. The deficit framing of MLL and SWD, based on artificially inflated estimates of “achievement gaps,” exerts real, negative impacts on students’ academic opportunities and outcomes. This is especially problematic when test scores are used to make high-stakes decisions regarding students’ access to academic content (Estrada & Wang, 2018) and eligibility for programs and services, such as gifted and talented education (Peters et al., 2019), extended school year, and scholarships. For example, extended school year is an aspect of free appropriate public education that is specifically for SWD who qualify for special education under IDEA. Students are usually recommended for services on an individual basis (Tatgenhorst et al., 2014), often with consideration of their test scores. If test scores do not accurately reflect students’ achievement or learning gains, or if scores are interpreted in misguided ways, students who need extended school year may be denied the service, while students who do not need the service may be given the service.

As another example, historical and some current approaches to identifying disabilities use Intelligence Quotient (IQ) tests, which have been shown to be biased against people of color (Terzian, 2021). Prior to IDEA 2004, students were identified as having a learning disability if there was a discrepancy between their IQ test score and their academic proficiency. The validity of this “discrepancy model” hinges on the validity of both the IQ test and the academic achievement test for the population of students with specific disabilities; yet the validity evidence for IQ tests and academic achievement tests specifically for SWD is severely lacking.

One concern with using IQ test scores and the discrepancy model is that very few standardized tests include SWD in their design and norming processes or clearly state implications for validity. Another concern involves the Flynn effect, which refers to the gradual and systematic rise in standardized IQ and some achievement test scores since the mid-20th century (Kanaya et al., 2022). As overall scores rise, the inflated mean is reset back a few points with every wave of norming. For a stylistic and hypothetical example, consider a test normed in 2010 with a mean of 100 points; the mean score rises from 100 to 105 in the five years that follow so that in 2015, the test is re-normed and the mean (now 105 according to the old norm) is set back to 100, and cut scores for diagnoses are adjusted accordingly. Over a long period of time, this Flynn effect might minimally impact students at the center of the score distribution, but implications can be more substantial for students at the lower tail and for students whose disability diagnoses depend on the cut scores. Cut scores are often used to distinguish types of disabilities, and moves in the cut score due to the Flynn effect can artificially increase the number of diagnoses for learning disabilities and decrease the number of diagnoses for intellectual disabilities, regardless of actual student needs (Kanaya et al., 2022). Yet another concern with using IQ test scores and the discrepancy model for determining disability and service qualification pertains to reading. Regardless of IQ and regardless of whether a discrepancy exists between IQ and achievement scores, students could still struggle with reading (Tanaka et al., 2011). Rather than relying on test scores alone for diagnoses and risking under-identification, students who struggle with reading should be provided with supports or interventions (Hutchinson et al., 2021).

Although the claim that intelligence is quantifiable, inherited, and fixed over a person’s life has been challenged and the bias and racism embedded in traditional IQ tests exposed over the years, there are still no guidelines for adjusting tests and test administration conditions to promote more valid disability assessment for historically marginalized groups (Sireci, 2020). While the reauthorization of IDEA 2004 stated that criteria “…must not require the use of a severe discrepancy between intellectual ability and achievement for determining whether a child has a specific learning disability as defined in 34 CFR 300.9(c)(10)” and despite the absence of empirical support for using IQ tests, many schools and states still implement the discrepancy model.[[4]](#footnote-4) Specifically, 39 states allow local education agencies to continue using the discrepancy model as an option for identification (Schultz & Stephens, 2023).

Relying on test scores without validity evidence for MLL and SWD, educators may overlook students’ linguistic and academic strengths and primarily consider labels such as “language” or “disability,” thus placing students in learning environments that limit their exposure to rigorous academic discourse and materials. Students can also respond to these diminished expectations by internalizing their status as low-proficiency learners of English and/or perceive themselves to be less intelligent than their peers (Kibler et al., 2018; Thompson, 2015). In this system that fails to provide appropriate supports, students who do not neatly fit into learning environments designed in service to a White, monolingual, non-disabled population must bear the onus of their own academic progress. When interpreting data for MLL and SWD, a question often asked is, “Is it language or disability?” (Kangas, 2021, p. 678), implying that either the student’s multilingual status or disability must be responsible for the failure. The onus of achievement and progress is shifted away from the disabling mechanisms within the educational system and placed entirely on the student. In relation to an ideology of normalcy, for which a command of the English language is a prerequisite, MLL with disabilities are perceived as doubly disabled (Cioè-Peña, 2017).

**Reframing Assessment and Research**

We believe that within education, assessment, research, and instruction are intertwined. These processes need to be aligned to the goals of equity and social justice to effectively address systemic oppression. Many educational policies aim to ensure access to equitable educational opportunities; however, efforts can remain theoretical if partners are not equipped with skills and tools for implementation. Below, we propose reframing approaches, including concrete steps, for (a) designing and administering assessments to improve validity and fairness, and (b) reporting and using data for instruction and policy-making.

**Assessment Design and Administration**

To address equity in assessment, first we need to recognize that assessment is subjective to the extent that its design and implementation are influenced by the experiences and beliefs of assessment creators and users. All tests are “inevitably cultural devices” (Solano-Flores & Trumbull, 2003, p.9), as the assessment process involves various organizations and individuals, each with their own sociocultural identities, positionalities, and biases (Heiser et al., 2018). Individually, every participant needs to reflect critically on their views and experiences (May & Sleeter, 2010); collectively, they must incorporate understandings of diverse languages and cultures throughout the entire assessment process (Solano-Flores & Trumbull, 2003). Another element that plays a role in the equity of assessment is the administration. Many decisions for test administration and use are made at the local level with insufficient guidance for implementation from state education agencies and assessment providers (Boyer & Landl, 2021).

Fairness requires that the contexts of test tasks be sufficiently familiar, appropriate, and accessible to all students. Currently, most assessments have accommodations. However, the research and design of accommodations are often after-the-fact: tests with uniform content are created first, then accommodations are retrofitted in an ad hoc manner (Mislevy et al., 2013). In the decade after the seminal work by Mislevy and colleagues, new assessments were built with universal design approaches, refining constructs being assessed and improving platforms that deliver the assessments. Universal design (UD) is “the design and composition of an environment so that it can be accessed, understood and used to the greatest extent possible by all people regardless of their age, size, ability or disability” (Centre for Excellence in Universal Design, 2024). Universal design ensures that accessibility of the assessment is considered and that student accommodations are available. Additionally, this enables students to enter the content in multiple ways, respond to the content in multiple ways, and engage with the content in multiple ways. However, current assessment practices have not realized a fully inclusive and equitable experience. For instance, a recent qualitative study examined 13 interim assessments and found that eight provided some information on universal design principles that were followed during test development and content review, and that there was a fair amount of variation in how universal design was implemented, ranging from minimal to comprehensive procedures (Boyer & Landl, 2021). These findings illustrate that while many assessments claim to be based on universal design, actual definitions, implementation, and functional usability vary widely.

We propose centering the assessment on diversity. Instead of starting with the same standardized formats and functions, which have centered on monolingual students without disabilities, and then retrofitting in an attempt to make the test *also* work for SWD and MLL, begin with a more flexible approach as the standard, with the following considerations.

1. Establish the purpose of the assessment and the population. What is the test measuring and for whom? Is the assessment intended to measure a specific skill, achievement, or language proficiency? Is the assessment for diagnosing difficulties or disabilities? For assessments intended for a diverse population, in addition to identifying what the assessment is intended to measure, it is just as vital to clearly state what the assessment is *not* intended to measure. For example, if an assessment is to measure reading fluency and not English language proficiency, this distinction should be stated. A clearly articulated purpose is central to the inclusivity process, as it will guide the test design to include steps to anticipate and reduce barriers to access and construct irrelevant variance.
2. Define standards and evidence. How will the skill be measured? What evidence needs to be observed? We urge test creators to use an evidence-centered design (ECD), which aims to make explicit the claims, or the inferences one intends to make, and the nature of the evidence that supports the claims (Mislevy & Haertel, 2006).
3. Critical to defining the evidence is identifying access to the assessment materials. For instance, text printed in English may not be accessible for MLL and braille users. ESSA (2015) stated that states must use principles of Universal Design for Learning (UDL) when creating assessments. Incorporating the three principles of UDL helps to address the needs of diverse student groups through establishing (1) multiple means of representation; (2) multiple means of action and expression; and (3) multiple means of engagement (Rose & Meyer, 2006).
4. Ensure that test content and language reflect diverse backgrounds and perspectives.

Test content and instructions should be written and reviewed for cultural inclusion and responsiveness. While maintaining alignment with state and national standards, the wording of items and cultural features, such as situations in reading passages and mathematical application items, may need to be adapted according to specific local contexts. Local educators and, when appropriate, students should be included in developing (e.g., reviewing, piloting, and adapting items) assessments for their communities (Solano-Flores & Trumbull, 2003; Solano-Flores, 2014). Unless vocabulary or reading comprehension is the construct of interest, items should be linguistically simplified in meaningful ways to reduce unnecessary language demand and construct irrelevant variance (Abedi et al., 1997). For example, passive verb forms should be replaced with active, and conditional or relative clauses can be recast to reduce sentence length (Abedi & Lord, 2001). These considerations are especially relevant to math and science tests.

(e) Design test accommodations with consideration for usability and heterogeneity. Test creators must provide not only assessment materials that maximize accessibility but also accommodations to remove any remaining barriers to access. Accommodations that are hard to use or do not meet the needs of individual students are likely to be ineffective or even impose unnecessary cognitive demand and cause test performance decline (Solano-Flores, 2022). Test creators must provide users with guidance and training for assigning accommodations to students. Accommodations should not be assigned broadly by disability category or English proficiency level; rather, the specific barriers that prevent the student from fully accessing the test materials should be addressed (Kettler & Dembitzer, 2022; Satsangi et al., 2019). For example, the use of screen magnification may be appropriate for some students who are low vision while the use of screen magnification with audio may be more appropriate for others.

(f) Include SWD and MLL in item- and test-level psychometric analyses. In crucial standard setting processes such as norming, SWD and MLL must be included if they are part of the test taker population. But inclusion alone is not sufficient: in the analysis they must not be treated as monoliths (Ortiz & Wong, 2022); rather, a truly valid comparison group with similar cognitive, linguistic, and cultural backgrounds and experiences must be constructed to reflect the heterogeneous abilities, proficiencies, and needs of SWD and MLL (Solano-Flores, 2014). Building a diverse and inclusive norming sample has many potential challenges and requires time and effort, but it is the basis for valid score comparisons and should remain a collective aspiration for assessment creation. As a first step, we urge test creators to report their sampling procedures and sample characteristics, including limitations, with transparency. This information will provide guidance on what kind of inferences should and should not be made based on the test scores. For example, if the norming sample did not include students who are deaf or hard of hearing, the technical documentation should specify this omission so that schools that serve students who are deaf or hard of hearing may take this into consideration when using the test or test scores. Test creators bear responsibility for producing evidence on test validity, reliability, and fairness for all students in the population. “Since score validation is a prerequisite to score interpretation, this requirement implies the reliability and validity evidence relevant to scores should also be disaggregated” for SWD and MLL (Kettler & Dembitzer, 2022, p. 340).

**Reporting and Data Use**

Test scores and other quantitative measures of student achievement are not neutral or objective (Castillo & Gillborn, 2022) but artifacts of measurement practices designed and implemented by humans with perspectives and biases. Thus, a critical approach is crucial to reporting and using student data (Gillborn et al., 2018). The inclusion of SWD and MLL in reporting is the first key step to transparency and accountability. However, simply reporting a pooled average for the whole group is not enough: we must attend to potential heterogeneity in the outcomes and experiences within these populations (Suzuki et al., 2021). No language or disability is universally experienced; and inequalities based on socially constructed categories such as race, income, and locale always affect individuals’ outcomes and experiences (Annamma et al., 2013). Where permitted by protections for students’ identity and privacy, data should be reported for subgroups by English language proficiency, home language, disability, race/ethnicity, and other social categories that might impact students’ opportunities and outcomes. This will help address the challenge that outcomes of students in smaller subgroups are often masked by larger groups. Importantly, reporting by subgroup will also address the intersectionality of students’ identities by elevating the experiences and needs of multiply-marginalized populations, such as students of color with disabilities and MLL with disabilities, to the forefront of policy discussions and decisions. In addition to accountability reporting, state and district leaders make key decisions that impact test score reports. These reports are often designed not according to the test creators’ theoretical orientations but client requests, as defined in the state or local education agencies’ requests for proposals. We therefore implore state and district leaders to elevate equity and inclusion in their requests for proposals and ensure that test score reports provide visibility for the achievement and growth demonstrated by MLL and SWD.

When examining data, inferences about students’ skills and proficiencies should only be made based on assessments validated for specific purposes, and with consideration for potential sources of construct-irrelevant variance. Conclusions about students’ English language proficiency should be based only on English language proficiency assessments (e.g., ACCESS, ELPAC), not on standardized tests of English reading and math skills. Scores on tests given without appropriate accommodations should not be the basis for high-stakes decisions that impact students’ access to programs and services. Similarly, comparisons of scores across student groups require attention to centering, reclassification, and interaction.

Research on SWD and MLL should center questions, data, and interpretations on the focal populations: SWD and/or MLL. Breaking from research traditions, treat SWD or MLL as the norm or reference group rather than as the outlier or “other.” Ask questions that are specific to disabilities, language, and other social and cultural experiences that are relevant for SWD and MLL. Collect, analyze, and interpret data in ways that affirm and value their experiences independently, without juxtaposing against a White, monolingual, non-disabled standard or majority (Covarirubias e al., 2018; Mayhew & Simonoff, 2015; Viano & Baker, 2020). Reclassification, or exiting language service, must be considered in light of the “revolving-door” phenomenon (Kieffer and Thompson, 2018; Saunders & Marcelletti, 2013). Students exit language service as they reach higher levels of English proficiency (and sometimes English reading in addition). Solely comparing students who receive services with those who are native or fluent English users inflates the achievement gap and diminishes the progress made by MLL as a group. Thus, outcomes for reclassified students should be reported separately.

Finally, decision-makers should exercise caution in drawing conclusions about students’ academic skills and potential based on language and/or disability, and vice versa. Providing the right instruction and access to education and opportunities is vital. Placements for special education and linguistic support should not perpetuate systemic racism but should serve as a mechanism for equity and social justice. To inform high-stakes decisions in program and service eligibility, we recommend considering multiple measures and voices and assigning weights to the multiple criteria in a consistent manner across students and time (Arbuthnot, 2020; Sireci, 2020). Scores should be considered from assessments that included SWD and MLL in all stages of the test design and the norming procedure. At a minimum, reports should state whether such inclusion applied to the assessment used in decision-making. To the maximum extent possible, measures of achievement in the student’s primary language or mode of communication should be used to supplement tests administered in English. In addition to math and English reading standardized test scores, course grades, behavioral and social-emotional outcomes, language proficiency, measures of other relevant life skills, and family input should be considered. Growth trajectories in the key outcomes over time also merit attention. Research has shown that during summer and other out-of-school periods, students may experience setbacks in learning due to interruptions in instruction and/or services (e.g., Johnson & Barker, 2022). Decisions that impact students’ opportunities should be made with consideration for these patterns.

**Reflection for Researchers**

As Smith (2021) pointed out, “Research in itself is a powerful intervention” (p. 228). Researchers, traditionally members of the dominant group in society, have been on the receiving end of privileged information and benefited (e.g., funding, promotion) from the research enterprise. Throughout the research process, researchers “have the power to distort, to make invisible, to overlook, to exaggerate and to draw conclusions” (p. 229). Our reframing approach rooted in QuantCrit, DisCrit, and critical multiculturalism perspectives suggests that researchers’ positionalities, responsibilities, and training are intricately and inextricably embedded in every stage of research. It is imperative that researchers acknowledge and reflect on these power dynamics and critically examine their own motivations, assumptions, value judgements, methodologies, questions, theories, and communication styles (Smith, 2021).

Through careful reflection, we, the authors, have come to recognize that our research is deeply impacted by our own identities. Two of us were trained in quantitative methods. Our backgrounds and professional responsibilities have contributed to our assumptions about test scores, shaped our preferences for large sample sizes, and even constrained our experience with community-based research and work that focuses on the lived experiences of individuals. We are humbled by such limitations and are aware that engaging with critical approaches is a challenging task.

Researchers must go beyond a general reflection on personal beliefs and assumptions and apply a critical framework to all aspects of their work. We recommend employing Villavicencio et al.’s (2022) culturally responsive research approach, which incorporates multiple partners. Power-sharing between researchers and educators, students, and their communities will democratize the research process and center voices and perspectives that have been silenced in the past. Where relevant, researchers should also dedicate resources to empowering educators and marginalized communities to develop projects that advance social justice. Ultimately, researchers need to commit to “continuously develop their understandings of ethics and community sensibilities, and critically examine their research practices” (Smith, 2021, p. 261).

**Concluding Remarks**

We champion Kettler and Dembitzer’s (2022) model, which “hold[s] the educational system responsible for teaching, and hold[s] tests responsible within that system for allowing students to show their learning, by yielding valid score inferences for those students.” (p. 350).

We implore researchers and test developers to further their understanding of structural barriers to equity and critically examine their own experiences, perspectives, and practices in relation to the students who are the test takers and the researched. Assessment creators should adopt principled, proactive approaches to embed research-based best practices throughout the development process, from infusing ECD and UDL into the design from the very beginning to writing culturally relevant content in the middle and to seeking evidence for reliability, validity, and fairness disaggregated by group membership for the completed assessment. “Because it is logical to establish the quality of a score before considering its quantity” (Kettler & Dembitzer, 2022, p.339), assessment creators must give careful consideration to ensure that the diversity in the student population is addressed in all stages of assessment development.

**References**

Abedi, J. (2004). The No Child Left Behind Act and English Language Learners: Assessment and Accountability Issues. Educational Researcher, 33(1), 4–14. <https://doi.org/10.3102/0013189X033001004>

Abedi, J., Courtney, M., and Leon, S. (2003). Research-supported accommodation for English language learners in NAEP. University of California, Center for the Study of Evaluation, National Center for Research on Evaluation, Standards, and Student Testing.

Abedi, J., Leon, S., & Mirocha, J. (2003). *Impact of students’ language background on content-based assessment: Analyses of extant data* (CSETech. Rep. No. 603). University of California, NationalCenter for Research on Evaluation, Standards, and Student Testing.

Abedi, J., & Lord, C. (2001). The language factor in mathematics tests. *Applied Measurement in Education*, *14*, 219–234.

Abedi, J., Lord, C., Hofstetter, C., & Baker, E. (2000). Impact of accommodation strategies on English language learners’ test performance. *Educational Measurement: Issues and Practice*, *19*(3), 16–26.

Abedi, J., Lord, C., & Plummer, J. (1997). *Language background as a variable in NAEP mathematics performance* (CSE Tech. Rep. No.429). University of California, National Center for Researchon Evaluation, Standards, and Student Testing.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for Educational and Psychological Testing.* American Educational Research Association.

Annamma, S. A. & Handy, T. (2019). DisCrit solidarity as curriculum studies and transformative praxis. Curriculum Inquiry, 49(4), 442-463. <https://doi.org/10.1080/03626784.2019.1665456>

Annamma, S. A., Connor, D., & Ferri, B. (2013). Dis/ability critical race studies (DisCrit): theorizing at the intersection of race and disability. *Race Ethnicity and Education, 16*(1), 1-31.

Agran, M., Lewis, J., Kurth, J.A., Ryndak, D., Burnette, K., Jameson, M., Zagona, A., Fitzpatrick., & Wehmeyer, M. (2020). Why aren’t students with severe disabilities being placed in general education classrooms: examining the relations among classroom placement, learner outcomes, and other factors. *Research and Practice for Persons with Severe Disabilities, 45(1)*, 4-13.

Arbuthnot, K. (2020). Reimagining assessments in the postpandemic era: Creating a blueprint for the future. Educational Measurement: Issues and Practice, 39(3), 97-99.

Boyer, M. & Landl, E. (2021). Interim assessment practices for students with disabilities. National Center on Educational outcomes (NCEO) Brief Number 22. https://eric.ed.gov/?id=ED613028

Brayboy, B. M. J., Castagno, A. E., Maughan, E. (2007). Equality and justice for all? Examining race in education scholarship. Review of Research in Education, 31(1), 159–194.

Briggs, D.C. (2022). *Historical and Conceptual Foundations of Measurement in the Human Sciences: Credos and Controversies*. Routledge.

Campanelle, J., Cerilli, C., Varadaraj, V., Sweeney, F., Smith, J., Zhu, J., Yenokyan, G., Swenor, B.K. (2022). Accessibility and disability inclusion among top-funded U.S. undergraduate institutions. *PLoS ONE, 17(11),1-13*: e0277249. https://doi.org/10.1371/journal. pone.0277249

Castillo, W. & Gillborn, D. (2022). How to “QuantCrit:” Practices and questions for education data researchers and users. (EdWorkingPaper: 22-546). Annenberg Institute at Brown University: <https://doi.org/10.26300/v5kh-dd65>

Carnoy, M. & Garcia, E. (2017). Five key trends in U.S. student performance. Economic Policy Institute Report. <http://www.epi.org/publication/five-key-trends-in-u-s-student-performance-progress-by-blacks-and-hispanics-the-takeoff-of-asians-the-stall-of-non-english-speakers-the-persistence-of-socioeconomic-gaps-and-the-damaging-effect/>

Centre for Excellence in Universal Design, (2024, January 18). Definition and overview of universal design (UD). https://universaldesign.ie/about-universal-design/definition-and-overview

Cioè-Peña, M. (2017). Bilingualism, Disability and What it Means to Be Normal. *Journal of Bilingual Education Research & Instruction*, *19*(1), 2017.

Cioè-Peña, M. (2020). Planning Inclusion: The Need to Formalize Parental Participation in Individual Education Plans (and Meetings). *Educational Forum*, *84*(4), 377–390. https://doi.org/10.1080/00131725.2020.1812970

Coghlan, D., & Brydon-Miller, M. (Eds.) (2014). The SAGE encyclopedia of action research. (Vols. 1-2). SAGE Publications Ltd, https://dx.doi.org/10.4135/9781446294406

Covarrubias, A., Nava, P.E., Lara, A., Burciaga, R., Vélez, V.N. & Solorzano, D.G. (2018). Critical race quantitavtive intersections: a testimonio analysis. *Race Ethnicity and Education, 21(2)*, 253-273, DOI: 10.1080/13613324.2017.1377412

Crenshaw, K. (1991). Mapping the margins: Intersectionality, identity politics, and violence against women of color. Stanford Law Review, 43(6), 1241–1299.

Cruz, R. A., & Rodl, J. E. (2018). An Integrative Synthesis of Literature on Disproportionality in Special Education. *The Journal of Special Education*, *52*(1), 50-63. <https://doi.org/10.1177/0022466918758707>

deBettencourt, L.U. (2002). Understanding the differences between IDEA and section 504. *Teaching Exceptional Children, 34(3)*, 16-23.

DeMatthews, D. (2020) Addressing Racism and Ableism in Schools: A DisCrit Leadership Framework for Principals, The Clearing House: A Journal of Educational Strategies, Issues and Ideas, 93:1, 27-34, DOI: 10.1080/00098655.2019.1690419

Dever, B.V., Raines, T. C., Dowdy, E., & Hostutler, C. (2016). Addressing disproportionality in special education using universal screening approach. *Journal of Negro Education, 85(1)*, 58-71.

Estrada, P., & Wang, H. (2018). Making English Learner Reclassification to Fluent English Proficient Attainable or Elusive: When Meeting Criteria Is and Is Not Enough. American Educational Research Journal, 55(2), 207-242. https://doi.org/10.3102/0002831217733543

Every Student Succeeds Act, 114 U.S.C. § 95 (2015).

Gallagher, C.J. (2003). Reconciling a tradition of testing with a new learning paradigm. *Educational Psychology Review, 15(1)*, 83-99.

Geisinger, K. F. (2007). Preface. In C. Cahalan Laitusis & L.L. Cook (Eds.), Large-Scale Accommodations: What Works? (pp. ix-xi). Council for Exceptional Children.

Georgiou, G.K., Martinez, D., Vieira, A.P.A., & Guo, K. (2021). Is orthographic knowledge a strength or a weakness in individuals with dyslexia? Evidence from a meta-analysis. *Annals of Dyslexia, 71*, 5-27. https://doi.org/10.1007/s11881-021-00220-6

Gillborn, D., Warmington, P. & Demack, S. (2018). QuantCrit: education, policy, ‘Big Data’ and principles for a critical race theory of statistics, *Race Ethnicity and Education*, *21*(2), 158-179, DOI: 10.1080/13613324.2017.1377417

Gilmour, A. F., Fuchs, D., & Wehby, J. H. (2019). Are students with disabilities accessing the curriculum? A meta-analysis of the reading achievement gap between students with and without disabilities. *Exceptional Children*, *85*(3), 329-346.

Gough, P.B. & Tunmer, W.E. (1986). Decoding, reading, and reading disability. Remedial and special education, 7(1), 6-10.

Harper, S. R. (2012). Race without racism: How higher education researchers minimize racist institutional norms. Review of Higher Education, 36(1), 9–29.

Heiser, C. A., Prince. K., & Levy, J. D. (2017). Examining critical theory as a framework to advance equity through student affairs assessment. The Journal of Student Affairs Inquiry, 3(1), 1-17.

Hutchinson, J., Timimi, S., McKay, N. (2021). Trends in SEN identification: contexts, causes and consequences. Journal of Research in Special Education Needs, 21(1), 19-38.

Johnson, A. (2019). A matter of time: Variations in high school course-taking by years-as-EL subgroup. *Educational Evaluation and Policy Analysis*, *41*(4), 461–482. <https://doi.org/10.3102/0162373719867087>

Johnson A., & Barker E., (2022). Understanding Differential Growth During School Years and Summers for Students in Special Education. Journal of Education for Students Placed at Risk (JESPAR), DOI: [10.1080/10824669.2022.2051145](https://doi.org/10.1080/10824669.2022.2051145)

Jonson, J.L. & Geisinger, K.F. (2022). *Fairness in Educational and Psychological Testing: Examining Theoretical, Research, Practice, and Policy Implications of the 2014 Standards.* Washington, DC: American Educational Research Association.

Kanaya, T., Wai, J., & Worrell, F. C. (2022). The “Flynn Effect” and Decision Making in Education: Addressing Fairness Concerns. In Jonson, J. L. & Geisinger, K. F. (Eds) *Fairness in Educational and Psychological Testing: Examining Theoretical, Research, Practice, and Policy Implications of the 2014 Standards* (pp.251-268*).* American Educational Research Association.

Kangas. (2021). “Is it Language or Disability?”: An Ableist and Monolingual Filter for English Learners with Disabilities. *TESOL Quarterly*, *55*(3), 673–683. <https://doi.org/10.1002/tesq.3029>Kanno., Y. & Cromley, J. (2013). English language learners’ access to and attainment in postsecondary education. *TESOL Quarterly*, *47*(1), 89-121.

Kauffman, J. M., Anastasiou, D. & Maag, J. W. (2018). Special Education at the crossroad: An identity crisis and the need for scientific reconstruction. *Exceptionality*, *25*(2), 139–155. doi:10.1080/09362835.2016.1238380.

Kettler, R. & Dembitzer, L. (2022). Fairness and Test Adaptations for Students With Functional Impairments. In Jonson, J. L. & Geisinger, K. F. (Eds) *Fairness in Educational and Psychological Testing: Examining Theoretical, Research, Practice, and Policy Implications of the 2014 Standards* (pp. 333-354). American Educational Research Association.

Kibbee, D. A. (2016). *Language and the law: Linguistic inequality in America*. Cambridge University Press.

Kibler, A. K., Karam, F. J., Ehrlich, V. A. F., Bergey, R., Wang, C., Elreda, L. M. (2018). Who Are ‘Long-term English Learners’? Using Classroom Interactions to Deconstruct a Manufactured Learner Label, Applied Linguistics, 39(5), 741–765, <https://doi.org/10.1093/applin/amw039>

Kieffer, M. J., & Thompson, K. D. (2018). Hidden Progress of Multilingual Students on NAEP. Educational Researcher, 47(6), 391-398. <https://doi.org/10.3102/0013189X18777740>

King, S. (2022). The education context for twice-exceptional students: An overview of issues in special and gifted education. *Neurobiology of Learning and Memory, 193*, 1-7. <https://doi.org/10.1016/j.nlm.2022.107659>

Kim, C.Y., Losen, D.J., & Hewitt, D.T. (2010). *The school-to-prison pipeline: structuring legal reform*. NYU Press.

Košak-Babuder, M., Kormos, J., Ratajczak, M., & Pižorn, K. (2019). The effect of read-aloud assistance on the text comprehension of dyslexic and non-dyslexic English language learners. *Language Testing*, 36(1), 51-75.

Lazarus, S. S., Quanbeck, M., Hinkle, A. R., & Albus, D. (2021). Text-to-speech (computer generated voice): States’ accessibility policies, 2020 (NCEO Accommodations Toolkit #4b). National Center on Educational Outcomes.

Lazarus, S. S., & Thurlow, M. L. (2016). *2015-16 high school assessment accommodations policies: An analysis of ACT, SAT, PARCC, and Smarter Balanced* (NCEO Report 403). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

Lewis, A. E. & Diamond, J. B. (2016) Despite the best intentions: How racial inequality thrives in good schools. 1st edition. Oxford University Press.

May, S. (Ed.) (1999). Critical multiculturalism: Rethinking multicultural and antiracist education. Falmer Press.

May, S. & Sleeter, C. E. (2010). Critical multiculturalism. 1st Edition. Routledge.

Mayhew, M.T. & Simonoff, J.S.(2015). Effect coding as a mechanism for improving the accuracy of measuring students who self-identify with more than one race. *Research in Higher Education, 56(6)*, 595-600.

McGrew, K.S., Thurlow, M.L., & Spiegel, A.N. (1993). An investigation of the exclusion of students with disabilities in national data collection programs. Educational Evaluation and Policy Analysis, 15(3), 339-352. <https://www.jstor.org/stable/1164346#metadata_info_tab_contents>

Messick, S. (1995). Validity of psychological assessment validation of inferences from persons’ responses and performances as scientific inquiry into score meaning. *American Psychologist, 50(9)*, 741 -749.

Mislevy, R.J. Haertel, G. D. (2006). Implications of evidence-centered design for educational testing. *Educational Measurement: Issues and Practice, 25(4)*, 6-20.

Mislevy, R. J., Haertel, G., Cheng, B. H., Ructtinger, L., DeBarger, A., Murray, E., Rose. D…, Vendlinski, T. (2013). A “conditional” sense of fairness in assessment*, Educational Research and Evaluation*, *19*(2-3), 121-140, DOI: 10.1080/13803611.2013.767614

Morgan, P. L., G. Farkas, M. M. Hillemeier, and S. Maczuga. (2017). “Replicated Evidence of Racial and Ethnic Disparities in Disability Identification in U.S. Schools.” Educational Researcher 46 (6): 305–322. doi:10.3102/0013189X17726282.

National Council on Disability (NCD). (2018). Every Student Succeeds Act and students with disabilities. <https://ncd.gov/sites/default/files/NCD_ESSA-SWD_Accessible.pdf>

Norton, E.S., & Wolf, M. (2012). Rapid automatized naming (RAN) and reading fluency: implications for understanding treatment of reading disabilities. *Annual Review of Psychology, 63*, 427-452.

Ortiz, A., Fránquiz, M., & Lara, G., (2020) Applying an intersectionality lens to education policy and practice for emergent bilinguals, Bilingual Research Journal, 43:4, 357-361, DOI: [10.1080/15235882.2020.1896310](https://doi.org/10.1080/15235882.2020.1896310)

Ortiz, S. O. & Wong, J. Y. T. (2022). Theoretical, empirical, and practical issues in testing English learners. In Jonson, J. L. & Geisinger, K. F. (Eds) *Fairness in Educational and Psychological Testing: Examining Theoretical, Research, Practice, and Policy Implications of the 2014 Standards* (pp. 293-332). American Educational Research Association.

Ozernov-Palchik, O., Sury, D., Turesky, T., Yu, X., Gaab, N. (2022) Longitudinal changes in brain activation underlying reading fluency. *Human Brain Mapping*. doi: 10.1002/hbm.26048. Epub ahead of print. PMID: 35984111.

Park, S. (2014). Bilingualism and children with autism spectrum disorders: Issues, research, and implications. *NYS TESOL Journal, 1*(2), 122-129.

Patel, L. (2015). Decolonizing educational research: From ownership to answerability. Routledge.

Peters, S. J., Gentry, M., Whiting, G. W., & McBee, M. T. (2019). Who Gets Served in Gifted Education? Demographic Representation and a Call for Action. *Gifted Child Quarterly*, *63*(4), 273–287. <https://doi.org/10.1177/0016986219833738>

Raj, C. (2016). The misidentification of children with disabilities: a harm with no foul, 48 Ariz. L.J. 373.

Rose, D. H., & Meyer, A. (Eds.). (2006). A practical reader in universal design for learning. Harvard Education Publishing Group.

Russell, M. (2023). *Systemic Racism and Educational Measurement Confronting Injustice in Testing, Assessment, and Beyond*. Routledge.

Rutherford, R.B., Bullies, M., Anderson, C.W., & Griller-Clark, H.M. (2002). Youth with disabilities in the correctional system: prevalence rates and identification issues. Office of Juvenile Justice and Delinquency Prevention.

Satsangi, R., Miller, B., & Savage, M. N. (2019). Helping teachers make informed decisions when selecting assistive technology for secondary students with disabilities. Preventing School Failure: Alternative Education for Children and Youth, 63(2), 97–104.

Saunders, W. M., & Marcelletti, D. J. (2013). The gap that can’t go away: The catch-22 of reclassification in monitoring the progress of English learners. *Educational Evaluation and Policy Analysis*, *35*(2), 139-156.

Shaywitz, S. (2003). Overcoming dyslexia: A new and complete science-based program for reading problems at any level. Alfred A. Knopf.

Schultz, E.K. & Stephens, T.L. (2023). SLD Identification: An analysis of state policies. The Advocacy Institute. Accessed November 6, 2023. <https://www.advocacyinstitute.org/resources/SLD.Identification.State.Policy.shtml>

Sireci, S. (2020). Standardization and understandardization in educational assessment. *Educational Measurement: Issues and Practice,* 39(3), 101-105.  <https://doi.org/10.1111/emip.12377>

Smith, L. T. (2021). Decolonizing methodologies: Research and indigenous peoples. Third Edition. Bloomsbury Publishing.

Solano-Flores, G. (2019). Examining cultural responsiveness in large-scale assessment: The matrix of evidence for validity argumentation. Frontiers in Education, 4:43

Solano-Flores, G. (2014). Probabilistic approaches to examining linguistic features of test items and their effect on the performance of English Language Learners. *Applied Measurement in Education*, *27*(4), 236-247. DOI: 10.1080/08957347.2014.944308

Solano-Flores, G., & Trumbull, E. (2003). Examining language in context: The need for new research and practice paradigms in the testing of English-language learners. *Educational Researcher*, *32*, 3–13.

Solórzano, D., Allen, W. R., Carroll, G. (2002). Keeping race in place: Racial microaggressions and campus racial climate at the University of California, Berkeley. *Chicano/Latino Law Review*, 23, 15.

Steele, J. L., Slater, R. O., Zamarro, G., Miller, T., Li, J., Burkhauser, S., & Bacon, M. (2017). Effects of Dual-Language Immersion Programs on Student Achievement: Evidence From Lottery Data. American Educational Research Journal, 54(1\_suppl), 282S-306S.

Suzuki, S., Morris, S. L., & Johnson, S. K. (2021). Using QuantCrit to Advance an Anti-Racist Developmental Science: Applications to Mixture Modeling. Journal of Adolescent Research, 36(5), 535-560. <https://doi.org/10.1177/07435584211028229>

Tanaka, H., Black, J.M., Hulme, C., Stanley, L.M., Kesler, S. R. Whitfield-Gabrieli, S., Reiss, A.L., Gabrieli, J.D.E., & Hoeft, F. (2011). The brain basis of the phonological deficit in dyslexia is independent of IQ. *Psychological Science, 22(11)*, 1442-1451. DOI: 10.1177/0956797611419521

Tatgenhorst, A., Norlin, J.W., & Gorn, S. (2014). *The answer book on special education law 6th edition*. LRP Publications.

Terzian, S.G. (2021). “Subtle, vicious effects”: Lillian Steele proctor’s pioneering investigation of gifted African American children in Washington, DC. *History of Education Quarterly, 61*, 351-371.

Thompson, K. D. (2015). Questioning the Long-Term English Learner Label: How Categorization Can Blind Us to Students’ Abilities. Teachers College Record, 117(12), 1-50. https://doi.org/10.1177/016146811511701203

Thorius, K. A. K. (2019). Facilitating en/counters with special education’s cloak of benevolence in professional learning to eliminate racial disproportionality in special education, International Journal of Qualitative Studies in Education, 32:3, 323-340, DOI: [10.1080/09518398.2019.1576945](https://doi.org/10.1080/09518398.2019.1576945)

Thorius K. A. K., Stephenson J. (2012). Racial and ethnic disproportionality in special education. In Noltemeyer A. L., Mcloughlin C. (Eds.), Disproportionality in education and special education: A guide to creating more equitable learning environments (pp. 25–44). Charles C Thomas.

Thurlow, L. (2022). Fairness in testing: accessibility in a new educational context. In Jonson, J.L. & Geisinger, K.F. (Eds.), *Fairness in educational and psychological testing: examining theoretical, research, practice, and policy implications of 2014 standards* (pp. 133-152). American Educational Research Association.

Thurlow, M. L., & Larson, J. (2011). Accommodations for state reading assessments: Policies across the nation. Minneapolis, MN: University of Minnesota, Partnership for Accessible Reading Assessment.

Umansky, I. M., Thompson, K. D., & Díaz, G. (2017). Using an Ever–English Learner Framework to Examine Disproportionality in Special Education. Exceptional Children, 84(1), 76-96. https://doi.org/10.1177/0014402917707470

U.S. Department of Education. (n.d.) *Our nation’s English Learners: What are their characteristics?* <https://www2.ed.gov/datastory/el-characteristics/index.html#four>

Viano, S., & Baker, D. J. (2020). How Administrative Data Collection and Analysis Can Better Reflect Racial and Ethnic Identities. Review of Research in Education, 44(1), 301-331. https://doi.org/10.3102/0091732X20903321

Villavicencio, A., Conlin, D., & Pagan, O. (2022). Research-practice partnerships in pursuit of racial justice in schools: Navigating a hostile sociopolitical climate. Education Policy. doi/10.1177/08959048221130353

Voulgarides C. K., Aylward A. (2022). Enduring equity questions: A sequence analysis of citations in response to racial inequity via the Individuals with Disabilities Education Act. Journal of Disability Policy Studies. <https://doi.org/10.1177/10442073221114102>

Voulgarides, C. K., Fergus, E., & Thorius, K. A. K. (2017). Pursuing Equity: Disproportionality in Special Education and the Reframing of Technical Solutions to Address Systemic Inequities. Review of Research in Education, 41(1), 61-87. https://doi.org/10.3102/0091732X16686947

Wolfe, P.S. & Harriott, W.A. (1998). The reauthorization of individuals with disabilities education act (IDEA): what educators and parents should know. *Focus on Autism and Other Development Disabilities, 12(2)*, 88-93.

Wu, Y.-C., Thurlow, M. L., & Liu. K. K. (2021, October). Understanding the characteristics of English learners with disabilities to meet their needs during state and districtwide assessments (NCEO Brief #24). National Center on Educational Outcomes.

1. Throughout this paper, we use the term ‘multilingual learners’ to refer to all students who use one or more languages in addition to English, regardless of their level of language proficiency. [↑](#footnote-ref-1)
2. Throughout this paper, we use the term ‘students with disabilities’ to refer to all students who qualify for special education or a Section 504 plan, regardless of their disability type. [↑](#footnote-ref-2)
3. MLL who are developing English proficiency are classified as English Learners, or ELs, in federal education policy. We use the term MLL throughout the paper to refer to this student population. [↑](#footnote-ref-3)
4. Some schools have adopted other approaches such as response to intervention (RTI). RTI is a way to implement instruction, monitor the student’s progress overtime, and understand whether the student needs intensive interventions to learn. This model may use multiple sources of data, such as classroom observations and behavioral measures, to supplement standardized test scores. However, to the extent that standardized test scores are used to determine baseline academic proficiency and monitor progress, the validity of the test for SWD remains an issue. [↑](#footnote-ref-4)