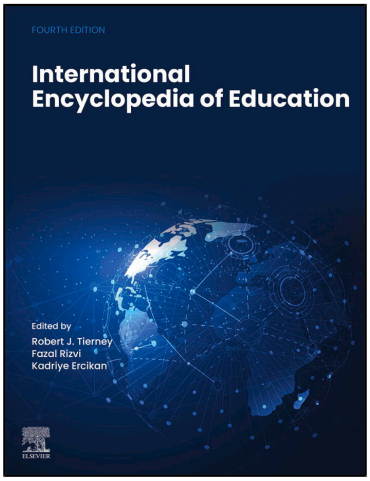




From Oliveri, M.E., Poe, M., Elliot, N., 2023. In: Tierney, R.J., Rizvi, F., Erkican, K. (Eds.), International Encyclopedia of Education, vol. 14. Elsevier.
https://dx.doi.org/10.1016/B978-0-12-818630-5.10006-5.
ISBN: 9780128186305

# Fairness

**Maria Elena Oliveri**[a], **Mya Poe**[b], **and Norbert Elliot**[c], [a] Buros Center for Testing, University of Nebraska, Lincoln, NE, United States; [b] Northeastern University, Harvard, MA, United States; and [c] New Jersey Institute of Technology, Newark, NJ, United States





## Introduction

In 2016, a discussion about fairness might have begun with a general acknowledgment of the varied ways that fairness is understood in social, legal, and measurement contexts and then focus on issues specific to test design: "The goal of the fairness argument is to present evidence that the test is fair for various groups within the test-taking population" (Zieky, 2016, p. 96). Today, in the wake of the COVID-19 global pandemic beginning in 2019 and international Black Lives Matter demonstrations in 2020, conversations about fairness must take these global issues into account. Inattention to social context and justice is not an option anymore. Consequently, in this chapter, we acknowledge that fairness cannot be achieved without jointly considering inequity, anti-racism, and justice. The conventional definition of fairness offered in test design guides and professional standards seem woefully insufficient for a post-pandemic, anti-racist worldview. Fairness cannot be circumscribed solely to discussions of test design but should also include discussions of the processes used to make evaluative judgments of the meaning of test scores, the stakeholders involved in making them, the cultural and linguistic backgrounds of test takers, and the many ways test-based decisions have disadvantaged and continue to disadvantage subgroups. To this end, like Gipps and Stobart (2010), we take an inclusive view of educational assessment, including classroom and program assessment as well as high-stakes testing. We also see fairness as an ongoing project of justice (i.e., using assessment toward addressing injustice; Poe et al., 2018; Randall, 2021).

This entry has four sections. In the first section, we begin our discussion of fairness by presenting examples of (un)fairness in assessment systems used around the world and the types of challenges stakeholders and the public face in relation to addressing fairness considerations within those systems. Those examples inform the key terms we define in the second section. In our definitions, we include both terms used traditionally in measurement and newer terms. In the third section, we describe considerations on the future of fairness in measurement. The fourth section is the conclusion in which we provide a brief reflection on issues discussed in the article.

## Setting the context: international examples of (un)fairness

Real-world examples, drawn from international contexts, such as Chile, the Middle East and North Africa, and the United States, demonstrate the difficulties of achieving fairness when resources are limited and tests are seen as barriers to accessing prestigious universities:

- On January 6 and 7, 2020, the Prueba de Selección Universitaria (PSU)—a battery of multiple-choice achievement tests taken by high school students applying for university admissions—was scheduled to be taken by approximately 300,000 applicants across Chile. Led by student organizations, the Coordinating Assembly of High-School Students, and the National Coordinator of High-School Students, protests erupted against the PSU. Protestors argued that the exams privilege wealthier students and increase economic inequality. Access to test centers were blocked and examination papers were burned. More than 27 people died, with thousands injured and arrested (Nugent, 2020). In October of 2021, large-scale demonstrations marked the anniversary of the protests (Cambero, 2021).
- On June 19, 2021, the Ministry of Education in Sudan ordered the nation's internet service providers to shut down all mobile internet connections in advance of national secondary school examinations. In Algeria, the practice of shutting down the nation's internet has occurred each year since 2016 during periods of national examinations. In Syria, the government stopped internet service across the country between May 31 and June 22 for approximately four and a half hours each day while students took their high-school examinations. And in Jordan, the government blocked the internet between June 24 and July 15, 2021, as students took their high-school examinations. Journalists noted that these practices violated United Nations Resolution 44/12 regarding internet shutdowns (Fatafta et al., 2021).
- On January 20, 2020, the Centers for Disease control confirmed the first case of COVID-19 in the US. By June 2022, more than one million people in the US had died of COVID 19 (Centers for Disease Control and Prevention, 2022a,b). The disease has had



International Encyclopedia of Education, Fourth Edition, 2023, 45–50

a disproportionate effect on historically marginalized groups, resulting in huge disparities in hospitalizations and deaths (Centers for Disease Control and Prevention, 2022a,b; Lopez et al., 2021). The pandemic upended classroom teaching, forcing students from preschool to college age into online learning contexts. In US K-12 schooling, test scores in math and reading declined (Kuhfeld et al., 2022). In US higher education, the pandemic brought about many issues, including declines in student enrollment, declines in students' post-college labor outcomes, especially among students of color (Department of Education, 2021, p. 39, 41), and changes to admissions testing. By March of 2020, approximately 1000 of the 2300 private nonprofit and public bachelor's-granting colleges and universities offered students the option to apply without submitting standardized test scores. By October of 2020, nearly two-thirds of colleges and universities had adopted test-optional, or test score masked, policies in place (Conley and Masa, 2020). A February 2021 ACT survey of 4-year colleges found that 80% ended up being test-optional for the previous year and between 60 and 70% said they would likely remain test-optional or test blind post-COVID (ACT, 2021).

In terms of fairness, the PSU in Chile has been criticized for equity deficits in terms of economic inequality and opportunity to learn; in 2018, approximately 80,000 students did not have the opportunity to learn the material covered in the technical and professional section of the test—the educational track taken by the poorest in the nation. Politically, the PSU has come to symbolize a country divided between those desiring social change and those devoted to conservative policies; an October PSU anniversary demonstration, held in advance of November 2021 presidential and legislative elections, reveals the polarized nature of the country. Internet shutdowns in Middle East and North Africa have been framed as violating the United National General Assembly resolution "*strongly condemning* the use of Internet shutdowns to intentionally and arbitrarily prevent or disrupt access to or dissemination of information online" (United National General Assembly, 2020, p. 3)—an example of how educational testing and human rights are conflated when technology becomes a feature of the testing environment. In the US, the decline of standardized college admissions testing during the global COVID-19 pandemic strengthened the long-standing concerns that standardized admissions tests may not provide information on students that is sufficient to compensate for inequitable barriers to access created by the tests themselves. One early study of 100 private institutions has shown that test-optional admissions increase Pell Grant applicants, women, and first-time students from underrepresented racial/ethnic backgrounds (Bennett, 2021).

The examples above illustrate the complexities of fairness when we consider the vary unfair conditions of assessment globally. If we are to determine complexities related to fairness, including the ways assessment is interwoven with larger social issues such as public health crises, further understanding can also be gained by identifying key terms related to fairness. To this end, in the next section, we provide definitions useful in conceptualizing fairness.

## Definitions: key terms related to fairness

Inherently, fundamental considerations of fairness are a part of education. When we narrow fairness applications to educational measurement alone, discussions of unfairness (fairness violations) arise across all settings, from elementary to secondary school group scoring programs to higher education admissions testing—as well as across purposes of testing, from classroom assessments to testing for licensure and certification. In this broad range of testing applications, it seems nearly impossible to begin a discussion of fairness without considering the larger context in which tests are used. This issue is ever more important when we use situated language frameworks in which local environments have a substantial influence on stakeholders impacted by tests and its scores: advisory board members, administrators, faculty, parents, professional organizations, students, tribes and tribal members, and the public (Gee, 2020; Mislevy, 2018). Within this context, we next describe key terms related to fairness. Our list of key terms includes both terms from the educational assessment community used traditionally in fairness related discussions and newer terms that reflect global calls for fairness and justice. These terms are organized in alphabetical order.

- *Anti-racist assessment.* These assessment practices disrupt racist beliefs around learning by actively confronting the economic, structural, and historical roots of inequality (Kendi, 2016). With anti-racist approaches to assessment, teachers and test designers "acknowledge both the current and historical role of race and racism in their own pedagogical and assessment practices" (Randall et al., 2021, p. 596).
- *Bias*: Sources of construct underrepresentation or construct-irrelevance variance may differentially disadvantage test-taker groups. In statistics, bias is identified through systematic error in a test score (American Educational Research Association et al., 2014).
- *Consequence*: Impact identifies the intended or unintended outcomes of using tests in particular ways, in certain contexts, with certain populations. Some consequences are directly related to the score interpretation and use, while others are related to testing contexts or unintended negative consequences (American Educational Research Association et al., 2014).
- *Culturally responsive assessment:* As explained by Montenegro and Jankowski (2017), culturally responsive assessment is "mindful of the student populations the institution serves, using language that is appropriate for all students when developing learning outcomes, acknowledging students' differences in the planning phases of an assessment effort, developing and/or using assessment tools that are appropriate for different students, and being intentional in using assessment results to improve learning for all students" (p. 10) (See also Hood et al., 2015).
- *Culturally responsive indigenous evaluation*: Also known as an Indigenous Self-Determination Evaluation Model, CRIE "respects, recognizes, and values the inherent worth of Indian culture; is responsive to the communities' needs as voiced by all members of

the tribal community; builds evaluation designs and processes around Indian assets and resources; and literally and figuratively employs Indians in every part of the process (program, policy, implementation, evaluation) to heal, strengthen, and preserve Indigenous societies for the next 7 generations" (Bowman, 2005, p. 8).

- *Culturally sustaining assessment*: As explained by Randall et al. (2021), "culturally sustaining approaches to assessment (a) draw on Black, Indigenous, People of Color (BIPOC) students' funds of knowledge (Vélez-Ibáñez and Greenberg, 1992), (b) are connected to the lives of these students, (c) allow them to demonstrate their competence in a variety of ways through community cultural wealth (Yosso, 2005), and (d) are embedded within a culturally sustaining curriculum. Such an approach would 'perpetuate and foster … linguistic, literate, and cultural pluralism as part of the democratic project of schooling and as a needed response to demographic and social change' (Paris and Alim, 2014, p. 88)" (p. 596) (See also Ladson-Billings, 1995).
- *Disparate impact*: Differences in educational outcomes may result from facially neutral policies or practices. When used as both a statistical and conceptual tool, disparate impact analysis allows detailed analysis educational consequences, as well as legal strategies that may be used to redress inequity (Poe and Cogan, 2016)
- *Equity*: Gee (2008) calls for assurances that every individual has the opportunity to develop an identity as a competent learner, participate in communities of learning, and sustain a learning trajectory. Equity-centered design acknowledges that students have rich backgrounds of experience that, when acknowledged, provide opportunities for knowledge demonstration (Oliveri et al., 2020; Pullin, 2008)
- *Fairness*: "Validity of test score interpretations for intended use(s) for individuals from all relevant subgroups. A test that is fair minimizes the construct-irrelevant variance associated with individual characteristics and testing contexts that otherwise would compromise the validity of scores for some individuals" (American Educational Research Association et al., 2014, p. 219).
- *Opportunity to Learn* (*OTL*): Exposure to particular knowledge, skills, and attitudes needed for curricular success. Consideration of OTL is important for the study of alignment between tests and curriculum, as well as allocation of resources needed to assure learning for all students (Moss et al., 2008).
- *Validity*: Extent to which evidence and theory support a specific interpretation and use argument. As is the case with fairness and reliability—the foundations of educational measurement—validity is considered as a process of evidence gathering (American Educational Research Association et al., 2014).
- *Reliability/Precision*: The degree to which test scores are consistent over repeated test applications. A test is said to provide evidence of reliability/precision when a measurement process is dependable, consistent, and free from random error (American Educational Research Association et al., 2014).
- *Social Justice*: Theories holding that fairness is understood through analysis of social contracts, distributive principles, and communal interaction. While theories of social justice vary, each maintains a commitment to principled strategies aimed at understanding and reducing inequality (Rawls, 1971/1999; Young, 2011).
- *Universal Design*: Approaches to educational measurement that maximize test accessibility for all test takers. Practices of universal design ensure that a test is fair for all test-takers, regardless of gender, age, language background, socioeconomic status, or disability (American Educational Research Association et al., 2014).

In general, some concepts—such as bias and reliability/precision—fall into technical work associated with educational measurement. Other concepts—such as culturally responsive assessment—are often associated with pedagogical practice. And yet other concepts, such as disparate impact, are associated with legal precedent. When studying fairness, it is important to emphasize commonalities among communities, from obligations to stakeholders to consideration of multidisciplinary collaboration. However, it is equally important to recognize that these definitions themselves are not without critics. For example, Randall et al. (2021) argue that OTL "has been used to justify construct definitions that exclude/attempt to erase the ways of knowing and understanding of BIPOC students. Simply put, constructs are defined in such a way that privilege whiteness and then BIPOC communities are informed that they simply have not had the opportunity to perfect/improve their whiteness" (p. 85). When defined with recognized qualifications, the terms defined above are one way of opening portals to common frames of reference (Meyer and Land, 2006).

Various fairness definitions have been offered in the *Standards for Educational and Psychological Testing* as well as educational measurement guidelines (American Educational Research Association et al., 2014; ETS, 2014; ITC, 2018). For a historical review of the *Standards*, see Sireci and Randall (2022). As Zwick (2019) has noted, increasing awareness of standards allow stakeholders to reach informed selection decisions, use multiple measures in making those decisions, and maintain decision transparency.

The best-known articulation of standards is that issued, in its present form, by three leading disciplinary organizations: American Educational Research Association [AERA], American Psychological Association [APA], and National Council on Measurement in Education [NCME]. With origins in 1952, the present version of the *Standards* established fairness as a center of evidence equal to validity and reliability as a foundational concept (American Educational Research Association et al., 2014; American Psychological Association and Committee on Test Standards, 1952). As noted in the definition provided above, fairness is understood as the validity of test score interpretations for intended uses applied to individuals from all relevant subgroups. Further, a test is said to be fair if it minimizes the construct-irrelevant variance associated with individual characteristics and testing contexts that otherwise would compromise the validity of scores for some individuals. In associating fairness with validity, the authors demonstrate the integrative value of fairness as a fundamental category of evidence. Beyond the definition, the *Standards* for fairness focus on minimizing barriers to valid score interpretations, providing detailed comment on accommodations and safeguards, and emphasizing the general conditions needed allow diverse student groups opportunities to demonstrate the construct under examination. As such,

the authors adopt a distributed perspective in which fairness is both the goal for all testing applications and the responsibility of all parties involved in testing.

Of equal interest are the *ITC Guidelines for the Large-Scale Assessment of Linguistically Diverse Populations* (International Test Commission, 2018) and the *ETS International Principles for Fairness Review of Assessments* (Educational Testing Service, 2016). Both are focused on testing within or across countries or regions that are linguistically and culturally diverse. For example, the International Test Commission's (2018) *Guidelines* "are designed to inform test developers, psychometricians, and test users of the considerations that should be made to help ensure test fairness and score comparability to support meaningful inferences in culturally and linguistically diverse contexts" (p. 3). Emphasizing diversity, the *ITC Guidelines* define fairness as evidence that an adapted assessment—changes made in content, format, or test administration to increase access for cultural or linguistic groups—will result in valid demonstration of test-takers construct knowledge. In its definition of fairness, ETS cites the *Standards* (2014). These guidelines and principles are important to understanding the factors that impact fair assessment in diverse populations of test-takers, including legal language status, language of instruction, and resources for adapting tests. Adopting the sociocognitive and sociocultural perspectives identified above are a great advantage in understating diversity.

## Considerations: The future of fairness in measurement

While Zwick (2019) has noted the importance of using existing standards to deepen stakeholder perspectives, she has also advocated the importance of candid and clear communication. As she observes, frank and honest communication allows identification of common ground among stakeholders, including discussions of test fallibility and avoidance of defensiveness. She is cautious, however, in recommending a single solution model to the solving the challenges of fairness in educational measurement. In her discussion of bias, for example, she quotes Cronbach (1976, p. 31): "Make no mistake. The issues will not be settled by mathematical specialists." Standards can't stop student protests, make governments restore the internet, or erase decades of racist, classist, sexist gatekeeping.

To try to imagine a discussion of fairness today without considering these larger social implications is shortsighted, but there are real questions as to what future of fairness is to be in educational assessment. To those ends, we offer four considerations as a way to reimagine the Gipps and Stobbart discussion of relationships among equity, access, curriculum, and assessment in the previous edition of this Encyclopedia (2010, Table 1, p. 58).

- *Consider fairness as a form of purpose pluralism.* Newton (2017) has called attention to the significance of purpose pluralism by noting that "assessment design should be driven by a multiplicity of assessment purposes simultaneously" (p. 5). In advancing purpose pluralism, he describes the tensions between purism (a fixed belief that an assessment should be used for a single purpose) and pluralism (a fluid belief acknowledging the many aims and perspectives involved in coordinating the activities of test designers and users). In proposing purpose pluralism, Newton identifies important perspectives related to information quality (targeting use of assessment results to make decisions, with emphasis on a trustworthy and transparent basis for making important decisions), expertise (emphasizing fidelity to a specific and prespecified cluster of learning outcomes, with attention to providing assessment justification), and engagement (focusing on the consequences arising from anticipating assessment impact for various participants, with variations regarding how positive consequences may be achieved). In applications of purpose pluralism to fairness, it may be useful to think about a central definition that can be applied to a specific assessment, with tailored versions of that definition shaped by various stakeholders. Evidence of fairness can then be gathered in accordance with these definitions, with claims made in cases where the definitions are congruent, and qualifications made in cases where the definitions are incongruent.
- *Consider fairness as an ongoing process.* While statisticians are likely correct in noting that answers related to bias may not be settled by mathematical specialists, it is equally true that much of our understanding of fairness has benefitted enormously by psychometric analysis in which statistical methods are used determine disparate impact. In seminal work, Cleary (1968) used general linear modeling techniques to target evidence of test bias based on whether or not the criterion score was consistently too high or too low for members of the subgroup. This work combined emphasis on both social goals of equity and statistical goals of production. Perhaps, in the future, it will be equally true that questions regarding fairness will be answered by more fully explicating interpretation and use of scores (Kane, 2013) as well as their social consequences. As Camilli (2006) has presently observed, test developers and teachers alike are embedded in social contexts. The future may therefore well be determined by reconciliation of that situatedness—in terms of transparent test construction and clear community communication on the part of test developers, careful test selection and anticipated consequences on the part of teachers, and sustained research and action on the social uptake of test results. Sustained research and action on the social uptake of test results would invite the linking of historical injustice to current action and ongoing work toward achieving justice for all test takers (Hood and Hobson, 2008).
- *Consider fairness as advanced through anticipatory design practices.* The 21st century has witnessed a variety of consequence-centered, anticipatory approaches to designing a variety of assessment genres. These approaches find their origin in work by Mislevy et al. (2003) in Western theory. Using a conceptual assessment framework, evidence-centered design uses model-based reasoning to identify conceptual, delivery, scoring, and reporting models for any given measurement episode. This assessment framework has been followed by related design approaches, including Integrated Design and Appraisal Framework (IDAF, Slomp, 2016),

expanded evidence-centered design (Arieli-Attali et al., 2019), and equity-centered design (Oliveri et al., 2020). Similarly, theory of action techniques have been used to conceptualize the educational effectiveness of an innovation in terms of governing variables (curricular models), action strategies (curricular delivery strategies) and impact (consequences of those strategies) (Argyris, 1997). When applied to educational tests, classroom assessment, and community education evaluation, this approach holds the potential to help identify patterns of interactions among stakeholders and various forms of consequences, both intermediate and long term (Hazelton et al., 2021). In addition, a decolonial approach to anticipatory design is culturally responsive indigenous evaluation in which evaluators do not "impose a model, evaluation design, instruments, or tools upon the members of that community" (Waapalaneexkweew, 2018, p. 546). Instead, through "recognition of the legal implications of Indigenous sovereignty," designers move beyond the "'beads and feathers' linguistic and cultural aspects of evaluation to work with tribal governments and communities as sovereign partners," in the design, adaptation, problem-solving, and reporting of evidence-based evaluations (Waapalaneexkweew, 2018, p. 550). Taken together, evidence-centered design approaches, theory of action techniques, and culturally responsive indigenous evaluation provide powerful ways for assessment specialists and classroom teachers to consider factors related to evidence of fairness, validity, and reliability in a given measurement episode.

- *Consider fairness as understood through multidisciplinary and community collaboration.* While the twentieth century may be said to have proposed solutions to social challenges, the 21st century may be said to have viewed those very solutions as problems. From the present vantage point, addressing these problems will require multidisciplinary collaboration. As Oliveri et al. (2021) have shown in the case of a complex digital assessment, coordination of multidisciplinary action involving a wide variety of community stakeholders and experts (from computer science to policy analysis) is needed to maximize the desired, intended consequences from the use of an assessments—and to minimize undesirable, unintended effects.

## Conclusion

Whether framed in terms of broad societal issues such as resource distribution or specific applications such as admissions testing, fairness is a critical subject to discuss in educational assessment. While the concepts involving evidence of fairness may be complex, they are nevertheless accessible. Even the most complex legal, statistical, and community-based discussions of fairness are possible, especially when these discussions are framed—at the outset—as accessible through multidisciplinary collaboration. Often, fairness is envisioned as an unattainable goal, but present international circumstances described in the introduction to this Encyclopedia entry reveal that position to be untenable. While it may be argued that it was impossible to prevent the adverse consequences identified in the case studies drawn from Chile, the Middle East and North Africa, and the United States, it can be equally argued that intentional use of the perspectives provided here could have lessened their impact. And in the wake of the COVID 19 pandemic and Black Lives Matter movement, discussions of fairness can never again be extracted from the social fabric in which fairness is woven.

## References


ACT, 2021. Summary findings: Survey of higher education enrollment and admissions officers. ACT. https://chronicle.brightspotcdn.com/a1/52/dcf530674ab59217fc5f025d6a76/210212-hedsurveysummaryfindings-externaluse.pdf.

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

American Psychological Association, Committee on Test Standards, 1952. Technical recommendations for psychological tests and diagnostic techniques: preliminary proposal. Am. Psychol. 7, 461–475.

Argyris, C., 1997. Learning and teaching: a theory of action perspective. J. Manag. Educ. 21, 9–26. https://doi.org/10.1177/105256299702100102.

Arieli-Attali, M., Ward, S., Thomas, J., Deonovic, B., von Davier, A.A., 2019. The expanded evidence-centered design (e-ECD) for learning and assessment systems: a framework for incorporating learning goals and processes within assessment design. Front. Psychol. 10, 1–17. https://doi.org/10.3389/fpsyg.2019.00853.

Bennett, C.T., 2021. Untested admissions: examining changes in application behaviors and student demographics under test-optional policies. Am. Educ. Res. J. https://doi.org/10.3102/00028312211003526.

Bowman, N.R., 2005. Government to government evaluation: Issues and strategies for conducting evaluation with tribal governments. Paper Presented at the Annual Conference of the American Evaluation Association, Toronto, Ontario, Canada.

Cambero, F., 2021. Polarized Chile marks anniversary of 2019 protests as election nears. Reuters. https://www.reuters.com/world/americas/chile-braces-protests-crossroads-election-nears-2021-10-18/.

Camilli, G., 2006. Test fairness. In: Brennan, R.L. (Ed.), Educational Measurement, fourth ed. American Council on Education/Praeger, pp. 221–256.

Centers for Disease Control and Prevention, 2022a. Health equity considerations and racial and ethnic minority groups. https://www.cdc.gov/coronavirus/2019-ncov/community/health-equity/race-ethnicity.html.

Centers for Disease Control and Prevention, 2022b. COVID Data Tracker. https://covid.cdc.gov/covid-data-tracker/#datatracker-home.

Cleary, T.A., 1968. Test bias: prediction of Negro and white students in integrated colleges. J. Educ. Meas. 5, 115–124.

Conley, B., Masa, B., 2020. The decline of testing affects more than testing. Inside High. Ed. https://www.insidehighered.com/admissions/views/2020/12/14/decline-standardized-testing-affects-more-testing-opinion.

Cronbach, L.J., 1976. Equity in selection: where psychometrics and political philosophy meet. J. Educ. Meas. 13, 31–42.

Department of Education, 2021. Education in a Pandemic: The Disparate Impacts of COVID-19 on America's Students. Office for Civil Rights. https://www2.ed.gov/about/offices/list/ocr/docs/20210608-impacts-of-covid19.pdf.

Educational Testing Service, 2014. ETS Standards for Quality and Fairness. Educational Testing Service.

Educational Testing Service, 2016. ETS International Principles for Fairness Review of Assessments. Educational Testing Service.

Fatafta, M., Mnejja, K., Anthonio, F., 2021. Internet shutdowns during exams: When MENA governments fail the test. accessnow. https://www.accessnow.org/mena-internet-shutdowns-during-exams/.
Gee, J.P., 2008. A sociocultural perspective on opportunity to learn. In: Moss, P.A., Pullin, D.C., Gee, J.P., Haertel, E.H., Young, L.J. (Eds.), Assessment, Equity, and Opportunity to Learn. Cambridge University Press, pp. 76–108.
Gee, J.P., 2020. What is a Human? Language, Mind, and Culture. Palgrave Macmillan.
Gipps, C., Stobart, G., 2010. Fairness. In: McGraw, B., Baker, E., Peterson, P. (Eds.), International Encyclopedia of Education, third ed. Elsevier, pp. 56–60.
Hazelton, L., Nastal, J., Elliot, N., Burstein, J., McCaffrey, D.F., 2021. Formative automated writing evaluation: a standpoint theory of action. J. Resp. Wri. 7, 37–91. https://scholarsarchive.byu.edu/cgi/viewcontent.cgi?article=1041&context=journalrw.
Hood, S., Hopson, R.K., 2008. Evaluation roots reconsidered: Asa Hilliard, a fallen hero in the "Nobody Knows My Name" Project, and African educational excellence. Rev. of Educ. Res. 78 (3), 410–426. https://doi.org/10.3102/0034654308321211.
Hood, S., Hopson, R.K., Kirkhart, K.E., 2015. Culturally responsive evaluation: Theory, practice, and future implications. In: Newcomer, K.E., Hatry, H.P., Wholey, J.S. (Eds.), Handbook of Practical Program Evaluation, fourth ed. Jossey-Bass, pp. 281–317.
International Test Commission, 2018. ITC Guidelines for the Large-Scale Assessment of Linguistically Diverse Populations. Version 4.2. Buros Center for Testing.
Kane, M.T., 2013. Validating the interpretation and uses of test scores. J. Educ. Meas. 50, 1–73. https://doi.org/10.1111/jedm.12000.
Kendi, I.X., 2016. Stamped from the Beginning: The Definitive History of Racist Ideas in America. Nation Books.
Kuhfeld, M., Soland, J., Lewis, K., 2022. Test score patterns across three COVID-19-impacted school years. In: Annenberg Institute at Brown University EdWorkingPaper 22–521. https://doi.org/10.26300/ga82-6v47.
Ladson-Billings, G., 1995. Toward a theory of culturally relevant pedagogy. Am. Educ. Res. J. 32 (3), 465–491.
Lopez, L., Hart, L.H., Katz, M.H., 2021. Racial and ethnic health disparities related to COVID-19. JAMA 325 (8), 719–720. https://doi.org/10.1001/jama.2020.26443.
Meyer, J.H.F., Land, R. (Eds.), 2006. Overcoming Barriers to Social Understanding: Threshold Concepts and Troublesome Knowledge. Routledge.
Mislevy, R.J., Steinberg, L.S., Almond, R.G., 2003. On the structure of educational assessment. Measurement 1, 3–62. https://doi.org/10.1207/S15366359MEA0101_02.
Mislevy, R.J., 2018. Sociocognitive Foundations of Educational Measurement. Routledge.
Montenegro, E., Jankowski, N.A., 2017. Equity and assessment: moving towards culturally responsive assessment (Occasional Paper No. 29). National Institute for Learning Outcomes Assessment. https://files.eric.ed.gov/fulltext/ED574461.pdf.
Moss, P.A., Pullin, D.C., Gee, J.P., Haertel, E.H., Young, L.J. (Eds.), 2008. Assessment, Equity, and Opportunity to Learn. Cambridge University Press.
Newton, P.E., 2017. There is more to educational measurement than measuring: the importance of embracing purpose pluralism. Educ. Meas. 36, 5–15. https://doi.org/10.1111/emip.12146.
Nugent, C., 2020. Why Chile's SATs have become the new frontline of inequality protests. Time. https://time.com/5770308/chilestudent-protests/.
Oliveri, M.E., Slomp, D., Elliot, N., Rupp, A., Mislevy, R., Vezzu, M., Tackitt, A., Nastal, J., Phelps, J., Osborn, M., 2021. Introduction: meeting the challenges of workplace English communication in the 21st Century. J. Wri. Analy. 5, 1–33. https://wac.colostate.edu/docs/jwa/vol5/intro.pdf.
Oliveri, M.E., Nastal, J., Slomp, D., 2020. Reflections on equity-centered design. In: ETS Research Report 20–22. Educational Testing Service.
Paris, D., Alim, H.S., 2014. What are we seeking to sustain through culturally sustaining pedagogy? A loving critique forward. Harv. Educ. Rev. 84 (1), 85–100.
Poe, M., Cogan, J.A., 2016. Civil rights and writing assessment: using the disparate impact approach as a fairness methodology to evaluate social impact. J. Writ. Assess. 9. http://journalofwritingassessment.org/article.php?article=97.
Poe, M., Inoue, Asao, A.B., Elliot, N., 2018. Writing Assessment, Social Justice, and the Advancement of Opportunity. The WAC Clearinghouse; University Press of Colorado. https://doi.org/10.37514/PER-B.2018.0155.
Pullin, D.C., 2008. Assessment, equity, and opportunity to learn. In: Moss, P.A., Pullin, D.C., Gee, J.P., Haertel, E.H., Young, L.J. (Eds.), Assessment, Equity, and Opportunity to Learn. Cambridge University Press, pp. 333–351.
Randall, J., 2021. 'Color-neutral' is not a thing: redefining construct definition and representation through a justice-oriented critical antiracist lens. Educ. Meas. 40, 82–90. https://doi.org/10.1111/emip.12429.
Randall, J., Poe, M., Slomp, D., 2021. Ain't Oughta Be in the dictionary: getting to justice by dismantling anti-black literacy assessment practices. J. Adolesc. Adult Literacy 64 (5), 594–599.
Rawls, J., 1999. A Theory of Justice. Cambridge University Press (Original work published 1971).
Sireci, S., Randall, J., 2022. Evolving notions of fairness in testing in the United States. In: Clauser, B.E., Bunch, M.B. (Eds.), The History of Educational Measurement. Routledge, pp. 111–135.
Slomp, D., 2016. An integrated design and appraisal framework for ethical writing assessment. J. Writ. Assess. 9 (1). http://journalofwritingassessment.org/article.php?article=91.
United National General Assembly, 2020. Promotion and Protection of All Human Rights, Civil, Political, Economic, Social and Cultural Rights, including the Right to Development. Resolution 44.12. https://undocs.org/en/A/HRC/RES/44/12.
Vélez-Ibáñez, C.G., Greenberg, J.B., 1992. Formation and transformation of funds of knowledge among U.S. Mexican households. Anthropol. Educ. Q. 23 (4), 313–335.
Waapalaneexkweew (Nicole R. Bowman-Farrell), 2018. Looking backward but moving forward: honoring the sacred and asserting the sovereign in indigenous evaluation. Am. J. Eval. 39 (4), 543–568.
Yosso, T.J., 2005. Whose culture has capital? A critical race theory discussion of community cultural wealth. Race Ethn. Educ. 8, 69–91.
Young, I.M., 2011. Responsibility for Justice. Oxford University Press.
Zieky, M., 2016. Developing fair tests. In: Lane, S., Raymond, M.R., Haladyna, T.M. (Eds.), Handbook of Test Development. Routledge, pp. 81–99.
Zwick, R., 2019. Fairness in measurement and selection: statistical, philosophical, and public perspectives. Educ. Meas. 38, 34–41. https://doi.org/10.1111/emip.12299.