F

Fairness in Educational Assessment

Robin D. Tierney
Research-for-Learning, San Jose, CA, USA

Synonyms

Bias; Care; Consistency; Differentiation; Equality; Equity; Justice; Opportunity to learn; Reliability; Respect; Transparency; Trust; Validity

Introduction

Fairness is a moral virtue and a fundamental quality in educational assessment. Understanding of fairness in educational assessment has evolved with developments in learning theory and measurement, and it has increasingly been recognized as a necessary quality for inclusivity in education. Fairness is closely related to, but distinct from, the societal concepts of equality, equity, and justice and the measurement concepts of bias, reliability, and validity. In practice, three conditions contribute to fairer educational assessment: opportunity to learn, a constructive environment, and evaluative thinking. Multiple strategies, revolving around the principle of transparency and the provision of opportunity to demonstrate learning, should be used to ensure fairer educational assessment for diverse learners.

The Concept of Fairness

Some English words are used fleetingly by one generation, snatched up like new toys and quickly abandoned, whereas others are used by generations for centuries. The roots of the word *fair* date back to the Germanic period in Northern Europe (400–800 AD). In Old English, fair meant beautiful, pleasant, or agreeable, and it was broadly applied from physique (e.g., fair hair) to sailing (e.g., fair wind). By Medieval Times, it was also used to describe good treatment of others (e.g., fair dealings), particularly in following the rules of competition (e.g., fair play). This etymological transition illuminates the link between two definitions of *fair* in English language dictionaries, which might otherwise seem disparate. Fair is a physical quality characterized by an absence. A fair day lacks inclement weather. Fair is also a behavioral quality, specifically interacting or treating others without self-interest, partiality, or prejudice. Both of these definitions underlie the meaning of fairness in educational assessment.

Common expressions relating to fairness help explain how educational stakeholders interpret the concept. A *fair shake*, for example, means to give someone or something a reasonable opportunity. The expression comes from games of chance and it holds two ideas stemming from the above etymology. The first is openness; game rules usually require dice to be visible as they roll. The opposite is reflected in expressions, such as under-the-table or shady deals, that

M.A. Peters (ed.), *Encyclopedia of Educational Philosophy and Theory*,
DOI 10.1007/978-981-287-588-4

refer to dishonest or questionable proceedings. The second element is balance; the playing surface and the sides of the dice should be even so that the results are not impeded by irregularities. This idea is expressed as fair and square in business. The association with balance suggests that fairness is not extreme. Fair weather is pleasant, neither very hot nor very cold. To do something fairly well means it is good, neither a poor nor an excellent performance. In contrast, favoring one side, person, or perspective is considered unfair, or biased. Fairness in human interactions involves reasoning, or weighing the balance. It is for this reason that unbiased opinions, independent investigators, and evenhanded judges are sought, particularly when high-stakes decisions are involved.

A complication for discussion across educational contexts is that the word fairness is not directly translatable, with the exception of Scandinavian languages. In other languages, the related terms equality, equity, and justice are necessarily used. *Equality* refers to sameness, or the state of being equal. Treating people equally can be fair in some situations (e.g., cutting a cake), and it is an important step for social justice (e.g., universal human rights), but its application can also have the reverse effect. For example, one-strength-fits-all eyeglasses would clearly be unfair for some. *Equity* is the state of being just, impartial, or fair. Something equitable is proportionately equal, meaning that it is distributed according to need. The three bears in Goldilocks, for instance, had an equitable arrangement in beds and bowls. *Justice* is the quality of being fair, reasonable, or just. The word *just* also has multiple meanings, and it can be used to mean morally right (e.g., a cause), based on sound reason (e.g., a decision), or deserved (e.g., desserts). As their definitions indicate, equity, justice, and fairness are closely related, and they are often used in tandem. However, their use in context shows that they are not completely synonymous. For example, just is more apt as a descriptor for a cause than for access, whereas the opposite is true for equitable. Furthermore, neither equity nor justice encompass the same sense of openness that is associated with fairness.

The universality of the concept of fairness has been questioned because of the missing cognates across languages. Several points suggest that at very minimum, fairness is widely considered desirable as a social quality. All world religions preach the ethic of reciprocity (i.e., the Golden Rule), which essentially encourages fair treatment of others. Fairer social and economic practices are promoted by international initiatives (e.g., fair trade), and similar concerns about fairness in education are voiced across national boundaries (e.g., research on grading). Additionally, research in behavioral sciences has shown that most humans and some animals have a sense of fairness that goes beyond the level of cooperation needed for survival. In counterpoint, there are radical differences in political opinion regarding the notion that *life is unfair*. From a liberal perspective, this idea provides initiative for social reform, whereas in conservative economics it is used as justification for perpetuating the status quo. Although the latter may simply be political rhetoric, there is an underlying equation of fairness with sameness, which is also how some educational stakeholders understand fairness. Thus, while fairness is widely appreciated, different interpretations exist within and across educational contexts.

Imperatives for Fairness in Educational Assessment

Educational assessment encompasses two main types: classroom assessment and external assessment. Both involve the process of collecting and evaluating information about student learning, but they serve different purposes. In *classroom assessment*, teachers and students ideally use various assessment strategies, tools, and tasks for two purposes: to support ongoing teaching and learning (i.e., formative assessment, or assessment for learning) and to report on the achievement of learning expectations (i.e., criterion-referenced summative assessment). Classroom assessment usually relies on teachers' knowledge and professional judgment. *External assessment* includes standardized tests and large-scale assessments developed commercially or by organizations

outside the classroom. Standardized tests are used to determine individual levels of achievement in reference to a norm group, usually for the purpose of certification or selection (e.g., program admission). Large-scale assessments are administered at regional, national, and international levels, either within or across educational systems. They are usually used for accountability, meaning that their purpose is to evaluate system effectiveness. The purpose of an assessment affects what should be considered during its development and use for the fairest results. Guiding principles or standards have been produced by various test developers, educational committees, and professional organizations since the mid-twentieth century. Although these documents do not always clearly define fairness, they reflect a strong interest in fair assessment in the educational community.

Three imperatives underpin the quest for fairness in educational assessment. The first is democratic. Various forms of assessment have been used through history for candidate selection. Government appointments in China were based on written examinations for centuries, from the Han Dynasty (202 B.C. to A.D. 200) to the early twentieth century. Written examinations were also used, from the mid-nineteenth century on, to select candidates for university admissions in western nations. The prestige conferred through a rigorous selection process was initially of more interest to the administering institutions than any democratization of the process. Examinations came to be seen as fairer because they allowed merit to be considered in the selection process. They thus appeared to broaden access to opportunity, which had previously been restricted to a privileged minority. With the development of psychological and educational measurement in the twentieth century came a dramatic rise in the prevalence of high-stakes testing, and the significant consequences for students became apparent. By the beginning of the twenty-first century, the long-lasting effect of classroom assessment on students' identities and learning trajectories was also recognized. Democratic values now call for inclusive educational systems that recognize and support student diversity. From a technical perspective, this requires attention to multiple factors in the design, administration, use and consequences of educational assessment to ensure that some students are not favored over others. Specifically, when student characteristics and abilities (e.g., gender, ethnicity, language) are not relevant to the construct (i.e., subject or concept) being assessed, they should not affect the results. In sum, while examinations have not historically always been fair or democratic in intent or consequence, a democratic imperative for fairness in assessment has evolved from the tradition of selection by examination.

The second imperative relates to measurement. The quality of information that assessments provide is affected by stakeholders' perceptions of fairness. Students' perceptions influence their motivation in the assessment process, their degree of engagement, and thus the degree to which they demonstrate learning. When principals and teachers administer externally developed assessments in schools, their perceptions of fairness influence how they present the materials to students and the degree to which they follow the developers' instructions. Perceptions of fairness should be taken into account in the development and revision of an assessment tool for the same reason that face validity and credibility are considered. When stakeholders view an assessment as fair and meaningful, genuine participation is more likely, as opposed to superficial compliance or disengagement. Additionally, information gathered from one group can be used for revision to improve subsequent perceptions of fairness, thus further increasing the quality of the results.

The third imperative is pedagogical. As theories and knowledge about learning have evolved, educational assessment has increasingly been seen as a social process that shapes identity and influences opportunity to learn. Interest in using assessment for learning has grown internationally. For assessment to serve this purpose effectively, students must be willing to reveal what they do and do not know, and teachers must be able to recognize and act on the information students provide. This necessitates a trusting, respectful, and engaging learning environment where students feel psychologically safe enough to make mistakes. Students are also more likely to engage

when learning expectations and assessment criteria are clear and relevant, and they are more likely to continue learning when assessment results (e.g., feedback) and interactions (e.g., peer comments) are meaningful. The beliefs and knowledge that teachers and students bring to teaching and learning interact with different elements of fairness (next section). The fairness of classroom assessment, regardless of purpose, is affected by classroom relationships and interactions, which in turn influences the learning environment. When this circular process is scaled up from classrooms to systems, based on the assumption that all educational assessments should ultimately inform teaching and learning, the pedagogical imperative for fair assessment is at the heart of the enterprise.

The Evolution of Fairness in Educational Assessment

The field of educational assessment emerged in the twentieth century as a result of two earlier developments. Interest in the mind had taken a decidedly scientific turn in the latter half of the nineteenth century as experimental psychologists attempted to measure human intelligence. At the same time, student numbers were rising dramatically in many nations with the growth of compulsory public education. Various systems were borrowed from universities for grading achievement, which resulted in mounting concern about the assignment of students' grades. Assessment methods at the time (i.e., essays and oral recitations) permitted a great deal of subjectivity in grading decisions, and there was wide variance in the criteria and evidence being used by teachers. New psychometric methods held promise for greater consistency, and the subsequent proliferation of educational tests were considered fairer because they could be objectively scored. In retrospect, the initial response to concerns about fairness in educational assessment focused primarily on the issue of reliability. Although progressive alternatives were proposed, they were not popular at the time, and objective tests won the day.

Scientific technique and standardization had considerable appeal for efficiency in expanding educational systems, and testing was increasingly common after WWI, particularly in American schools. However, earlier charges of racism following the development of intelligence tests continued to surface. By the end of the 1960s, attempts to build "culture-free" tests had failed, and measurement specialists were looking for algorithms for fair use of test results by employers and educational institutions. The 1974 edition of the *Standards for Educational and Psychological Tests* explicitly acknowledged the potential unfairness of testing, particularly for women and minority groups, and called on developers to examine test items for bias. The terms fair and unbiased were often used as synonyms up to this point, but the meaning of bias narrowed with advancements in statistics. Bias in measurement refers specifically to systematic error in test scores, which occurs when factors that are irrelevant to the intended construct of an assessment influence the results. Because evidence of bias can be found through statistical analysis of differences in how groups of test takers respond (e.g., differential item functioning), unfairness in external assessment still tends to be equated with bias.

Toward the end of the twentieth century, the technical focus that had dominated discourse and research in educational assessment began to shift. The value-laden nature of assessment and the need to consider its consequences were emphasized in validity theory. Debate ensued on several fronts, particularly about ownership of responsibility for assessment consequences (i.e., test developers or users) and about the feasibility of validation for all uses of an assessment, especially those beyond what the developers intended. Measurement and assessment specialists continue to disagree about the scope of validity, diverging on whether it is purely a measurement concept or a matter of ethics. This raises questions regarding the relationship between validity, reliability, and fairness. Overlapping areas on a Venn diagram with three circles representing these key qualities could change considerably according to differing perspectives or purposes in assessment. Nonetheless, clear definitions are needed for key qualities

to guide assessment practices. At present, general agreement exists on two points. First, fairness is similar to validity and reliability in that it is not dichotomous, which means that it is determined by degree. Second, unlike validity and reliability, fairness is not in itself a technical quality, but it is affected by technical quality. Fairness is a requirement for the ethical practice of educational assessment, which in turn contributes to the broader matter of social justice.

Achieving Fairer Educational Assessment

Fairness in educational assessment is complex, and it cannot be ensured through any one practice. Fairer assessment can be achieved in different ways, with some conditions or strategies being more important in some situations than others, depending on the purpose of the assessment and the individuals assessed. Conditions and strategies for fairness should be considered proactively in the design and development of assessment tools and tasks, continually through assessment interactions, and retrospectively in reviewing the assessment process.

Three conditions for fairer educational assessment are opportunity to learn, a constructive environment, and evaluative thinking. *Opportunity to learn* is a seemingly self-defining term that can vary considerably in breadth. It can simply mean exposure to test content or refer more broadly to the alignment between curriculum and assessment. It can also refer to a gamut of socioeconomic and educational factors that enable learning, including the availability and quality of resources (i.e., teachers, learning materials, technology, etc.) and students' ability to use them within an environment or system. While ensuring opportunity to learn for every student in the fullest sense is a social justice issue beyond the scope of any one assessment, it should be considered in planning and interpreting results in both classroom and external assessment. A *constructive environment* is one that respectfully encourages students to fully participate and disclose their knowledge and learning through assessment. This requires an extent of buy-in, meaning that an assessment must be perceived as worthwhile, or at least necessary. Interactions between students, teachers, principals, and parents can affect the environment for external assessments and hence the quality of results. For classroom assessment to genuinely and openly serve learning, high levels of trust and respect must be nurtured, not only between teachers and students but also between classroom peers. Creating constructive environments for fairer assessment requires *evaluative thinking*. This involves asking questions, identifying assumptions, seeking evidence and considering different explanations, or in brief, critically evaluating assessment practices. In external assessment, evaluative thinking should be part of a formal process (i.e., validation) that draws on qualitative and quantitative evidence for fairness (e.g., panel reviews, DIF). In classroom assessment, teachers' self-evaluation about assessment tools, tasks, and interactions should be part of reflective practice. Reflection is particularly important for recognizing assumptions or beliefs that might lead to bias, and for receptivity to the knowledge and learning of diverse students, even when it diverges from the expected. Most importantly, all educational assessments benefit from the acceptance of responsibility for fairness and thoughtful planning, administration, and interpretation.

Multiple strategies can and should be used at each phase to ensure fairer educational assessment. Several strategies relate to the principle of *transparency*. It is widely accepted that students should know how their work (i.e., performance, product, responses) will be judged before an assessment begins. A basic strategy is to provide clear instructions. Assessment criteria that flow logically from learning expectations should also be made explicit. A challenge in this process is to specify criteria sufficiently for students to understand it without constraining what they demonstrate or what is evaluated. Additional criteria should not come into play in the marking process (or scoring, grading). This is especially problematic in performance assessments where differing responses are possible (e.g., problem solving) and personal preferences creep into the judgment

process. Transparency makes assessments fairer by reducing the potential influence of irrelevant factors. It also supports the primary purpose of classroom assessment, particularly when the formative process should in itself be an opportunity to learn.

A second set of strategies revolves around the provision of *opportunity to demonstrate learning*. Students should have multiple, varied, equitable, and meaningful opportunities to demonstrate their learning. Offering students multiple opportunities allows teachers to gather sufficient information to make fairer decisions (i.e., increasing reliability), especially in summative assessment that affects placement or certification. Offering students varied opportunities allows learning to be demonstrated in different ways and enables different types of learners to succeed. Because assessment formats (e.g., multiple choice, essay) affect students differently, using varied methods prevents any one type from overly benefiting or penalizing students. Providing students with equitable opportunities, where appropriate accommodations allow individuals or groups to demonstrate what they know, also makes assessment fairer. A tension here goes back over a century to when it was noted that fair comparison of scores was not possible with inconsistent grading. The need for consistency (i.e., equal treatment) versus the need for differentiation (i.e., equitable treatment) depends on whether the purpose of the assessment is to compare scores or support learning. Whenever assessment results are used to distribute a benefit or opportunity, there should be a consistent basis for comparison. This is why the comparability and appropriateness of procedures, including accommodations, are important in external testing and when summative grades are used for external decisions (e.g., admissions). In contrast, when the goal is to advance learning, assessment can be individually tailored, such as in adaptive computer assessment. A final strategy for fairness, offering students meaningful opportunities to demonstrate learning, requires a balance between two core qualities in the ethics of teaching: care and respect. Care must be tempered with respect for students' learning to ensure that opportunities are engaging without being superficial, and that they are genuinely challenging without being impossible. Assessment that recognizes and responds to the knowledge and learning of diverse students is fairer in the long term because it enables subsequent opportunities to learn. In essence, educational assessment should give all students a fair shake at success in their learning careers.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Buckley, J., Archibald, T., Hargraves, M., & Trochim, W. M. (2015). Defining and teaching evaluative thinking: Insights from research on critical thinking. *American Journal of Evaluation, 36*(3), 375–388. doi:10.1177/1098214015581706.

Cowie, B. (2014). Equity, ethics and engagement: Principles for quality formative assessment in primary science classrooms. In C. Milne, K. G. Tobin, & D. Degenero (Eds.), *Sociocultural studies and implications for science education* (Vol. 12, pp. 117–133). Dortrecht: Springer.

Dorn, S. (2014). Testing like William the conqueror: Cultural and instrumental uses of examinations. *Education Policy Analysis Archives, 22*(119), 1–11. doi:10.14507/epaa.v22.1684.

Fischer, D. H. (2012). *Fairness and freedom: A history of two open societies, New Zealand and the United States*. New York: Oxford University Press.

Giordano, G. (2005). *How testing came to dominate American schools: The history of educational assessment*. New York: Peter Lang Publishing.

Gipps, C., & Stobart, G. (2009). Fairness in assessment. In C. Wyatt-Smith & J. J. Cumming (Eds.), *Educational assessment in the 21st century* (pp. 105–118). Netherlands: Springer.

Klenowski, V. (2015). Fair assessment as social practice. *Assessment Matters, 8*, 76–93. doi:10.18296/am.0005.

Moss, P. A., Pullin, D. C., Gee, J. P., Haertel, E. H., & Young, L. J. (2008). *Assessment, equity and opportunity to learn*. Cambridge: Cambridge University Press.

Newton, P. E., & Shaw, S. D. (2014). *Validity in educational and psychological assessment*. London: SAGE.

Tierney, R. D. (2013). Fairness in classroom assessment. In J. H. McMillan (Ed.), *SAGE handbook of research on classroom assessment* (pp. 125–144). Thousand Oaks, CA: SAGE.