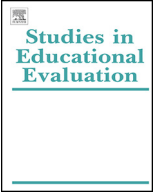




Contents lists available at [ScienceDirect](http://www.sciencedirect.com)

Studies in Educational Evaluation

journal homepage: www.elsevier.com/stueduc



Fairness as a multifaceted quality in classroom assessment

Robin D. Tierney*

University of British Columbia, MERM, Faculty of Education, Vancouver, British Columbia, Canada

ARTICLE INFO

Article history:

Received 15 May 2013
Received in revised form 8 November 2013
Accepted 17 December 2013

Keywords:

Fairness
Quality assurance
Classroom assessment
Assessment for learning
Phronesis
Qualitative research

ABSTRACT

Fairness is an essential and enduring ideal in education, but it has not been clearly defined for the dynamics of classroom assessment. This paper aims to contribute to the reconceptualization of fairness as multifaceted quality in classroom assessment where the primary purpose is to support student learning. This multi-case study elicited the phronesis (practical wisdom) of six purposefully selected teachers in Ontario, Canada. They responded to fairness issues in written vignettes, and then discussed their concerns and gave recommendations for fair assessment during interviews. The participants emphasized different aspects of fairness with the most prominent involving students' opportunities to learn and demonstrate learning, transparency, the classroom environment, critical reflection, and the tension between equal and equitable treatment in assessment.

© 2014 Published by Elsevier Ltd.

Quality assurance is a proactive process for determining and monitoring, either internally or externally, that a program or practice fulfills its purpose and meets stakeholder requirements (Doherty, 2008; Martin & Stella, 2007). Qualities that are desirable for educational assessment are usually identified in measurement theory as validity, reliability, and fairness (e.g., American Education Research Association [AERA], American Psychological Association [APA], and National Council on Measurement in Education [NCME], 1999). These qualities can be considered for different levels of educational assessment, from individual learner diagnostics to system-wide accountability testing. In this paper, I focus specifically on reconceptualizing fairness as a multifaceted quality in classroom assessment (CA) that aims to support learning.

CA is an ongoing process that involves teachers and students in gathering information (assessing) and making judgments (evaluating) about student learning. CA results have traditionally been used to determine and report on achievement in order to place or certify students. This is referred to as summative assessment or assessment of learning (AofL). The use of formative assessment or assessment for learning (AFL) has been increasingly endorsed in the educational assessment community (e.g., Assessment Reform Group [ARG], 1999; Earl & Katz, 2006; Stiggins & Chappuis, 2005). AFL involves sharing clear expectations and criteria, using varied methods to elicit learning, giving task-specific feedback to students, involving them in assessment processes, and using the

results to inform teaching (Tierney & Charland, 2007). William (2011) emphasizes that for assessment to effectively support learning, it must provide specific information, not only to direct further teaching, but also to encourage student engagement in productive learning. The need for AFL is now broadly recognized (e.g., Gordon Commission, 2013; Ontario Ministry of Education, 2010), but little research has been done on fairness issues in its practice.

Going back and moving forward

During most of the 20th century, the educational measurement community focused on the development of standardized tests. To a large extent, this was because of the widely held belief that objectivity could be attained through the application of scientific technique (Clarke, Madaus, Horn, & Ramos, 2000), and inferences from test results were thought to have a higher degree of validity, reliability and fairness than CA results. As the century turned, shifting social ideals, evolving ideas about the nature of knowledge, developments in understanding human learning, and rapid technological advancements changed the educational landscape. The negative impact of high-stakes testing on teaching and learning was increasingly recognized (e.g., Abrams, Pedulla, & Madaus, 2003; Frederiksen & Collins, 1989; Gipps & Murphy, 1994; Heubert & Hauser, 1999; Popham, 2003), which not only heightened concerns about quality in testing, but also generated interest in developing and using large-scale performance assessments that would support student learning and measure achievement. The challenges of ensuring high technical quality in performance assessments quickly became apparent, and expert

* Tel.: +1 604 401 6217.

E-mail addresses: robin.tierney@ubc.ca, tierney@research-for-learning.com

voices repeatedly cautioned that an assessment should not be considered fairer simply because it aims to support learning (Elwood, 2006a; Gipps, 2005; Linn, Baker, & Dunbar, 1991).

Interest in the pedagogical potential of CA increased through this period, particularly with the emergence of research on the benefits for student learning (e.g., Black & Wiliam, 1998). In comparison to standardized testing, CA was commonly considered to be low-stakes. However, a growing body of research has shown that CA does affect student motivation and self-regulation in learning (Brookhart, 2013), and there can be long-term social, emotional, and academic consequences for students (Brookhart, 1994; Dalbert, Schneidewind, & Saalbach, 2007; Morgan & Watson, 2002). Research on the quality of 'newer' assessment methods shows mixed results (Dierick & Dochy, 2001), and some are potentially less fair than traditional tests because of their personal nature (e.g., reflective response journals) (Gynnild, 2011; Schendel and O'Neill, 1999). Recognizing the high-stakes of CA for learners brings us full circle to concerns about quality similar to those expressed by early 20th century edumetricians before the heyday of standardized testing (Finklestein, 1913; Monroe, 1923; Rinsland, 1937).

Efforts to assess the quality of CA from a measurement perspective have generally resulted in teachers' technical ability or assessment literacy being "found wanting" (Brookhart, 2004, p. 447). Rather than assuming that quality problems were caused entirely by poor practice, many assessment specialists at the end of the 20th century began to question the relevance of measurement theory for the dynamics of CA (Brookhart, 1993; Delandshere, 2001; Gipps & Murphy, 1994; Stiggins, Frisbie, & Griswold, 1989; Whittington, 1999; Wiggins, 1993; Wiliam, 1994). This fueled the development of two documents containing principles or standards for CA in North America (Joint Advisory Committee [JAC], 1993; Joint Committee on Standards for Educational Evaluation [JCSEE], 2003). The idea that CA should be fair is inherent to these documents, but neither explicitly defines the concept. This fuzziness around fairness remains to date despite the sustained thrust to re-conceptualize quality for CA. Considerable discussion pertains to validity and reliability (e.g., Black & Wiliam, 2006; Bonner, 2013; Moss, 2003; Parkes, 2007, 2013; Smith, 2003; Stobart, 2006), but much less focuses on fairness, and little guidance for fair AfL is given. A better understanding of fairness in CA is needed, especially for AfL with diverse students. I aim to contribute to that understanding with this multi-case study on teachers' phronesis about fairness. In the balance of this paper, I explain the rationale for my research approach, and I identify existing interpretations of fairness. I then describe the methodology and results, and link them back to existing literature in the discussion.

Rationale for turning to teachers for phronesis

For quality assurance in education, Doherty (2008) suggests that when "you want to improve something, ask the people who do it" (p. 82). Baartman, Bastiaens, Kirschner, and van der Vleuten (2006, 2007) took this approach when they consulted 12 international measurement experts to re-frame quality criteria for competency-based assessment programs, and subsequently surveyed Dutch vocational teachers regarding their framework. My research has a similar goal, with two main differences. First, I am concerned more specifically with AfL, which involves both planned events and spontaneous interactions (Cowie & Bell, 1999). Thus my interest in fairness extends beyond tests to the nebulous space between teachers, students and students' learning where inferences and decisions are made, often quickly and tacitly. Second, rather than seeking opinions on a framework developed by

experts, I turned to teachers to shed light on a concept in their practices in order to improve the relevancy of CA theory.

The philosophical perspective underpinning this work is a form of critical pragmatism. Warms and Schroeder (1999/2009) describe pragmatism as a "way of doing philosophy that weaves together theory and action, each continuously modifying the other and maintain their mutual relevance" (p. 271). From this perspective, varied forms of knowledge are valuable for educational research (Biesta & Burbules, 2003; Maxcy, 2003), including ethical and practical knowledge. These are undervalued relative to theoretical and empirical knowledge (Fenstermacher, 1994; Kessels & Korthagen, 1996), and greater consideration should be given to the phronesis that guides practice (Bernstein, 1985; Biesta, 2007; Dunne, 1993). Phronesis is translated from classical Greek as practical wisdom, reasoning, or judgment (Dottori, 2009; Fenstermacher, 1994; Flyvbjerg, 2004). Aristotle defined it as a "reasoned and true state of capacity to act with regard to human goods" (Aristotle, trans. 1925/1975, p. 143). In contemporary terms, phronesis draws on a mental network that includes technical knowledge, theoretical knowledge, moral beliefs and professional ethics, personal characteristics, experience, and understanding of particulars (Connolly, 2005; Dunne, 1993). Essentially, it provides the ability "to judge correctly, all things considered, the best action to perform in any given situation" (Warne, 2006, p. 15). Phronesis shares characteristics with other conceptualizations of teachers' knowledge (e.g., Clandinin & Connelly, 1996; Shulman, 2004a), but it also includes a moral dimension that is key for research into the ethics of practice.

Interpretations of fairness in assessment theory

In the literature that aims to reconceptualize quality for CA, what or where we are moving from is not always clear, possibly because describing this is like painting a cloudy sky on a windy day. The meaning of fairness continues to evolve, along with other key qualities, in measurement theory. In early editions of testing standards (APA, AERA, & NCME, 1966, 1974), the terms *biased* and *unfair* were used interchangeably, but now fairness is recognized as a broader social concern that goes beyond technical issues of bias (Camilli, 2006; Moss, Pullin, Gee, & Haertel, 2005; Stobart, 2005). Four interpretations of fairness are discussed in the last edition (AERA, APA & NCME, 1999). The first two, being the absence of statistical bias and the equitable treatment of test takers in the testing process, have generally been accepted in the measurement community. The third interpretation, where fairness is associated with equality of test outcomes, is rejected based on the long-established point that group score differences do not necessarily indicate test bias (e.g., Messick & Anderson, 1970). The fourth interpretation, where fairness is associated with opportunity to learn is considered problematic, but it is acknowledged that prior access to test material is necessary for decisions based on test results to be fair. Discussion about fairness in testing is ongoing in the measurement community (e.g., responses to Xi, 2010). Nonetheless, there is general agreement on two points. First, fairness cannot be determined dichotomously because it is a matter of degree (Cole & Zieky, 2001; Lane & Silver, 1999). And second, it is an important quality that is distinct from, but related to validity (Camilli, 2006; Messick, 2000; Stobart, 2005).

While the meaning of fairness continues to evolve in testing, it is not defined in the principles and standards for CA (JAC, 1993; JCSEE, 2003), and research focusing specifically on fairness in CA is limited (Tierney, 2013). In preparation for this multi-case study, I culled existing interpretations from a range of texts (i.e., peer reviewed articles, CA textbooks, doctoral dissertations, and joint committee documents). Some interpretations conflict while others

Download English Version:

<https://daneshyari.com/en/article/6849186>

Download Persian Version:

<https://daneshyari.com/article/6849186>

[Daneshyari.com](https://daneshyari.com)