

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/260144263>

Tierney, R. D. (2013). Fairness in classroom assessment. In J. H. McMillan (Ed.). SAGE Handbook of Research on Classroom Assessment. Thousand Oaks, CA: Sage Publications.

Chapter · January 2013

CITATIONS

0

READS

3,281

1 author:



[Robin Tierney](#)

Research-for-Learning

11 PUBLICATIONS 204 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Research on Grading in Ontario, Lead Researcher Marielle Simon [View project](#)



Formative Assessment Literature Review [View project](#)

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 Publications.

8

FAIRNESS IN CLASSROOM ASSESSMENT

ROBIN D. TIERNEY

Fairness is an ideal that has pervaded the history of education in democratic societies, from the establishment of merit-based systems centuries ago to the organization of contemporary classrooms. Common classroom practices, such as raising hands and taking turns, reflect shared beliefs about fairness (Campbell, 2003; Jackson, Boostrom, & Hansen, 1993; Shulman, 1987/2004). Fairness is considered a desirable quality in a wide range of educational assessments, from the diagnosis of individual learning needs to large-scale international surveys of student achievement. As Green, Johnson, Kim and Pope (2007) note in their work on the ethics of assessment, fairness is a "general principle that no one contests in the abstract" (p.1001). In practice, however, the fairness of assessment is less readily assumed and far more contentious.

Concern has long been expressed about the fairness of educational assessment. Early in the 20th century, objective tests were hailed as a means of bypassing the "injustice" caused by teachers' inconsistent grading practices (Finklestein, 1913, p. 6). In subsequent decades, educational measurement experts concentrated on the development and use of standardized tests, but research on classroom assessment (CA) was limited. During the late 1960's and 1970's, political upheaval, shifting social ideals, and new insights in learning theory affected the educational landscape (Bredo, 2006; Giordano, 2005; Gipps, 1999). By the close of the century, reform was underway in many educational

systems. As part of this process, there was a growing interest in performance-based or authentic assessment (Clarke, Madaus, Horn & Ramos, 2000; Horn, 2002; Lam, 1995), emphasis on the pedagogical potential of formative assessment (Black & Wiliam, 1998; Shepard, 2006), increasing recognition of the need to serve diverse students (Kornhaber, 2004; Nettles & Nettles, 1999), and greater awareness of the social, political, and value-laden nature of assessment (Gipps, 1994; Horn, 2002). These changes led to questions about the relevance of measurement theory for the dynamics of CA, and they gave rise to new concerns about fairness for the 21st century.

At present, consensus regarding the meaning of fairness in CA is limited in both practice and theory. While teachers generally aim to be fair in their assessment practices (Brookhart, 1994; Tierney, Simon & Charland, 2011), their interpretations of fairness vary considerably (e.g., Yung, 2001; Zoeckler, 2005). The statistical techniques that help determine fairness in testing are not useful for teachers because of the time constraints and the relatively small number of students typically involved in CA (Camilli, 2006). While measurement theory supports the development and use of standardized and large-scale assessment, its appropriateness for CA was strenuously questioned during the 1990's (e.g., Brookhart, 1993; Delandshere, 2001; Gipps, 1994; Moss, 1994; Shepard, 2000; Stiggins & Bridgeford, 1985; Whittington, 1999; Wiggins,

1993; Wiliam, 1994). The reconceptualization of validity and reliability for CA has since been ongoing (e.g., Black & Wiliam, 2006; Moss, 2003; Parkes, 2007; Smith, 2003; Stobart, 2006; see also Chapters 6 and 7, this volume), but less attention has been paid to fairness. Although the moral intent underlying the quest for fairness in standardized testing and CA may not be all that different, the concept of fairness needs reworking to take the dynamics and purposes of CA into account.

Greater interest in formative and performance-based assessment in recent years is in keeping with the shift from a testing culture to a learning culture, as described by Gipps (1994) and Shepard (2000). However, caution has been repeatedly voiced regarding the assumption that these forms of assessment are necessarily fairer than traditional written tests (Elwood, 2006; Gipps, 2005; Linn, Baker & Dunbar, 1991). Some alternative methods, such as discussion circles, reflective journals, and self-assessments, may even have a greater potential for unfairness because they involve personal contact or ask students to reveal weaknesses (Brookfield, 2001; Gynnild, 2011; Schendel & O'Neill, 1999). The conceptual shift from a testing to a learning culture alters the relationships and power dynamics among peers and between students and teachers, resulting in a learning environment that has yet to be fully explored in terms of fair assessment. Furthermore, it is now widely recognized that teachers' assessments, which were once considered low-stakes in comparison to standardized tests, can also have significant and long-lasting personal, social, and educational effects for students. For example, the results of summative CAs, such as final examinations or culminating projects, weigh heavily in the selection and placement decisions made for students in many educational systems. Consequently, they can play the same "gatekeeping role" (Nagy, 2000, p.262) as high-stakes tests. Informal and spontaneous judgments by teachers can also affect students' self-identity, motivation, and approach to learning, which in turn influence teachers' perceptions in ongoing classroom interactions (Brookhart, 1994; Cowie, 2005; Gipps, 1994; Morgan & Watson, 2002; Watson, 1999). Given current ideals regarding

the use of assessment for multiple purposes, changing classroom dynamics, and the impact this may have on diverse learners, strong imperative now exists for investigating fairness as a quality of CA.

The purpose of this chapter is to review how fairness has been interpreted from a range of perspectives in order to inform its conceptualization for CA. Specifically, the aim is to answer the following two questions: (1) How has the concept of fairness been interpreted in different types of texts? (2) Which interpretations of fairness are most relevant for the purposes of CA? The goal is not to prescribe how fair assessment should be practiced by all teachers, but to provide avenues for further research, discussion, and guidance relating to fairness in CA. An assumption in undertaking this review is that fairness is for the student. What might or might not be fair for teachers, school or system administrators, and test developers involved in educational assessment is beyond the scope of this work.

The next two sections of this chapter provide a background for understanding fairness in CA, first by looking at common usage of the word *fair* and then by tracing the development of the concept in educational measurement. While this may seem peripheral to research on CA, ideas about fairness have not developed in isolation. Examining the link to daily language and measurement terminology illuminates the "a priori reasoning" (Bredo, 2006, p.4) embedded in thinking about fairness for CA. This is followed by an overview of the different types of literature that contain interpretations of fairness, and by a discussion of research relating to key aspects of fairness in CA.

THE MULTIPLE MEANINGS OF *FAIR*

Like many words in the English language, the word *fair* has multiple definitions. As an adjective or adverb, its meaning changes according to the noun or verb it describes. Fair hair is light, fair weather is temperate, and a fair deal is straight, especially if it's square. Fair play involves following the rules of the game, whereas anything that is fair game is open to attack. When someone

gets a fair shake or a fair go it means they have a clear chance. In contrast, rating a service as fair indicates that it is less than perfect, but still passable (fair to middling). When a statement is deemed fair enough it is acceptable, or just not sufficiently radical to elicit disagreement. Beneath these disparate uses of the word there often lies a sense of openness, or absence - no clouds, no shady business, no undue impedance in the pursuit of an appropriate target. This is reflected in the dictionary definition of *fair* as that which is "marked by impartiality or honesty, and free from self-interest, prejudice, or favouritism." What is most important about this openness or absence is that opportunity is not constrained.

A second sense that underlies how the word *fair* is used is that of neutral alignment. In carpentry, to square something means to check the accuracy of angles and straighten all sides. In business, fair and square means playing above board and not swaying from the law. In any kind of decision making, the opposite of fair is biased. When fabric is cut on the bias, it is cut on an angle, diagonally to the warp and weft. Thus a biased line leans one way or another, whereas an unbiased report takes a neutral stance, and an unbiased jury remains objective. To be unbiased or fair in this sense is to adhere to established principles, or a framework, in the same way that a square window fits its casing. Related to the idea of alignment is the sense of being between extremes, or balanced. Fair weather is pleasant, meaning neither too hot nor too cold. Fairly well off is comfortable, meaning neither rich nor poor. Fair trade, even-handedness, the scales of justice, and a level playing field all involve the idea of balance. As such, despite what appears to be a multiplicity of definitions, common use of the word *fair* generally conveys a sense of openness, neutrality, or balance.

THE EVOLUTION OF FAIRNESS IN TESTING

Testing is pervasive in education, and concerns about its fairness are perennial. Questions were raised about the fairness of psychological and educational tests at their inception more than a century ago (Cole & Zieky, 2001; Giordano, 2005; Kornhaber, 2004), and

fairness issues continue to figure prominently in discussions about large-scale assessment systems (e.g., National Research Council, 2010). It is, therefore, an irony in the history of educational assessment that written examinations were initially devised as a method for improving fairness. For example, entry into government positions in ancient China was gained through examinations (Ebel, 1972; Gipps, 1994; Popham, 1981), and European and North American universities began using them to select candidates for admissions during the 19th century (Camilli, 2006; Gipps, 1994). While it may seem "naive" in retrospect (Stobart, 2005, p.285), the intent of these examinations was to remove social class as a barrier and equalize opportunity in education and employment. It is now evident that by controlling access, they have the reverse effect as gatekeepers in that they bar some students from further learning opportunities.

With the advent of public education during the late 19th and early 20th centuries came an unprecedented number of students, which resulted in a need to distinguish between those capable of higher education and those destined for fields and factories. Psychology was emerging as a branch of science (rather than philosophy) around the same time, and the creation of intelligence tests strongly influenced the development of testing in education (Giordano, 2005; Kornhaber, 2004). High regard for scientific technique during this period supported the general assumption that standardized tests were inherently fairer than the subjective methods (i.e., oral recitation, essays) that had been used previously to assess student achievement (Clarke et al., 2000). However, by the 1960's, questions about the fairness of testing were sufficiently vociferous to command attention from the measurement community. In response to the growing "public hostility" to testing (Messick, 1965, p. 136), the prospect of culture-fair tests garnered interest, and there was a concerted effort to develop an algorithm for fair use of test results in the selection of candidates (Cole & Zieky, 2001). By the end of the 1970's various statistical models for fair decision making had been put forward, but were found wanting within the measurement community. Although educational reforms continued to augment the role of testing

in education (Clark et al., 2000), public faith in the inherent fairness of standardized tests was eroding.

Four revisions of the *Standards for Educational and Psychological Testing* (the first two were named the Standards for Educational and Psychological Tests) (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 1985, 1999; APA, AERA, & NCME, 1966, 1974) reflect how ideas about fairness changed during the latter half of the 20th century. While the 1966 edition does not mention fairness, the 1974 edition acknowledges that "some unfairness may be built into a test" (p. 2). The words *biased* and *unfair* were often used interchangeably at this point; essentially one was tantamount to the other, and both were subsumed by validity. However, as advancements in statistical techniques allowed more vigorous examination of how test items functioned, biased items that favoured or disadvantaged different groups of test-takers could be identified (Cole & Zieky, 2001). As a result of this development, clearer distinction was made between bias and fairness. The 1985 edition of the *Standards* reflects this in stating that unlike bias, "fairness is not a technical psychometric term" (p.13). The word *fairness* continued to be used in reference to the utilization of test results, but it was excluded from the technical terminology of measurement. In contrast, bias was increasingly used as a measurement term (e.g., Childs, 1990), but its meaning was restricted to the idea of neutrality rather than the sense of openness that is also seen in common usage of the word *fair*.

With the evolution of validity theory and Messick's (1989) emphasis on consequential evidence, awareness was heightened in the measurement community about the role of values and ethics in testing (e.g., Moss, 1998; Shepard, 1997). While fairness had been vaguely positioned under the broader umbrella of validity, by the end of the century it had become comparable to validity, and discussion about the quality of assessment had become as much about ethics as metrics. The revision of the *Standards* in 1999 illustrates this change with four chapters in a section devoted specifically to fairness.

Previous editions contained similar content, but the gathering of these chapters under one heading reflects the change in the significance accorded to fairness. The *Standards* (AERA et al., 1999) emphasize the "importance of fairness in all aspects of testing and assessment" (p.73). Four interpretations are discussed in defining fairness. Two of these – (1) the absence of statistical bias and (2) equitable treatment in the testing process – are described as having generally been accepted in the measurement community. A third interpretation, equality of test outcomes, is rejected. This rejection is hardly radical given that measurement experts have argued for decades that group score differences do not necessarily indicate test bias (e.g., Messick & Anderson, 1970). The fourth interpretation, opportunity to learn, is less straightforward. Several problems with this interpretation are identified for the context of testing, but the *Standards* (AERA, et al., 1999) acknowledge that opportunity to learn test material is necessary for test-based decisions to be fair. As a whole, these interpretations indicate that fairness is larger than and inclusive of bias. Fairness in testing also includes equitable treatment and access to test materials. Equitable treatment involves a balance, and access requires some degree of openness. Thus, the qualities that underlie common usage of the word *fair* are also reflected in the interpretations of fairness for testing.

At present, it is generally accepted in measurement theory that fairness is an important quality that is distinct from but related to validity (Camilli, 2006; Messick, 2000; Stobart, 2005). Like validity, fairness cannot be determined dichotomously because it is a matter of degree (Cole & Zieky, 2001; Lane & Silver, 1999). Research and debate on the relationship between validity and fairness is ongoing (e.g., Xi, 2010 and responses by Davies, 2010; Kane, 2010; Kunnan, 2010), but the earlier distinction made between bias and fairness has been maintained. Camilli (2006), for example, explains that bias is a measurement issue whereas fairness relates to "factors beyond the scope of the test" (p. 225). According to Stobart (2005), fairness is a "social process" and a "qualitative judgment" (p. 285) that must take the curriculum and students' learning opportunities into account. Moss, Pullin,

Gee and Haertel (2005) argued further that the influence of measurement and testing on constructs such as opportunity and success in education must be considered in terms of social justice. Through ongoing revisions to the standards, and as thinking about testing evolves, it has become evident that fairness is a complex issue and an enduring concern in education.

OVERVIEW OF LITERATURE RELATING TO FAIRNESS IN CLASSROOM ASSESSMENT

Until recently, most research on fairness in educational assessment has pertained to standardized testing or large-scale assessment. Changes in the educational landscape have provided the imperative for investigating fairness as a quality of CA, but three challenges are apparent in this process. The first is that no single context exists for CA. An enormous diversity of learners and subjects is involved in classroom learning, which makes possible a wide range of circumstances and methods for CA. What might be fair, for example, in determining the literacy of an English-language learner (ELL) at the beginning of term in a primary classroom, might not be fair when medical students take their final pharmacology exam in a university classroom. To discuss the characteristics of the concept as a whole, it is necessary to generalize across classroom contexts, but the particulars of practice should be kept in mind.

The second challenge is that despite being frequently mentioned in CA literature, fairness is not often defined. This may be because fairness is considered an essential virtue in teaching practice (Campbell, 2003) or because fair assessment is commonly accepted as a student's right (Dann, 2002; Robinson, 2002). The definitions and interpretations of fairness that are found in the extant literature vary, and they are sometimes conflicting. This section examines these definitions and interpretations in comparison to those in testing and common usage in order to provide a fuller sketch of the concept than seen to date.

A third complicating factor in this process is the plethora of terminology. Related terms -

specifically objectivity, bias, values, ethics, equality and equity - are often used in conjunction or interchangeably with fairness. This sometimes results in circularity, such as defining fair assessment as "equitable assessment" immediately after stating that equity is "about fairness" (Szpyrka, 2001, p.5). As precision and consistency are currently lacking in the terminology, literature using related terms that pertains substantively to fair CA is included.

Research literature documents the questions asked and answers given, the discussion of ideas, and the statement of ideals within a community. As such, it is a form of dialogue among stakeholders, and different types of literature capture different aspects of this dialogue. In the following sections, literature is discussed in two main categories: (1) conceptual and (2) empirical. Within these are subcategories that reveal differences in how fairness in CA is interpreted by groups of stakeholders, particularly students, teachers, and researchers.

Conceptual Literature

Three different types of texts are identified as conceptual here because they discuss concepts, usually for the purpose of directing research or practice. The first, committee documents, are developed by members of educational organizations to provide principles or standards for practice. The second contains practical guidance on fair assessment for classroom teachers, and the third aims to further scholarly discussion within the academic community. While some texts are clearly of one type or the other, many contain both practical guidance and theoretical discussion, and others are situated in the movement between practice and theory. As such, the second and third types of texts are discussed together under the rubric of CA theory.

Committee Standards

Two major documents containing principles or standards specifically for CA are the *Principles for Fair Assessment Practices for Education in Canada* (Joint Advisory Committee, 1993) and *The Student Evaluation Standards* (Joint Committee on Standards for Educational

Evaluation, 2003). Because these documents were collaboratively constructed with input from many educational organizations, they carry the weight of consensus in reflecting the ideals of the educational community at the time of their publication.

The *Principles* (Joint Advisory Committee, 1993) were developed primarily in response to inappropriate use of large-scale assessment results in Canada (T. Rogers, personal communication, May 10, 2011), but they also include a set of guidelines specifically for CA. Fairness is not explicitly defined, but the principles are said to be "indicative of fair assessment practice" (p.3) overall. Although they were forward-looking at the time, their sequenced structure is suitable for planned, summative assessment, and there is very little guidance for fairness in informal assessment interactions. The evolution of assessment ideals over the following decade is reflected in *The Student Evaluation Standards* (Joint Committee on Standards for Educational Evaluation, 2003), which presents assessment as "central to student learning" (p.1). Once again, fairness is not explicitly defined, but 21 of the 28 standards are cross-referenced to fairness in the functional table of contents. As a result, fairness is associated with a wide range of related assessment qualities, including sound, equitable, consistent, balanced, ethical, feasible, accurate, and useful. This breadth supports the idea that ensuring fairness is not a straightforward process, but such a nebulous definition limits the usefulness of the guidance.

Both the *Principles* (Joint Advisory Committee, 1993) and *The Student Evaluation Standards* (Joint Committee on Standards for Educational Evaluation, 2003) emphasize that the purpose, criteria, and results of any assessment should be explicitly and clearly communicated. This is consistent with the sense of openness in the common usage of the word *fair*, and it associates fairness with the concept of transparency in assessment (Frederiksen & Collins, 1989). As the *Principles* (Joint Advisory Committee, 1993) and *The Student Evaluation Standards* (Joint Committee on Standards for Educational Evaluation, 2003) were written a decade apart, it also suggests that the association between fairness and transparency has been

accepted in the educational community for some time. In addition, both documents are concerned with administrative issues that extend beyond assessment interactions in classrooms, such as maintaining the confidentiality of assessment results and records at the school and system levels. Fairness is presented in both documents as a general ideal that pervades all assessment phases and concerns all stakeholders. In comparison to the *Standards* (AERA et al., 1999) for testing, the principles and standards for CA delineate fairness less clearly, but the breadth of associated issues far exceeds the few interpretations that are accepted for testing.

Classroom Assessment Theory

The ethics of CA have not been addressed comprehensively in assessment theory. Discussion pertaining to fairness in CA is scattered through journal articles, book chapters, textbooks, monographs, and professional magazines. In this overview, references are generally given as examples as more comprehensive discussion of the key aspect of fairness follows later in this chapter.

CA theory differs considerably in terms of perspective. For example, it includes measurement (e.g., Camilli, 2006), post-modern (e.g., Schendel & O'Neil, 1999), critical theory (e.g., Tierney, 1998/2005), social justice (e.g., Drummond, 2003), and sociocultural (e.g., Poehner, 2011) interpretations of fairness. There is also a difference in the degree of focus; in some texts fairness issues are mentioned as part of a broader discussion (e.g., Shepard, 2000), whereas others relate specifically to one aspect of fairness (e.g., Gee, 2003). Nonetheless, patterns emerge in terms of the fairness issues that are identified. In two ways, these texts are similar to the *Principles* (Joint Advisory Committee, 1993) and *The Student Evaluation Standards* (Joint Committee on Standards for Educational Evaluation, 2003). First, fairness is most frequently discussed without explicit definition. Only a handful of authors attempt to operationalize the concept with strategies or steps for practice (Camilli, 2006; McMillan, 2011; Russell & Airasian, 2012; Suskie, 2002; Volante, 2006). Second, fairness is often associated with validity and reliability.

Given the centrality of these concepts in measurement theory, this in itself is not surprising, especially in texts written by assessment specialists (e.g., Andrade, 2005; Lam, 1995; Whittington, 1999). What is interesting is the evolution of the discussion from its initial concerns about validity and reliability (e.g., Brookhart, 1994; Gipps, 1999; Schendel & O'Neill, 1999) to the reconceptualization of these qualities for CA (e.g., Moss, 2003; Smith, 2003), which has led to the realization that different criteria may be needed for fair CA (e.g., Camilli, 2006, Cowie, in press).

In CA theory, four needs that are often raised in relation to fairness are for transparent communication (e.g., Guskey & Jung, 2009), for assessments to be appropriate for individual learners (e.g., Poehner, 2011), for thoughtfulness about bias (e.g., Russell & Airasian, 2012), and for students to have opportunities to learn (e.g., Shepard, 2000). Two of these needs – (1) opportunity to learn and (2) bias – appear in the definition of fairness for testing but with significant differences. Opportunity to learn is described in the *Standards* (AERA, et al. 1999) as a social issue beyond the confines of testing, and it is accepted as a fairness issue only in the narrow terms of test preparation. In CA theory, opportunity to learn is more widely associated with the alignment of instruction and assessment, the provision of time, access to materials, the quality of student-teacher interactions, and the teacher's ability to understand and support diverse students (e.g., Camilli, 2006; DeLain 1995/2005; Drummond, 2003; McMillan, 2011). Some CA theorists discuss bias in terms of avoidance (e.g., Volante, 2006), which is similar to how it is addressed in testing (i.e., controlling construct-irrelevance). Rather than avoiding a negative, others suggest that what is needed is increased cultural awareness (e.g., Tierney, 1998/2005) and recognition of the knowledge, skills and interests that diverse students bring to the classroom (e.g., Badger, 1999; Jiménez, 2004/2005; Cowie, in press; Shepard, 2000; Tomlinson, 2005). Thus, being fair in CA requires more than being unbiased; it also involves being receptive to diversity. While a sense of neutrality underlies the meaning of fairness in both testing and CA, a greater degree of openness is reflected in the

latter. Overall, CA theory is influenced by a wide range of perspectives, and the interpretation of fairness is broader than in the definition for testing.

In contrast, the overall interpretation of fairness in CA theory is not as comprehensive as in the *Principles* (Joint Advisory Committee, 1993) and *The Student Evaluation Standards* (Joint Committee on Standards for Educational Evaluation, 2003). Some of the recommendations associated with fairness in the committee documents, such as writing clear policies, securely storing results, and analyzing statistical bias, receive scant mention in CA theory. Discussion about fairness in CA theory relates mainly to sharing expectations and criteria, providing abundant opportunity to learn, and ensuring that learners are appropriately assessed. The influence of social values and cultural expectations are frequently acknowledged, and consequently, the importance of teachers' critical reflection during and about assessment is repeatedly underscored.

Empirical Literature

While the conceptual literature represents thinking about how fair assessment should be practiced, empirical literature provides insight into how it is or is not practiced in various classroom contexts. Two groups of studies in this section serve different purposes. Studies in the first group are not about fairness per se, but fairness issues emerge from their results. They are presented here as examples, and not as an exhaustive collection. The second group consists of studies that focus intentionally on fairness in CA. These studies were identified through searches in three databases (ERIC, Scopus and Scholar's Portal) for the purpose of taking stock of what has explicitly been done or learned in this area of research.

Fairness Issues in Classroom Practice

Fairness issues are seen in two areas of classroom research: one focusing on the ethics of teaching and the other on assessment practices. Teachers who participate in classroom-based

research voice similar concerns about fairness, but these two areas of research reveal different views of fairness.

Two major works that draw on ethnographic field work to discuss ethics are by Jackson and colleagues (1993) and Campbell (2003). While these texts are primarily concerned with the influence of human morality and ethical knowledge on classroom life, they also shed light on the complexity of assessing learning. A significant issue they observe is that students' learning opportunities are restricted when uniformly low expectations are held for student learning, or when a teacher's feedback is either indiscriminant or absent. Teachers are viewed as moral agents who negotiate, whether consciously or not, between conflicting demands that relate to different perceptions or purposes of CA. Fairness from this perspective is an essential quality of both individuals and interactions in the classroom, and it is closely associated with other moral qualities such as honesty and respect. As such, fairness is seen to be in the hands of teachers, with varying degrees of ethical awareness, who strongly influence the quality of students' opportunities to learn.

Research on teachers' assessment practices tends to focus more on methods than morality. A few studies discuss transparency and differentiation in terms of fairness (e.g., Graham, 2005), but concern is most frequently expressed about reflection and consistency. The need for critical reflection, which is emphasized in CA theory, is taken up by several researchers. While the teachers in these studies do reflect about their assessment practices, their thinking is often hampered by misconceptions and traditions that are incompatible with current theory (e.g., Allen & Lambating, 2001; Crossouard, 2011; McMillan, 2003). Nonetheless, they tend to express strong convictions about their practices, particularly the notion that consistency improves the fairness of assessment.

Although consistency is ideal for system-wide assessments that rely on teachers' judgments (e.g., Wyatt-Smith, Klenowski & Gunn, 2010; Yip & Cheung, 2005), the studies discussed here involve teachers who equate fairness with using the same assessment tasks or criteria for all of the students in their classrooms (e.g., Brighton, 2003;

Eggen, 2004; Gummer, 2000; Ryan, 2000). Treating students equally may give the appearance of neutrality, and avoiding favoritism is one definition of fair, but it results in the problem highlighted by the ethical inquiries (Campbell, 2003; Jackson et al., 1993). Essentially, when equal treatment is highly valued, the appropriateness of an assessment is compromised for some students. Whether it be in terms of the method used or the feedback given (e.g., smiley faces on every science poster), students' opportunity to learn and/or demonstrate learning may subsequently be reduced. In summary, classroom-based research provides a glimpse of teachers who struggle with fairness issues in their assessment practices, but a narrower understanding of fairness is evidenced in comparison to the conceptual literature.

Research on Fairness in Classroom Assessment

The studies discussed in this section vary considerably in scope, ranging from a single reflective case (Phillips, 2002) to a nationally-sampled survey (Resh, 2009). They are gathered here because they focus on fairness issues, and they provide further evidence regarding the interpretation of the concept in CA.

A growing body of research on students' perceptions about fair CA has gained impetus by the changing needs of universities, for example, to revise programs (e.g., Duffield & Spencer, 2002), understand student evaluations (e.g., Wendorf & Alexander, 2005), reduce student aggression (e.g., Chory-Assad, 2002), and support English language learners (Knoch & Elder, 2010). A few studies are also concerned with the effect of grading practices on the experiences of secondary students (Bursuck, Munk & Olson, 1999; Dalbert, Schneidewind & Saalbach, 2007; Resh & Dalbert, 2007). Differences in students' perceptions about fairness sometimes relate to gender and culture (Mauldin, 2009; Resh & Dalbert, 2007; Tata, 2005), but two other patterns are also evident. First, students appreciate transparency, specifically in relation to the provision of study guides, explicit expectations and grading criteria, unambiguous test items, and clear feedback (e.g., Duffield & Spencer, 2002; Gordon & Fay, 2010; Robinson,

2002; Tata, 2005). Students want transparency primarily because it gives them greater opportunity to learn and/or demonstrate their learning. Second, students tend to prefer equal treatment in grading, with some exceptions for students with learning disabilities (Bursuck et al., 1999). This means that criterion-referenced assessment is considered fairest (Dalbert et al., 2007). Some students' perceptions of fairness are related to their grades (Resh & Dalbert, 2007), but procedural justice is more salient in several contexts, particularly in terms of respectful assessment interactions (Chory-Assad, 2002; Tata, 2005; Wendorf & Alexander, 2005). In keeping with these patterns, long-term memories of fair assessment stem from situations where teachers combine high expectations and clear communication, whereas memories of unfair assessment involve problems in assessment procedures and in teachers' treatment of students (Guskey, 2006).

Teachers' reflections about their assessment practices reveal that they are aware of students' perceptions of fairness (e.g., Phillips, 2002; Tierney, 2010). This may be why consistency is frequently discussed when teachers are involved in research on CA. However, in focusing on fairness, these studies reveal a greater understanding of the idea that equal treatment is not always fair (Szpyrka, 2001; Yung, 2001; Zoeckler, 2005). For example, almost all of the teachers surveyed by Green and colleagues (2007) felt that it would be ethical to accommodate a student with a learning disability during an assessment. Fairness is threatened, however, when differentiation is based on factors other than learning needs. Some teachers' judgments are influenced by student characteristics that should be irrelevant, such as gender (Peterson & Kennedy, 2006), ethnicity (Elhoweris, Mutua, Alsheikh & Holloway, 2005), socio-economic status (Miner, 1990), perceptions of ability (Resh, 2009), and behavioral stereotypes (Koch, Steelman, Mulkey & Catsambis, 2008). Teachers in these studies also continue to weigh student effort in grading (Green et al., 2007; Zoeckler, 2005; Tierney et al., 2011), which is consistent with previous research on grading from Stiggins, Frisbie and Griswold's (1989) pioneering study to the present. While

teachers consider effort out of a sense of being fair to students, especially lower-achieving students, it creates a larger problem that Brookhart (1993) describes as the "double standard of just deserts" (p.140). Some teachers openly state that they consider effort when they believe a student deserves it. However, they don't seem to question the underlying values, the currency of the information, or the quality of the process that they use to determine whether a student is deserving of special consideration. Given that students display effort differently, and that teachers' observations of effort are limited by time, class size, and their own beliefs about learning, the fairness of conflating effort and achievement is highly questionable, particularly when assessment decisions have irreversible consequences.

Almost all of this research involves practices within the realm of summative assessment, such as test preparation, grading, and making placement decisions. Very few studies have looked at fairness issues in the assessment interactions that are ongoing in classrooms. Those that do come to a similar conclusion, which in essence is that assessment information is necessarily incomplete (Bulterman-Bos, Terwel, Verloop & Wardekker, 2002; Morgan & Watson, 2002; Tierney, 2010). Incompleteness suggests missing information, which could threaten fairness, especially given the recommendations for using multiple assessments (e.g., Camilli, 2006) and the reconceptualization of reliability as a "sufficiency of information" (Smith, 2003, p.30). Students and teachers alike raise concerns about this issue (Duffield & Spencer, 2002; Tierney, 2010). However, while the completeness of information does affect the fairness of final grades or placement decisions, a shift occurs for formative assessment. What is needed is no longer a snapshot, but real-time images that are frequently refreshed with new information. Fairness in formative assessment is threatened when the information that teachers use about students' learning is not continuously renewed. This requires receptivity that goes beyond neutrality in classroom interactions, as suggested in the conceptual literature. It tasks teachers with far more than lecturing; it asks for their active and critical engagement as they continuously update

and use information about individual learners while directing and facilitating learning in the classroom.

In conclusion, this overview of the literature reveals similarities and differences in how fairness is understood in CA. In the committee documents, fairness is associated with all phases and purposes of assessment, and as a result, it is interpreted more broadly than in any other type of literature examined, including the standards for testing. Interpretations of fairness in CA theory relate primarily to four areas: (1) transparency, (2) appropriateness, (3) students' opportunities, and (4) teachers' critical reflection. Different types of empirical research reveal variations. In the ethical inquiries, fairness is an essential moral quality and teachers are moral agents who influence classroom interactions, for better or worse. In contrast, CA research tends to discuss fairness as a quality of tasks or tools, particularly those for summative assessment. A tension is evidenced between equality and equity. Equality is maintained with the same tasks and criteria for all learners, while equity involves differentiating according to students' needs. Taken together, the students and teachers' voices heard through empirical research give evidence to the complexity of fair CA. Differences in how fairness is interpreted by the various stakeholders represented in the extant literature suggest that it is a matter of perspective, meaning that perceptions of fairness will always depend on who you are and where you stand. However, some interpretations appear so frequently, especially when the flip side is taken into account (e.g., equity-equality, transparency-privacy), that fairness may be more productively considered as a single concept with multiple interrelated facets. This notion is explored further in the following section by looking more closely at specific aspects of fairness in CA.

ASPECTS OF FAIRNESS IN CLASSROOM ASSESSMENT

There is no single or direct path to guarantee fair CA; the complexity of the concept defies simple solutions. Nonetheless, it is clear from the literature that certain practices relating to

transparency, students' opportunities, equitable treatment, critical reflection, and the classroom environment can lead to fairer assessment. Given the diversity of classroom contexts, these should not be taken as a universal checklist, but as aspects to consider for the theory and practice of fair CA. The following elaboration on each of these aspects draws on the extant literature and provides suggestions for further research.

What's Not Clear About Transparency

The merits of transparency are extolled in the theory and practice of educational assessment. Standards for both CA and testing state that clear information should be provided to learners and test-takers, and the principles of assessment for learning (AFL) contain a similar statement (Assessment Reform Group, 2002). Assessment specialists have recommended for decades that explicit learning expectations and assessment criteria be shared with students (e.g., Andrade, 2005; Guskey & Jung, 2009; McMillan, 2011; Sadler, 1989; Speck, 1998; Suskie, 2002), and teachers claim wide agreement in principle (Brookhart, 1994; Green et al., 2007; Tierney, 2010; Tierney et al., 2011; Zoeckler, 2005). This would seem to suggest a lack of substantive issues relating to transparency as an aspect of fairness in assessment. Nonetheless, three concerns about transparency arise from the literature.

First, transparency can conflict with opportunity to learn and demonstrate learning, which are also aspects of fairness. On one hand, explicit criteria support learning because they provide students with a framework for metacognition and encourage learning-oriented dialogue about the desirable qualities of a task (Shepard, 2000). On the other, they restrict or normalize not only the knowledge gained and displayed by students, but also that which is recognized and valued by teachers (Crossouard, 2011; Morgan & Watson, 2002). This has made the use of rubrics controversial for teaching and assessing writing (e.g., Badger, 1999; Newkirk, 2004; Wilson, 2006). Several strategies, such as using exemplars and student-generated criteria, have been recommended by experienced teachers

to ensure that explicit criteria enhance fairness without constraining learning (Tierney, 2010). Empirical evidence regarding the effectiveness of these strategies for learning is emerging (Andrade, Du & Wang, 2008; see also Chapter 2, this volume), but has yet to be established in diverse contexts.

A second concern is that despite widespread agreement with the principle of transparency, teachers' assessment practices are still not entirely transparent. CA literature tends to focus narrowly on the explicitness of criteria, which does not necessarily bring the values and biases embedded in an assessment "out into the open" (Gipps, 1999, p. 385). Wyatt-Smith and colleagues (2010) found in their investigation of moderation meetings in Australia that teachers applied their own idiosyncratic values and biases in assessing student work even though task-specific descriptors with exemplars were provided. The addition of unofficial or unacknowledged criteria clouds the meaning of assessment results. Although this issue has long been identified in research on grading (See Chapter 15, this volume), teachers persist in mediating assessment criteria that should emanate from learning expectations or standards.

A third concern about transparency relates to the communication between students and teachers about assessment. The quality of assessment conversations may vary to a degree that threatens fairness. Some secondary teachers in one study, for example, reported having regular assessment conversations with students, while others seemed to distribute rubrics without explanation (Tierney et al., 2011). The effectiveness of merely handing out rubrics is questionable (Andrade, Du & Wang, 2008), and different degrees of transparency in communicating about assessment may affect other aspects of fairness, particularly opportunity to learn and demonstrate learning.

Based on these concerns, three questions are recommended for further research on transparency as an aspect of fairness. First, what is the comparative effectiveness of strategies that aim to help diverse students understand and use assessment criteria for learning? Second, how and why are explicit, mandated assessment criteria mediated by tacit values, personal characteristics, and contextual factors? And third, how do the

amount, frequency and clarity of information communicated by teachers affect students' understanding of an assessment task?

Differences in Students' Opportunities

Opportunity to learn and opportunity to demonstrate learning are often discussed in tandem because they are tightly related in practice. However, they are quite distinct phenomena. In the literature, opportunity to learn is discussed temporally, as a phenomenon in the past, present, or future lives of students. When students' backgrounds are considered, opportunity to learn is a social justice issue that involves political, social and economic forces beyond the classroom. As differences in students' opportunities are central to assessment reform, change has been sought in terms of principles (e.g., Gee, 2003; Tierney, 1998/2005) and practical intervention in schools (e.g., Weeden, Winter & Broadfoot, 2002). Formative assessment is sometimes seen as a means of equalizing disparate opportunities (e.g., Camilli, 2006; Pryor & Crossouard, 2008; Wormeli, 2006). In this light, assessment is more than a technical or moral process; it is a "form of activism" (Poehner, 2011, p.103). An example of research in this vein is Crossouard's (2011) study on the use of a formative assessment task in schools serving deprived areas in Scotland.

Most frequently, researchers, teachers and students discuss opportunity to learn in terms of alignment. Sometimes this is specifically as a matter of test preparation (Camilli, 2006; Eggen, 2004; Gordan & Fay, 2010; Green et al., 2007), and other times alignment is described more generally as the fit between what is taught and what is assessed (Drummond, 2003; Duffield & Spencer, 2002; Eggen, 2004; Gummer, 2000; McMillan, 2011; Speck, 1998; Suskie, 2002). For some students and teachers, assessment is fairer when it provides an immediate opportunity to learn, either through completing a task or from subsequent feedback (Robinson, 2002; Duffield & Spencer, 2002; Tierney, 2010). Concerns are also expressed about the long-term consequences of teachers' assessment decisions, particularly when they influence access to scholarships,

educational placements and future employment opportunities (Morgan & Watson, 2002; Tierney, 2010). Differences in the beliefs and expectations that teachers bring to their decision making can result in unevenly distributed opportunities for students within the same school or system (e.g. Nagy, 2000; Watson, 1999; Yung, 2001; Zoeckler, 2005).

The notion that students should have multiple opportunities to demonstrate learning generates little disagreement among teachers (e.g., Green et al., 2007). The rationale for providing students with multiple opportunities is based on the technical need for sufficient information, especially for high-stakes decisions. It is also recommended that opportunities to demonstrate learning be varied (Camilli, 2006; Gipps, 1999; Shepard, 2000; Suskie, 2002). The rationale for varied opportunities is based on the understanding that learning is not a uniform process, especially in the social environment of a classroom. The characteristics of assessment tasks can influence students' perceptions of fairness (Duffield & Spencer, 2002; Knoch & Elder, 2010; Mauldin, 2009), and interact with their demonstration of learning (Shepard, 2000; Tomlinson, 2005; Weeden et al., 2002). Circumstances that can limit opportunity to demonstrate learning include, for example, when expectations are uniformly low (Campbell, 2003; Wiggins, 1993), when student happiness and fun activities are valued over self-efficacy and substantive learning (Brighton, 2003), or when assessment tasks are designed without accurate information about student learning (Moon, 2005). A lack of opportunity to demonstrate learning threatens fairness in both ethical and technical terms; it devalues the learner (Buzzelli & Johnston, 2002), and it limits the opportunities that are subsequently provided (Lam, 1995).

Considering opportunity to learn and opportunity to demonstrate learning as aspects of fairness raises two questions for further investigation. First, how are teachers' assessment practices influenced by their beliefs about students' past opportunities and their expectations for students' future opportunities? Second, to what extent might formative assessment mitigate the effects of social inequity in different educational contexts? Taking multiple

perspectives into account, especially those of students and policy-makers, could reveal a greater role for CA research in the social justice agenda.

Equal or Equitable for All

Students in varied educational contexts indicate that equal treatment, or consistency in assessment, is highly valued in classrooms (Bursuck et al., 1999; Dalbert et al., 2007; Duffield & Spencer, 2002; Robinson, 2002; Smith & Gorard, 2006). Some teachers agree, and they steadfastly dismiss any accommodation for one student as being unfair to the other students, even if it might be productive for learning (e.g., Allen & Lambating, 2001; Yung, 2001; Zoeckler, 2005). Others are cautious about appearing to differentiate in assigning tasks or giving feedback because they are aware of students' perceptions (e.g., Brighton, 2003; Tierney, 2010; Torrance & Pryor, 2001). A common strategy, reported across contexts, is to go back over a graded set of papers to ensure that assessment criteria is evenly applied without the effect of mood, fatigue, or other rater characteristics (e.g., Eggen, 2004; Gynnild, 2011; Ryan, 2000). In essence, this strategy aims to eliminate construct irrelevant variance, which is central to the definition of fairness in the *Standards* (AERA et al., 1999) for testing, and in keeping with the sense of neutrality and balance underlying common usage in the English language. As Whittington (1999) noted, the imperative for consistency in testing is readily transferred by teachers to their assessment practices because it is congruent with "basic ideas of fairness" (p.17).

While belief in the need for consistency is strong, teachers' descriptions of their assessment practices reveal that exceptions are frequently made in the name of fairness. Teachers' decisions are often based on their understanding of particular students or situations, and some teachers articulate moral rationale for differentiating assessment tasks or criteria for individual students (e.g., Campbell, 2003; Tierney, 2010; Zoeckler, 2005). The idea that assessment is fairer when it is appropriate for individual learners emerges from three areas. First, special education guidelines (e.g., Guskey

& Jung, 2009) have increasingly been adopted in assessment policies aiming to accommodate identified groups, such as students with learning disabilities or English-language learners (e.g., Ontario Ministry of Education, 2010). Second, this idea emerges from theories that integrate teaching, learning, and assessing (e.g., Moon, 2005; Poehner, 2011; Tomlinson, 2005). Third, viewing cultural diversity as a wealth has led to a greater awareness not only that students bring varied knowledge and skills to classrooms, but also that this must be recognized in assessment practices (e.g., Cowie, in press; Jiménez, 2004/2005). Overall, the literature reveals that a shift is underway as convictions about equal treatment yield slowly to the realization that equitable practices result in fairer assessment, particularly in terms of supporting student learning.

The literature also illuminates considerable confusion in this change process. Teachers face a challenge in knowing when and how varying degrees of consistency and differentiation can enhance the fairness of an assessment, especially when it serves multiple purposes. At present, differences in teachers' judgments often seem to be based on students' group membership (e.g., Elhoweris et al., 2005; Miner, 1990; Peterson & Kennedy, 2006), either in officially recognized groups (e.g., gender, special needs), or groups constructed and perceived in varying degrees by teachers (e.g., hardworking, capable). Fairness is threatened when assessment decisions are influenced by stereotypes and personal values, but these are often so entrenched that they are overlooked. This threat is exacerbated when guidance for fair assessment is unclear about the difference between equality and equity, and the need for both consistency and differentiation.

Research is needed in this area on both theoretical and practical levels. Tomlinson (2005) highlights the commonalities in the principles for grading and differentiated instruction (DI) (see Chapter 23, this volume). Similar discussion is needed to delve further into the relationship between elements of AFL and aspects of fairness. Grounding theoretical discussion with practical examples, such as Mrs. Baca's reliable assessment (Parkes, 2007), would be helpful. Case studies with teachers who are successful in implementing

current principles, similar to the one by Ernest, Heckaman, Thompson, Hull and Carter (2011), would also provide insight for practice. How do successful teachers approach the tension between equality and equity in CA, and what knowledge do they use to assess the learning of diverse students?

Critical Reflection On the Go

All of the aspects of fairness discussed thus far suggest that fairness in CA requires thought. The importance of teachers' reflection is often emphasized in the literature, but discussion about the focus of this reflection is somewhat of a hodgepodge. Exactly what should teachers' reflection be about, and how should they go about it? One topic that clearly stands out for fair assessment is bias. Empirical research has raised concerns for decades about how teachers use, or don't use their knowledge or perceptions of students (e.g., from Miner, 1990 to Tierney et al., 2011). The literature that aims to guide teachers' practices is replete with recommendations for thoughtfulness about the biases and stereotypes that constrain fair practice (e.g., Joint Committee on Standards for Educational Evaluation, 2003; McMillan, 2011; Russell & Airasian, 2012). Popham (2008) identifies two types of bias: (1) offensiveness in the content of an assessment (i.e., negative stereotypes) and (2) unfair penalization that occurs when the content of an assessment negatively affects students' performance because of their group membership (i.e., gender, ethnicity). Popham's guidance for reflection on this aspect of fairness is particularly helpful because he provides explicit strategies for classroom teachers.

A second topic for teachers' reflection about fair assessment is purpose. A criterion-referenced framework is considered the most appropriate for the purpose of grading (Brookhart, 2004b). While teachers and educational institutions tend to identify their framework as being criterion-referenced, grading often involves informal norm-referencing and student-referencing (e.g., Markus, 2002; McMillan, 2001; Sadler, 2005). As several studies have suggested, confusion about assessment purposes threatens fairness (e.g.,

Delandshere, 2001; Yip & Cheung, 2005; Yung, 2001). Earl and Katz (2006) highlight three assessment purposes: (1) formative, (2) metacognitive and (3) summative. This document is also helpful because it provides questions and tips for teachers' reflection about these purposes. In brief, for fairness, teachers need to actively engage in critical reflection about the purposes of their assessments and about the biases that play into their assessment interactions and decisions.

Research on teachers' grading practices over the past few decades shows that while teachers want to be fair, they generally do not think of their practices in terms of principles or theory (Brookhart, 2004a; Dixon & Williams, 2003; McMillan, 2003; Tierney et al., 2011). However, some studies have found that teachers can provide clear descriptions and ethical rationale for their practices (Campbell, 2003; Tierney, 2010). This suggests that teachers' assessment decisions may be guided more strongly by practical knowledge and moral beliefs than theoretical knowledge. Additionally, there is indication that teachers' current vocabularies are insufficient both for the complexity of formative assessment (Crossouard, 2011) and for critical discussion of summative assessment practices (Black, Harrison, Hodgen, Marshall, Serreta, 2010). Sadler (1989) noted that thinking is the source and instrument for professional judgment; it is "not reducible to a formula" (p.124). However, without opportunities to learn critical assessment literacy, teachers may indeed resort to "formula-based methods" (Shepard, 2000, p.15) because they provide an illusion of objectivity, or they may unquestioningly perpetuate the practices that were practiced upon them.

While the implication for teacher education seems evident, several questions remain for research. First, to what degree do teacher education materials and programs support critical assessment literacy? Is CA presented as a collection of tools and techniques or as a process of equitable decision making? How is the development of professional judgment encouraged, and to what extent are teacher candidates prepared for the inevitable tensions in the use of assessment for multiple purposes? Researchers who have worked with teachers in thinking about assessment quality suggest that

assessment literacy needs to begin with the knowledge and values at the heart of teachers' working philosophies (Black et al., 2010; McMillan, 2003; Whittington, 1999). What types of knowledge and values are involved in teachers' thinking as they assess students' learning on the go in a classroom? For research in this area to advance further, it will be important to use methods that probe teachers' responses, such as narrative analysis and think aloud interviews. The studies by Wyatt-Smith and colleagues (2010) and Crisp (2010) on teachers' judgments in the context of moderation and exam marking provide excellent examples. Similar studies that focus on teachers' thinking processes in CA are needed.

Classroom Environment

With growing recognition of the social nature of learning, the importance of the classroom environment has repeatedly been discussed in assessment literature (Brookhart, 2004a; Shepard, 2000; Stiggins, Conklin & Bridgeford, 1986; Torrance & Pryor, 1998). However, ideas diverge about the nature of a classroom environment that encourages AFL. For example, Pryor and Crossouard (2008) suggest attending to students' personal narratives as a means of nurturing an affective environment, whereas *The Student Evaluation Standards* (Joint Committee on Standards for Educational Evaluation, 2003) recommend the use of balanced feedback (i.e., strengths and weaknesses) to develop a constructive environment. While different ideas may relate to underlying aims or epistemologies, it may also be that the ideal features of a classroom environment depend on the characteristics of the learners and learning involved. As such, the exact nature of the environment needed for assessment to be fair and support learning may depend on the broader context in which it is situated.

Two themes are quite clear, nonetheless, in looking at the relationship between the classroom environment and fair assessment. First is the role of power dynamics. The teachers' authority in classroom relationships has traditionally been recognized, particularly in summative assessment where teachers examine, judge, and grade student

achievement (Buzzelli & Johnston, 2002). While students are subjected to this process, they also have some agency in CA. As Bulterman-Bos and colleagues (2002) note, students “contribute actively to what can be observed” and they sometimes “even compel teachers to take notice” (p.1085). Involving students in assessment can empower students further (e.g., McDonald & Boud, 2003), but this does not necessarily disrupt traditional power dynamics (Brookfield, 2001; Schendel & O'Neill; 1999), particularly with young children who are not prepared to challenge teachers' decisions (Dann, 2002). Fairness cannot be assumed simply because an assessment aims to support learning.

A second theme is that the virtues of respect and trust are essential for fairness. Tertiary students associate fair assessment with respectful interactions between students and instructors (Gordan & Fay, 2010; Tata, 2005; Wendorf & Alexander, 2005). Trust and respect are also central to secondary students' willingness to disclose their knowledge and engage in AFL (Cowie, 2005). Some experienced teachers emphasize the importance of being proactive in the development of a constructive environment for assessment to be fair and support learning (Tierney, 2010). Many of their recommended strategies center on trust and respect (e.g., monitoring the tone of peer assessment) and take the open nature of classroom interaction into account. This suggests that the relationship between fair assessment and AFL is symbiotic. They are not the same, but they *can* be mutually supportive. The possibility is stressed because student involvement in assessment can also have a negative effect in terms of “poisoning” peer relations (Pryor & Lubisi, 2002, p.682). The classroom environment is in itself affected by all the other aspects of fairness. Brookhart (2003) notes that the outcomes of CA “fold back into and become part of the classroom environment” (p.7). Thus, without clear understanding of expectations and criteria, without opportunities to learn and demonstrate learning, without appropriate assessments and some degree of consistency, and without a reflective and responsive approach to interactions and decision making, the classroom environment is unlikely to be one in which desirable learning is achieved through

assessment.

Classroom environment research has clearly linked the characteristics of assessment tasks to students' self-efficacy and attitudes about learning (e.g., Dorman, Fisher & Waldrup, 2006). An association between the quality of AFL and the culture in classrooms and schools has also been found (e.g., Birenbaum, Kimron, & Shilton, 2011). The relationship between assessment and the culture or environment should be explored further in terms of fairness. How do aspects of fairness interact in different environments, and how does this interaction influence learning? To what extent do contextual factors, such as class size and system requirements, facilitate or constrain teachers' efforts to be fair in assessing student learning, and what are the subsequent effects on the classroom environment? A better understanding the characteristics of classroom environments that encourage fair assessment and support learning would be valuable for both theory and practice.

Closing Thoughts

There are multiple and conflicting interpretations of the concept of fairness in the literature on educational assessment. To some extent, perspectives on fairness are informed by position, meaning that what is thought to be fair depends on a person's role in a particular context. While test developers focus on eliminating differential item functioning, tertiary students hope for respectful communication about grading. Significant differences are seen between fairness in testing and CA. Opportunity to learn, for example, is clearly more salient in classrooms where the primary purpose of assessment is to support learning. Fairness is particularly complex in AFL because it is a quality of human interaction in perpetually evolving circumstances. A posteriori evidence of fairness determined through investigation or argument is helpful from a theoretical perspective, but it is insufficient for the immediacy of classroom practice, especially for the learners who experience the effects of an assessment. The complexity of fairness is further compounded by the blurring of purposes in CA, which varies from context to context. As a result,

fairness cannot be prescribed in universal terms for all students in all classrooms. Ongoing dialogue between stakeholders is absolutely critical to encourage the fairest CA possible. While ideas about what is fair in any given circumstance may often seem at odds, the underlying meaning is often the same. This provides much common ground for the conceptualization and practice of fair assessment in the educational community.

REFERENCES

- Allen, J. D., & Lambating, J. (2001, April). *Validity and reliability in assessment and grading: Perspectives of preservice and inservice teachers and teacher education professors*. Paper presented at the Annual Meeting of the American Educational Research Association Seattle, WA.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1985). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- American Psychological Association, American Educational Research Association, and the National Council on Measurement in Education. (1966). *Standards for educational and psychological tests*. Washington, DC: American Educational Research Association.
- American Psychological Association, American Educational Research Association, and the National Council on Measurement in Education. (1974). *Standards for educational and psychological tests*. Washington, DC: American Educational Research Association.
- Andrade, H. G. (2005). Teaching with rubrics: The good, the bad, and the ugly. *College Teaching* 53(1), 27-30.
- Andrade, H. L., Du, Y., & Wang, X. (2008). Putting rubrics to the test: The effect of a model, criteria generation, and rubric-referenced self-assessment on elementary school students' writing. *Educational Measurement: Issues and Practice*, 27(3), 3-13.
- Assessment Reform Group. (2002). *Assessment for learning: 10 principles*. Retrieved from <http://www.assessment-reform-group.org/publications.html>
- Badger, E. (1999). Finding one's voice: A model for more equitable assessment. In A. L. Nettles & M. T. Nettles (Eds.), *Measuring Up: Challenges minorities face in educational assessment* (pp. 53-69). Norwell, MA: Kluwer Academic Publishers.
- Birenbaum, M., Kimron, H., & Shilton, H. (2011). Nested contexts that shape assessment for learning: School-based professional learning community and classroom culture. *Studies in Educational Evaluation*, 37, 35-48.
- Black, P., Harrison, C., Hodgen, J., Marshall, B., & Serret, N. (2010). Validity in teachers' summative assessments. *Assessment in Education: Principles, Policy & Practice*, 17(2), 2215-2232.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7-74.
- Black, P., & Wiliam, D. (2006). The reliability of assessments. In J. Gardner (Ed.), *Assessment and learning* (pp.119-131). London: SAGE Publications.
- Bredo, E. (2006). Philosophies of educational research. In J. L. Green, G. Camilli & P. B. Elmore (Eds.), *Handbook of complementary methods in education research* (pp. 3-31). Mahwah, NJ: Lawrence Erlbaum Associates for American Educational Research Association.
- Brighton, C. M. (2003). The effects of middle school teachers' beliefs on classroom practices. *Journal for the Education of the Gifted*, 27(2/3), 177-206.
- Brookfield, S. (2001). Unmasking power: Foucault and adult learning. *Canadian Journal for Studies in Adult Education*, 15(1), 1-23.
- Brookhart, S. M. (1993). Teachers' grading practices: Meaning and values. *Journal of Educational Measurement*, 30(2), 123-142.
- Brookhart, S. M. (1994). Teachers' grading: Practice and theory. *Applied Measurement in Education*, 7(4), 279-301.
- Brookhart, S. M. (2003). Developing measurement theory for classroom assessment purposes and uses. *Educational Measurement: Issues and Practice*, 22(4), 5-12.
- Brookhart, S. M. (2004a). Classroom assessment: Tensions and intersections in theory and practice. *Teachers College Record*, 106(3), 429-458.
- Brookhart, S. M. (2004b). *Grading*. Upper Saddle River, NJ: Pearson Education, Inc.
- Bulterman-Bos, J., Terwel, J., Verloop, N., & Wardekker, W. (2002). Observation in teaching: Toward a practice of objectivity. *Teachers College Record*, 104(6), 1069-1100.
- Bursuck, W. D., Munk, D. D., & Olson, M. (1999). The fairness of report card grading adaptations: What do students with and without learning disabilities think? *Remedial and Special Education*, 20(2), 84-92.
- Buzzelli, C. A., & Johnston, B. (2002). *The moral dimensions of teaching: Language, power, and culture in classroom interaction*. New York: RoutledgeFalmer.

- Camilli, G. (2006). Test fairness. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 221-256). Westport, CT: Praeger Publishers.
- Campbell, E. (2003). *The ethical teacher*. Maidenhead, England: Open University Press.
- Childs, R. A. (1990). Gender bias and fairness. Washington, DC: ERIC Clearinghouse on Tests Measurement and Evaluation, American Institutes for Research. (ERIC Document Reproduction Service No. ED328610)
- Chory-Assad, R. M. (2002). Classroom justice: Perceptions of fairness as a predictor of student motivation, learning, and aggression. *Communication Quarterly*, 50(1), 58-77.
- Clarke, M. M., Madaus, G. F., Horn, C. L., & Ramos, M. A. (2000). Retrospective on educational testing and assessment in the 20th century. *Journal of Curriculum Studies*, 32(2), 159-181.
- Cole, N. S., & Zieky, M. J. (2001). The new faces of fairness. *Journal of Educational Measurement*, 38(4), 369-382.
- Cowie, B. (2005). Student commentary on classroom assessment in science: A sociocultural interpretation. *International Journal of Science Education*, 27(2), 199-214.
- Cowie, B. (in press). 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: The experiential and the virtual*. Dordrecht, The Netherlands: Springer.
- Crisp, V. (2010). Towards a model of the judgement processes involved in examination marking. *Oxford Review of Education*, 36(1), 1-21.
- Crossouard, B. (2011). Using formative assessment to support complex learning in conditions of social adversity. *Assessment in Education: Principles, Policy & Practice*, 18(1), 59-72.
- Dalbert, C., Schneidewind, U., & Saalbach, A. (2007). Justice judgments concerning grading in school. *Contemporary Educational Psychology*, 32(3), 420-433.
- Dann, R. (2002). *Promoting assessment as learning: Improving the learning process*. London: RoutledgeFalmer.
- Davies, A. (2010). Test fairness: A response. *Language Testing*, 27(2), 171-176.
- DeLain, M. T. (1995/2005). Equity and performance-based assessment: An insider's view. In S. J. Barrentine & S. M. Stokes (Eds.), *Reading assessment: Principles and practices for elementary teachers* (2nd ed., pp. 52-55). Newark, DE: International Reading Association.
- Delandshere, G. (2001). Implicit theories, unexamined assumptions and the status quo of educational assessment. *Assessment in Education: Principles, Policy & Practice*, 8(2), 113-133.
- Dixon, H., & Williams, R. (2003). Teachers' understanding and use of formative assessment in literacy learning. *New Zealand Annual Review of Education*. Retrieved from <http://assessment.tki.org.nz/Research/Research>
- Dorman, J. P., Fisher, D. L., & Waldrup, B. G. (2006). Classroom environment, students' perceptions of assessment, academic efficacy and attitude to science: A LISREL analysis. In D. L. Fisher & M. S. Khine (Eds.), *Contemporary approaches to research on learning environments: Worldviews* (pp.1-28). Hackensack, NJ: World Scientific.
- Drummond, M. J. (2003). *Assessing children's learning* (2nd ed.). London: David Fulton Publishers, Ltd.
- Duffield, K. E., & Spencer, J. A. (2002). A survey of medical students' views about the purposes and fairness of assessment. *Medical Education*, 36(9), 879-886.
- Earl, L. M. & Katz, S. (2006). *Rethinking classroom assessment with purpose in mind*. Governments of Alberta, British Columbia, Manitoba, Northwest Territories, Nunavut, Saskatchewan, and Yukon Territory: Western and Northern Canadian Protocol for Collaboration in Education.
- Ebel, R. L. (1972). *Essentials of educational measurement* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall, Inc.
- Eggen, A. B. (2004). *Alfa and omega in student assessment: Exploring identities of secondary school science teachers*. Universitetet i Oslo (Norway). Retrieved from www.ils.uio.no/forskning/pdh-drgrad/.../AstridEggenAvhandling1.pdf
- Elhoweris, H., Mutua, K., Alsheikh, N., & Holloway, P. (2005). Effect of children's ethnicity on teachers' referral and recommendation decisions in gifted and talented programs. *Remedial and Special Education*, 26(1), 25-31.
- Elwood, J. (2006). Formative assessment: Possibilities, boundaries and limitations. *Assessment in Education: Principles, Policy & Practice*, 13(2), 215-232.
- Ernest, J. M., Heckaman, K. A., Thompson, S. E., Hull, K. M., & Carter, S. W. (2011). Increasing the teaching efficacy of a beginning special education teacher using differentiated instruction: A case study. *International Journal of Special Education*, 26(1).
- Finklestein, I. E. (1913). *The marking system in theory and practice*. Baltimore, MD: Warwick &

- York, Inc.
- Frederiksen, J. R., & Collins, A. (1989). A system approach to educational testing. *Educational Researcher*, 18(9), 27-32.
- Gee, J. P. (2003). Opportunity to learn: A language-based perspective on assessment. *Assessment in Education: Principles, Policy & Practice*, 10(1), 27-46.
- Giordano, G. (2005). *How testing came to dominate American schools: The history of educational assessment*. New York: Peter Lang Publishing.
- Gipps, C. (1994). *Beyond testing: Towards a theory of educational assessment*. London: Falmer Press.
- Gipps, C. (1999). Socio-cultural aspects of assessment. *Review of Research in Education*, 24, 355-392.
- Gipps, C. (2005). Commentary on "the idea of testing: Psychometric and sociocultural perspectives". *Measurement*, 3(2), 98-102.
- Gordon, M. E., & Fay, C. H. (2010). The effects of grading and teaching practices on students' perceptions of grading fairness. *College Teaching*, 58(3), 93-98.
- Graham, P. (2005). Classroom-based assessment: Changing knowledge and practice through pre-service teacher education. *Teacher and Teacher Education*, 21, 607-621.
- Green, S., Johnson, R. L., Kim, D. H., & Pope, N. S. (2007). Ethics in classroom assessment practices: Issues and attitudes. *Teaching and Teacher Education*, 23, 999-1011.
- Gummer, E. S. (2000). Rhetoric and reality: Congruence between the knowledge and practice of assessment of the classroom science teacher and the reform initiative of the national science education standards. Purdue University (USA). *Proquest Digital Dissertations*. UMI No. AAT 3018203.
- Guskey, T. R. (2006, April). "It wasn't fair!" *Educators' recollections of their experiences as students with grading*. Paper presented at the American Educational Research Association, San Francisco.
- Guskey, T. R., & Jung, L. A. (2009). Grading and reporting in a standards-based environment: Implications for students with special needs. *Theory Into Practice*, 48(1), 53-62.
- Gynild, V. (2011). Student appeals of grades: A comparative study of university policies and practices. *Assessment in Education: Principles, Policy & Practice*, 18(1), 41-57.
- Horn, R. A. (2002). Reform in curriculum, instruction, and assessment, *Understanding educational reform: A reference handbook*. Santa Barbara, CA: ABC-CLIO, Inc.
- Jackson, P. W., Boostrom, R. E., & Hansen, D. T. (1993). *The moral life of schools*. San Francisco: Jossey-Bass/John Wiley & Sons, Inc.
- Jiménez, R. T. (2004/2005). More equitable literacy assessments for Latino students. In S.J. Barrentine & S.M. Stokes (Ed.), *Reading assessment: Principles and practices for elementary teachers* (pp. 49-51). Newark, DE: International Reading Association.
- Joint Advisory Committee. *Principles for fair student assessment practices for education in Canada*. (1993). Retrieved from www2.education.ualberta.ca/educ/psych/crame/files/eng_prin.pdf
- Joint Committee on Standards for Educational Evaluation. (2003). *The student evaluation standards: How to improve evaluations of students*. Thousand Oaks, CA: Corwin Press, Inc.
- Kane, M. (2010). Validity and fairness. *Language Testing*, 27(2), 177-182.
- Knoch, U., & Elder, C. (2010). Validity and fairness implications of varying time conditions on a diagnostic test of academic English writing proficiency. *System*, 38(1), 63-74.
- Koch, P. R., Steelman, L. C., Mulkey, L., & Catsambis, S. (2008). Naughty or nice?: Equity, gender and behavior. *Social Psychology of Education: An International Journal*, 11(4), 409-430.
- Kornhaber, M. L. (2004). Assessment, standards, and equity. In J. A. Banks (Ed.), *Handbook of research on multicultural education* (2nd ed., pp. 91-109). San Francisco, CA: John Wiley & Sons, Inc.
- Kunnan, A. J. (2010). Test fairness and Toulmin's argument structure. *Language Testing*, 27(2), 183-189.
- Lam, T. C. (1995). Fairness in performance assessment. Greensboro, NC: ERIC Clearing house on Counseling and Student Services =. (ERIC Document Reproduction Service No. ED391982.
- Lane, S., & Silver, E. A. (1999). Fairness and equity in measuring student learning using a mathematics performance assessment: Results from the quasar project. In A. L. Nettles & M. T. Nettles (Ed.), *Measuring up: Challenges minorities face in educational assessment* (pp. 97-120). Norwell, MA: Kluwer Academic Publishers.
- Linn, R. L., Baker, E. L., & Dunbar, S. B. (1991). *Complex, performance-based assessment: Expectations and validation criteria*. (CSE Technical Report 331). Los Angeles, CA: Center for the Study of Evaluation, UCLA.
- Markus, J. (2002). Student assessment and evaluation in studio art. University of Toronto (Canada).

- Proquest Digital Dissertations. UMI No. AAT NQ69131.
- Mauldin, R. K. (2009). Gendered perceptions of learning and fairness when choice between exam types is offered. *Active Learning in Higher Education*, 10(3), 253-264.
- McDonald, B., & Boud, D. (2003). The impact of self-assessment on achievement: The effects of self-assessment training on performance in external examinations. *Assessment in Education: Principles, Policy & Practice*, 10(2), 209-220.
- McMillan, J. H. (2001). Secondary teachers' classroom assessment and grading practices. *Educational Measurement: Issues and Practice*, 20(1), 20-32.
- McMillan, J. H. (2003). Understanding and improving teachers' classroom assessment decision making: Implications for theory and practice. *Educational Measurement: Issues and Practice*, 22(4), 34-43.
- McMillan, J. H. (2011). *Classroom assessment: Principles and practice for effective standards-based instruction* (5th ed.). Boston, MA: Pearson Education Inc.
- Messick, S. (1965). Personality measurement and the ethics of assessment. *American Psychologist*, 20(2), 136-142.
- Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher*, 18(2), 5-11.
- Messick, S. (2000). Consequences of test interpretation and use: The fusion of validity and values in psychological assessment. In R. D. Goffin & E. Helmes (Eds.), *Problems and solutions in human assessment* (pp. 3-20). Norwell, MA: Kluwer Academic Publishers
- Messick, S., & Anderson, S. (1970). Educational testing, individual development, and social responsibility. *The Counseling Psychologist*, 2(2), 80-88.
- Miner, S. (1990). Teacher evaluations of case descriptions of LD gifted children. *Gifted Child Quarterly*, 34(1), 37-39.
- Moon, T. (2005). The role of assessment in differentiation. *Theory into Practice*, 44(3), 226-233.
- Morgan, C., & Watson, A. (2002). The interpretive nature of teachers' assessment of students' mathematics. *Journal for Research in Mathematics Education*, 33(2), 78-110.
- Moss, P. (1994). Can there be validity without reliability. *Educational Researcher*, 23(2), 5-12.
- Moss, P. (1998). The role of consequences in validity theory. *Educational Measurement: Issues and Practice*, 17(2), 6-12.
- Moss, P. (2003). Reconceptualizing validity for classroom assessment. *Educational Measurement: Issues and Practice*, 22(4), 13-25.
- Moss, P., Pullin, D., Gee, J. P., & Haertel, E. H. (2005). The idea of testing: Psychometric and sociocultural perspectives. *Measurement*, 3(2), 63-83.
- Nagy, P. (2000). Three roles of assessment: Gatekeeping, accountability, and instructional diagnosis. *Canadian Journal of Education*, 25(4), 262-279.
- National Research Council. (2010). *State assessment systems: Exploring best practices and innovations: Summary of two workshops*. Washington, DC: The National Academies Press.
- Nettles, A. L., & Nettles, M. T. (1999). *Measuring up: Challenges minorities face in educational assessment*. Norwell, MA: Kluwer Academic Publishers.
- Newkirk, T. (2004). A mania for rubrics. In A. S. Canestrari & B. A. Marlowe (Eds.), *Educational foundations: An anthology of critical readings* (pp.199-201). Thousand Oaks, CA: SAGE Publications.
- Ontario Ministry of Education. (2010). *Growing success: Assessment, evaluation and reporting in Ontario schools* (1st ed). Toronto, ON: Queen's Printer for Ontario.
- Parkes, J. (2007). Reliability as argument. *Educational Measurement: Issues and Practice*, 26(4), 2-10.
- Peterson, S., & Kennedy, K. (2006). Sixth-grade teachers' written comments on student writing. *Written Communication*, 23(1), 36-62.
- Phillips, M. (2002). Is 'same' treatment 'fair' treatment? In J. Shulman, A. K. Whittaker & M. Lew (Eds.), *Using assessments to teach for understanding* (pp.15-17). New York: Teachers' College Press.
- Poehner, M. E. (2011). Dynamic assessment: Fairness through the prism of mediation. *Assessment in Education: Principles, Policy & Practice*, 18(2), 99-112.
- Popham, W. J. (1981). *Modern educational measurement*. Englewood Cliffs, NJ: Prentice-Hall, Inc.
- Popham, W. J. (2008). *Classroom assessment: What teachers need to know* (5th ed.). Boston, MA: Pearson Education, Inc.
- Pryor, J., & Crossouard, B. (2008). A socio-cultural theorisation of formative assessment. *Oxford Review of Education*, 34(1), 1-20.
- Pryor, J., & Lubisi, C. (2002). Reconceptualising educational assessment in South Africa — testing times for teachers. *International Journal of Educational Development*, 22(6), 673-686.
- Resh, N. (2009). Justice in grades allocation: Teachers' perspective. *Social Psychology of Education: An International Journal*, 12(3), 315-325.

- Resh, N., & Dalbert, C. (2007). Gender differences in sense of justice about grades: A comparative study of high school students in Israel and Germany. *Teachers College Record*, 109(2), 322-342.
- Robinson, J. M. (2002). In search of fairness: An application of multi-reviewer anonymous peer review in a large class. *Journal of Further & Higher Education*, 26(2), 183-192.
- Russell, M. K., & Airasian, P. W. (2012). *Classroom assessment: Concepts and applications* (7th ed.). New York: McGraw-Hill Higher Education.
- Ryan, T. G. (2000). An action research study of secondary science praxis. University of Toronto (Canada). *Proquest Digital Dissertations*. UMI No. AAT NQ58601.
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems, *Instructional Science*, 18, 119-144.
- Sadler, D. R. (2005). Interpretations of criteria-based assessment and grading in higher education. *Assessment & Evaluation in Higher Education*, 30(2), 175-194.
- Schendel, E., & O'Neill, P. (1999). Exploring the theories and consequences of self-assessment through ethical inquiry. *Assessing Writing*, 6(2), 199-227.
- Shepard, L. (1997). The centrality of test use and consequences for test validity. *Educational Measurement: Issues and Practice*, 16(2), 5-8,13, 24.
- Shepard, L. (2000). *The role of classroom assessment in teaching and learning*. CSE Technical Report 517. Washington, DC: Office of Educational Research and Improvement.
- Shepard, L. (2006). Classroom assessment. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 623-646). Westport, CT: American Council on Education/Praeger Publishers.
- Shulman, L. S. (1987/2004). The wisdom of practice: Managing complexity in medicine and teaching. In S. M. Wilson (Ed.), *The wisdom of practice: Essays on teaching, learning, and learning to teach* (pp. 249-271). San Francisco, CA: Jossey-Bass.
- Smith, E., & Gorard, S. (2006). Pupils' views on equity in schools. *Compare*, 36(1), 41-56.
- Smith, J. K. (2003). Reconsidering reliability in classroom assessment and grading. *Educational Measurement: Issues and Practice*, 22(4), 26-33.
- Speck, B. W. (1998). Unveiling some of the mystery of professional judgment in classroom assessment. In R. S. Anderson & B. W. Speck (Eds.), *Classroom assessment and the new learning paradigm* (pp. 89-96). San Francisco, CA: Jossey-Bass/John Wiley & Sons, Inc.
- Stiggins, R. J., & Bridgeford, N. J. (1985). The ecology of classroom assessment. *Journal of Educational Measurement*, 22(4), 271-286.
- Stiggins, R. J., Conklin, N. F., & Bridgeford, N. J. (1986). Classroom assessment: A key to effective education. *Educational Measurement: Issues and Practice*, Summer, 5-17.
- Stiggins, R. J., Frisbie, D. A., & Griswold, P. A. (1989). Inside high school grading practices: Building a research agenda. *Educational Measurement: Issues and Practice*, 8(2), 5-14.
- Stobart, G. (2005). Fairness in multicultural assessment systems. *Assessment in Education: Principles, Policy & Practice*, 12 (3), 275-287.
- Stobart, G. (2006). The validity of formative assessment. In J. Gardner (Ed.), *Assessment and learning* (pp.133-146). London: SAGE Publications.
- Suskie, L. (2002). Fair assessment practices: Giving students equitable opportunities to demonstrate learning. *Adventures in Assessment*. Retrieved from <http://www.sabes.org/resources/publications/adventures/vol14/14suskie.htm>
- Szpyrka, D. A. (2001). Exploration of instruction, assessment, and equity in the middle school science classroom. University of Central Florida (USA). *Proquest Digital Dissertations*. UMI No. AAT 3029061.
- Tata, J. (2005). The influence of national culture on the perceived fairness of grading procedures: A comparison of the United States and China. *Journal of Psychology*, 139(5), 401-412.
- Tierney, R. D. (2010). Insights into fairness in classroom assessment: Experienced English teachers share their practical wisdom. University of Ottawa (Canada). *Proquest Digital Dissertations*. UMI No. AAT NR69109.
- Tierney, R. D., Simon, M., & Charland, J. (2011). Being fair: Teachers' interpretations of principles for standards-based grading. *The Educational Forum*, 75(3), 210- 227.
- Tierney, R. J. (1998/2005). Literacy assessment reform: Shifting beliefs, principled possibilities, and emerging practices. . In S. J. Barrentine & S. M. Stokes (Ed.), *Reading assessment: Principles and practices for elementary teachers* (pp. 29-40). Newark, DE: International Reading Association.
- Tomlinson, C. A. (2005). Grading and differentiation: Paradox or good practice? *Theory Into Practice*, 44(3), 262-269.
- Torrance, H., & Pryor, J. (1998). *Investigating formative assessment: Teaching, learning and*

- assessment in the classroom*. Philadelphia, PA: Open University Press.
- Torrance, H., & Pryor, J. (2001). Developing formative assessment in the classroom: Using action research to explore and modify theory. *British Educational Research Journal*, 27(5), 615-631.
- Volante, L. (2006). Reducing bias in classroom assessment and evaluation. *Orbit*, 36, 34-36
- Watson, A. (1999). Paradigmatic conflicts in informal mathematics assessment as sources of social inequity. *Educational Review*, 51(2), 105-115.
- Weeden, P., Winter, J., & Broadfoot, P. (2002). *Assessment: What's in it for schools?* New York: Routledge Falmer.
- Wendorf, C. A., & Alexander, S. (2005). The influence of individual- and class-level fairness-related perceptions on student satisfaction. *Contemporary Educational Psychology*, 30(2), 190-206.
- Whittington, D. (1999). Making room for values and fairness: Teaching reliability and validity in the classroom context. *Educational Measurement: Issues and Practice*, 18 (1), 14-22.
- Wiggins, G. (1993). *Assessing student performance: Exploring the purpose and limits of testing*. San Francisco, CA: Jossey-Bass, Inc.
- Wiliam, D. (1994). *Toward a philosophy for educational assessment*. Paper presented at the British Educational Research Association, Oxford, England.
- Wilson, M. (2006). *Rethinking rubrics in writing assessment*. Portsmouth, NH: Heinemann, Reed Elsevier, Inc.
- Wormeli, R. (2006). *Fair isn't always equal: Assessing and grading in the differentiated classroom*. Portland, MN: National Middle School Association.
- Wyatt-Smith, C., Klenowski, V., & Gunn, S. (2010). The centrality of teachers' judgement practice in assessment: A study of standards in moderation. *Assessment in Education: Principles, Policy & Practice*, 17(1), 59-75.
- Xi. (2010). How do we go about investigating test fairness? *Language Testing*, 27(2), 147-170.
- Yip, D. Y., & Cheung, D. (2005). Teachers' concerns on school-based assessment of practical work. *Journal of Biological Education*, 39 (4), 156-162.
- Yung, B. H. W. (2001). Three views of fairness in a school-based assessment scheme of practical work in biology. *International Journal of Science Education*, 23 (10), 985-1005.
- Zoeckler, L. G. (2005). Moral dimensions of grading in high school English University of Indiana (USA). *Proquest Digital Dissertations*. UMI No. AAT3183500.