



Rethinking Traditional Assessment Concepts in Classroom-Based Assessment

As teachers and students have taken more active roles in assessment practices, our field has begun to pay more attention to classroom-based assessment (CBA). As a result, we have gained a better understanding of CBA principles but we have also become aware of challenges, particularly in applying traditional assessment concepts such as reliability and validity to the classroom context. In this article I continue the argument that viewing assessment through a sociocultural perspective will help us broaden our understanding of classroom learning environments and rethink traditional assessment concepts for CBA.

Introduction

Classroom-based assessment can be broadly defined to include all activities that teachers and students undertake to gather information for the purpose of teaching and learning. The increased role of the teacher and student in assessment practices has placed a greater focus on the action of classroom-based assessment compared to traditional assessment scholarship, which tended to focus more on large-scale standardized tests. The terms *classroom-based assessment*, *teacher-based assessment*, *school-based assessment*, *formative assessment*, and *alternative assessment* are often used interchangeably to refer to the same classroom-based practices and procedures. They highlight different aspects of the assessment process but all signify a more teacher-mediated, student-involved, context-based, and learning-focused assessment practice in contrast to large-scale formal assessments that are summative in nature and are used for accountability purposes. In this article, the term *classroom-based assessment* will be used to signify all of the practices and procedures that are included in the above terms. The practices and procedures listed below distinguish classroom-based assessment from other forms of traditional assessment. They are more than just who is doing the assessing but also for what, where, how, and why the assessment takes place.

1. [The] teacher is involved from beginning to end in assessment activities.
2. A number of samples of student work are collected over time.

3. The assessment can be modified and adapted by the teacher to match the teaching and learning goals of the class being assessed.
4. Assessments are carried out in classrooms by the teacher.
5. Assessments involve students more actively in the assessment process.
6. Assessments allow for immediate feedback to the students.
7. Assessments stimulate continuous evaluation of teaching and learning.
8. Assessments complement other forms of assessments, including formal/external ones. (Adapted from Leung & Davison, 2009)

In this sense, classroom-based assessment shares many of the characteristics with *assessment for learning*, a term distinct from *assessment of learning*, which is used for the purposes of grading and reporting progress. Assessment *for learning* (AfL) in the past 10 years has gained considerable recognition in education-assessment literature (Assessment Reform Group, 2002; Black & Wiliam, 1998; Broadfoot & Black, 2004; Leung, 2004, 2005; Leung & Rea-Dickins, 2007; Rea-Dickins 2001, 2006). A core principle of AfL is as follows:

Assessment for learning should be recognized as central to classroom practice—[m]uch of what teachers and learners do in classrooms can be described as assessment. That is, tasks and questions prompt learners to demonstrate their knowledge, understanding and skills. What learners say and do is then observed and interpreted, and judgments are made about how learning can be improved. These assessment processes are an essential part of everyday classroom practice and involve both teachers and learners in reflection, dialogue and decision making. (Assessment Reform Group, 2002)

Broadly speaking then, AfL is committed to improving student learning through assessment activities that use the students' current knowledge and ability and teacher intervention through interactive feedback. Based on the above descriptions, the goal of assessment for learning or classroom-based assessments is that it improves learning because:

- It is embedded in a view of teaching and learning;
- The learning goals are shared with students;
- It helps students know and recognize the goals they are aiming for;
- The students are involved in self-assessment;
- It provides feedback and helps students recognize their next step;
- It is underpinned by confidence that every student can improve; and
- Both teacher and student review and reflect on assessment data. (The Assessment Reform Group, n.d., p. 7)

While we have gained much knowledge about the purpose of assessment practices in the classroom, we have also become aware of challenges in applying traditional concepts of assessment to classroom-based assessment (CBA).

Conventional assessments such as large-scale formal assessments, or what we traditionally call *tests*, used for the assessment of learning are very different from CBA. This difference is more than merely the way the assessment occurs or the purpose for the assessments; the difference is also in how CBA is conceptually viewed in the conventional assessment field. This is mainly because of the paradox of CBA—that traditional testing theory construes CBA's inherent strengths as its greatest weaknesses. In traditional assessments, the context is regarded as irrelevant and the assessor is someone who must remain objective and uninvolved throughout the process of assessment. In contrast, CBA derives much of its reliability and validity from the classroom, where assessment activities are part of the curriculum, and from the teacher, who is familiar with the students' work and has a stake in their improvement. This theorization of CBA is obviously very different from that associated with traditional testing, in which the primary goal is to make generalizable explanations or predictions based on learning outcomes. As such, one can even say that the traditional concepts of assessment associated with the psychometric tradition of testing are seen as a potential threat to the development of CBA, which as described above is highly contextualized and dialogic in practice. To exploit the features of CBA effectively, we then need to rethink concepts such as reliability and validity for CBA. To help us in this regard, perhaps we need to look at other theoretical resources to understand learning/teaching.

A Sociocultural Perspective

Traditionally, second-language learning theories have viewed learning as something that takes place inside the head of the learner. However, from a sociocultural perspective, learning is perceived in the evolving relationships between learners and other participants, tools (symbolic and material), and the settings in which they conduct their activities (Lantolf, 2000, 2006; Lave & Wenger, 1991; Mehan, 1998; Wertsch, 1998). From this perspective, learners do not simply receive, internalize, and construct knowledge in their minds but enact it as persons-in-the-world participating in the practices of a particular community (Wenger, 1998). This understanding has important implications for the classroom, including our assessment practices. Perceiving the learning process as a socially constructed activity embedded in the local context means teachers and students alike are recognized as meaningful assessment partners (Leung, 2004; Lynch, 2001; Lynch & Shaw, 2005; McNamara & Roever, 2006).

To understand this perspective, we can juxtapose some traditional concepts of assessment with CBA practices in a college-level ESL academic-writing class. The class is an advanced writing class that prepares students to write in their disciplines. In this class, students read academic texts around a number of topics and through writing are expected to synthesize ideas in response to a prompt. Students also engage in writing research papers both collaboratively and individually on topics of their choosing. As the teacher, I conference with students to provide feedback on their writing and students also provide feedback by reading each other's essays.

“Assessment Is a Distinct Activity”

Traditionally assessment is viewed as a distinct activity and if we think of assessment as an activity that is different from teaching and learning, it affects how we design the learning environment. Sometimes we consider an “assessment” as a distinct activity such as an end-of-term test or final paper, but in a learning environment to make assessment a separate activity is artificial. Of course assessments as distinct activities can be used to provide feedback on learning, inform instructional decisions, and hold students accountable, but as activities that are part of the learning environment, assessments also need to afford students with opportunities to engage in the practices, in this case those of academic writers. As such, these activities have function and meaning within, but more important, outside the ESL writing class. For example, some activities generate the need for clarifying information in the reading texts, which then become questions for class discussion. In addition, the feedback and conclusions that students get from their individual essays and group research paper provide resources for further learning and teaching. From a sociocultural perspective, “building these complex social relationships around meaningful activities” (Wenger, 1998, p. 272) becomes an activity all participants assume. Furthermore, as Wenger points out, these activities engage the classroom community in practices that have impact beyond the boundary of the class. Thus, CBA becomes not so much a distinct set of activities, but rather assessment activities that provide evidence of students’ learning, that is, “practice” as writers/researchers.

“Assessment Is Context Irrelevant”

Typically, assessments are used to make interpretations about individuals or groups of individuals. To be comparable, contexts must be standardized to make the assessment experience the same for each student and are considered irrelevant in making inferences about the individual. However, in a classroom, the learning environment plays a part in the students’ experience; students make meaning through the various activities that occur in the learning context and thereby the context may provide different learning experiences for each student. Traditional assessments assume generalizability of test scores but as Messick (1989) notes, “It is important to recognize that the intrusion of context raises issues more of generalizability than of interpretive validity” (p. 14). Thus the context is viewed as a threat to test-score reliability in traditional assessments. However, the social context plays an important role in classroom assessments, that is, the classroom shapes the essay grade and in turn the interpretations. For example, students’ experience of a writing assignment can be varied, complex, and sometimes even unique. Student performances may be shaped by the writing prompt, which all students see, but they could also be shaped by other aspects such as choices students make (e.g., what to focus on in their essay, the class readings they use, by the sources students find on their own or in the group research paper, and by their ongoing interactions). A change in any of these factors may affect the nature of the “assessment” for the whole class, a group of students, or even one student. So interpreting a student’s performance

also requires understanding the context it was produced in and the factors that shaped it. In keeping with a sociocultural perspective, it is more fitting in CBA, then, that assessment should include the individual *and* the social situation or interaction of the classroom (Mehan, 1998; Wertsch, Del Rio, & Alvarez, 1995).

“Validity Is a Test Score–Based Interpretation and Use”

Usually, validity refers to an interpretation or use based on a test score. Messick (1989) defines validity as “an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores or other modes of assessment” (p. 13). From this definition we can say that a validity argument is made on an interpretation or action that is based on a test score. While this focus is useful (from a traditional assessment perspective), it does not capture what teachers do. Most classroom teachers need to make decisions moment-to-moment, lesson-to-lesson, class-to-class to help students learn, and teachers also need to be able to study the effects of their own decisions. As such, evidence supporting an interpretation decision draws from multiple, varied sources about students’ learning and the supporting environment, rarely from a single “test.” For example, when designing a class, I do not think of any one assessment and the particular interpretation I hope to draw from it but in terms of all the activities in the course. Using a single assessment activity to make an interpretation is limiting as it does not serve what I need to do in the writing class. I need to know how the assessment fits in with other assessment practices. While writing activities provide opportunities for both me and students to assess progress, each activity also serves to make interpretations about student learning and the effects of my decisions. In other words, these interpretations inform next steps for them and for me. This may happen differently for different assignments. For example, for the group research paper, I give written feedback on areas in need of work. Obviously, the feedback is not the same as it is in response to the specifics of each paper. For the individual essays, I conference with the students one-on-one to help draw out what they want to accomplish, discuss issues in writing they need/want to address, and/or suggest ways to manage the writing task. Students also peer review assignments to give one another feedback, so that they learn from each other’s writing. In a situation in which time is limited, I might resort to an assessment-based interpretation such as using an in-class, mid-/end-of-term writing assignment, but in the classroom evidence is always available and it would be somewhat unproductive as well as pedagogically unsound to rely only on a single assessment-based interpretation. So, while teachers may make a particular interpretation based on a test score, interpretations must be made also of all the activities that engage students’ learning in the classroom.

“Impact or Consequences Are an Aspect of a Test”

The relationship between tests and impact or consequences is somewhat unclear within the field of assessment. Some (AERA, APA, & NCME, 1999; Messick, 1989) see test consequences narrowly in that a sound test (absent of

any effects) has very little impact, while others (Cronbach, 1988; Moss, 1998; Shepard, 1993, 1997) see consequences broadly when a test serves the intended purpose (which can have both positive/negative consequences). Both these explanations identify impact or consequences as derived from the test/assessment itself. However, classroom-based assessment—in which every activity is geared toward improving learning—is all about consequences. If our interpretations of classroom activities inform us of students' learning, and those interpretations in turn entail our instructional decisions, then the effects of all these interpretations become evidence for our validity argument. Thus, from a sociocultural understanding of learning the nature of consequences needs a wider consideration than is typically given in traditional assessment practice. Typically, the success of a class is measured by the extent to which students are able to display what they know and can do—the construct, the traditional focus of assessment—but it also includes how students are situated, that is, how students relate to knowledge and to others in their social situation (Wenger, 1998). This understanding has major implications for classroom activities and the development of assessments. First, it means that reflecting on classroom practices is important. For example, when I study the effects of my teaching (which include assessment practices), I am not only looking for evidence about my students' knowledge acquisition, but also for the kind of people I am helping them to become. What shapes their sense of being a writer or a researcher or simply a learner?

Second, like most teachers, I do not want my class to be about grades because it will not help my students beyond the classroom. If we did not have grades (or in my case, scores for essays), we could focus on improving the learning. Some students may like this practice (because it allows risk taking without worry about the grade), but some may not (needing grades for motivation). To make our classroom-based assessment have a positive impact, we should try to encourage other kinds of responsibility that are more in keeping with the practices of the classroom and practices as writers/researchers. For example, writing about topics that are meaningful to students and that have a purpose beyond the writing assignment, and by creating collaborative/participatory arrangements (e.g., the group research paper) so that students are accountable to one another, will, one hopes, provide the impetus students need to help them as they leave the classroom and join their own professional communities. Therefore, *all* classroom assessment activities have an impact as they have meaning and are useful in the context of the classroom as well as in the relevant communities to which students will go. Therefore, it is the evidence of short-term, long-term, and cumulative impact or consequences of the activities on which validity primarily rests in classroom-based assessment.

Conclusion

The discussion and description above illustrate how classroom-based assessment involves the teacher, student, and the context in which learning occurs. As a result, it becomes challenging to impose traditional concepts of assessment—reliability and validity—in the same way as the purpose of class-

room assessment practices is not to generalize performance. Instead the goal of classroom-based assessment is to improve student learning. In fact, if teachers revise their practices to more faithfully enact the traditional concepts of reliability and validity, teachers would be compromising the environment they have planned as a resource for students' learning. Assessment activities are not separate activities but are viewed as an integral part of the learning context, which gives students a sense of learner agency. Making inferences of students' performance must also take into consideration the context in which the assessment activities occur, because students draw on their individual experiences as well as on the collective expertise of the class to develop knowledge through participation and apprenticeship. In this sense there is greater reliability in decisions that are made when they are based on many pieces of evidence and not just discrete assessment activities. Similarly, a validity argument for CBA is more than just interpreting an assigned score or grade from an assessment instrument. It requires both teachers and students jointly making interpretations about learning and understanding the effects of decisions based on those interpretations. Finally, the impact of assessments is not limited to discrete ones but encompasses all assessment practices in the classroom. So in that sense all assessment activities in the classroom have consequential validity. Based on this need to broaden or rethink the traditional concepts of reliability and validity for CBA, a more useful approach is to take a sociocultural perspective that views assessment practices as more interpretive, that is, understanding "meaning in context" (Moss, 1996, p. 21).

The ideas presented here are not new. Research acknowledges that traditional concepts of assessment privilege traditional forms of assessment and research also points to their limitations for less formalized assessment, which includes classroom-based assessment (Leung & Davison, 2009; Lynch, 2001; Moss, 2003; Taylor & Nolen, 1996). Even within the field of traditional assessment, many scholars of classroom-based assessment suggest using interpretive approaches in less formal assessments and qualitative research in general to comprehend assessment in specific contexts. For instance, Shepard (2001) argues, "Evaluating open ended tasks and drawing valid inferences from both formal and informal data sources requires new methods of data analysis and interpretation" (p. 1088).

The first step toward further study of CBA may be to complement "quantitative measures with richer qualitative studies of processes and interactions within the classroom" (Black & Wiliam 1998, p. 44). Perhaps another approach is to develop a theory of assessment that works effectively within the CBA framework (Leung & Davison, 2009). Where traditional assessment focuses on generalizable learning outcomes, CBA assessment may have to focus on an assessment standard based on that particular context of learning. Simply put, the focus needs to move from product to process. Till then, scholars need to continue study of CBA principles and practices. This will lead to even closer examination and better understanding of applying traditional assessment concepts and to further exploration of the interrelationship between learning and assessment.

Author

Priyanvada Abeywickrama is an assistant professor at San Francisco State University, where she teaches in the MA TESOL Program. Her research interest is in language assessment, specifically examining issues of validity. This paper was originally given as the plenary talk at the Southern California Association of Language Assessment Research (SCALAR) conference at UCLA in 2010.

References

- AERA, APA, & NCME (American Educational Research Association, American Psychological Association, & National Council of Measurement in Education). (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Assessment Reform Group. (2002). *Assessment for learning: 10 principles*. Retrieved April 3, 2010, from <http://www.aiaa.org.uk/afl/assessment-reform-group>
- Assessment Reform Group. (n.d). *Assessment for learning: Beyond the black box*. Cambridge, England: University of Cambridge School of Education. Retrieved April 3, 2010, from <http://www.aiaa.org.uk/afl/assessment-reform-group>
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy and Practice*, 5, 7–74.
- Broadfoot, P., & Black, P. (2004). Redefining assessment? The first ten years of Assessment in Education. *Assessment in Education*, 11, 7–27.
- Cronbach, L. J. (1988). Five perspectives on validity argument. In H. Wainer (Ed.), *Test validity* (pp. 3-17). Hillsdale, NJ: Lawrence Erlbaum.
- Lantolf, J. P. (2000). Second language learning as a mediated process. *Language Teaching*, 33, 79–96.
- Lantolf, J. P. (2006). Sociocultural theory and L2: State of the art. *Studies in Second Language Acquisition*, 28, 67-109.
- Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge, England: Cambridge University Press.
- Leung, C. (2005). Classroom teacher-based assessment of second language development: Construct as practice. In E. Hinkel (Ed.), *Handbook of research in second language learning and teaching* (pp. 869–888). Mahwah, NJ: Erlbaum.
- Leung, C. (2004). Developing formative teacher-based assessment: Knowledge, practice, and change. *Language Assessment Quarterly*, 1(1), 19–41.
- Leung, C., and Davison, C. (2009). Current issues in English language teacher-based assessment. *TESOL Quarterly*, 43(3), 393-416.
- Leung, C., & Rea-Dickins, P. (2007). Teacher assessment as policy instrument: Contradictions and capacities. *Language Assessment Quarterly*, 4(1), 6–36.
- Lynch, B. (2001). Rethinking assessment from a critical perspective. *Language Testing*, 18, 351–372.
- Lynch, B., and Shaw, P. (2005). Portfolios, power, and ethics. *TESOL Quarterly*, 39(2), 263–297.

- McNamara, T., and Roever, C. (2006). *Language testing: The social dimension*. Oxford, England: Blackwell.
- Mehan, H. (1998). The study of social interaction in educational settings: Accomplishments and unresolved issues. *Human Development*, 41, 245-269.
- Messick, S. (1989). Validity. *Educational measurement* (3rd ed., pp. 13-103). Washington, DC: The American Council on Education and the National Council on Measurement in Education.
- Moss, P. A. (1996). Enlarging the dialogue in educational measurement: Voices from interpretive research traditions. *Educational Researcher*, 25(1), 20-28, 43.
- Moss, P. A. (1998). The role of consequences in validity theory. *Educational Measurement: Issues and Practice*, 17(2), 5-12.
- Moss, P. A. (2003) Reconceptualizing validity for classroom assessment. *Educational Measurement: Issues and Practice*, 22(4), 13-25.
- Rea-Dickins, P. (2001). Mirror, mirror on the wall: Identifying processes of classroom assessment. *Language Testing*, 18, 429-462.
- Rea-Dickins, P. (2006). Currents and eddies in the discourse of assessment: A learning-focused interpretation. *International Journal of Applied Linguistics*, 16, 164-189.
- Shepard, L. A. (1993). Evaluating test validity. *Review of Research in Education*, 19, 405-450.
- Shepard, L. A. (1997). The centrality of test use and consequences for test validity. *Educational Measurement: Issues and Practice*, 16(2), 5-8.
- Shepard, L. A. (2001). The role of classroom assessment in teaching and learning. In V. Richardson (Ed.), *Handbook of research on teaching* (4th ed., pp. 1066-1101). Washington, DC: American Educational Research Association.
- Taylor, C. S., & Nolen, S. B. (1996). What does the psychometrician's classroom look like? Reframing assessment concepts in the context of learning. *Education Policy Analysis Archives*, 4(17), 1-39. Retrieved February 11, 2012, from <http://epaa.asu.edu/ojs/issue/view/4>
- Wenger, E. (1998). *Communities of practice: Learning, meaning, and identity*. Cambridge, England: Cambridge University Press.
- Wertsch, J. V. (1998). *Mind as action*. Oxford, England: Oxford University Press.
- Wertsch, J. V., Del Rio, P., & Alvarez, A. (1995). Sociocultural studies: History, action, and mediation. In J. V. Wertsch, P. Del Rio, & A. Alvarez (Eds.), *Sociocultural studies of mind* (pp. 1-36). Cambridge, England: Cambridge University Press.