CHAPTER 33

CLASSROOM-BASED ASSESSMENT: POSSIBILITIES AND PITFALLS

PAULINE REA-DICKINS

The University of Bristol, United Kingdom

ABSTRACT

This chapter examines the possibilities and pitfalls of classroom-based English language assessment, drawing on both the language testing and classroom assessment literature in English language education as well as educational assessment more generally. The chapter opens with a brief overview of different contexts for language testing and assessment: external, classroom-based, and second language acquisition research. The second part of the chapter presents research findings that highlight different facets of classroom-based assessment: the different meanings of and purposes for assessment, relationships between formative and summative assessment, approaches and frameworks used in teacher assessment, teacher perceptions and implementation of assessment, and the extent to which conventional measurement paradigms are appropriate for assessing the worth of instructional embedded assessment. These research findings lead into a discussion of current concerns and issues, as well as some of the potential pitfalls associated with classroom-based assessment. The final part of the chapter outlines future directions for the field and highlights some of the challenges for both research and professional practice in relation to classroom-oriented assessment.

INTRODUCTION

There is a long and well-established tradition of research in the area of testing as a measure of language proficiency. This continues to be the case, with significant developments in, for example, our understanding of validity (e.g., Kunnan, 1998; Read & Chapelle, 2001; Mislevy, Steinberg, & Almond, 2002; Bachman, 2004) influenced by Messick's now classic article (Messick, 1989), greater technical sophistication in the statistical analysis of test performance (e.g., Purpura, 1999) and multi-faceted Rasch measurement (e.g., McNamara, 1996; O'Loughlin, 2001), advances in the use of qualitative approaches in the test validation process (e.g., Banerjee & Luoma, 1997; Green, 1998), together with a greater understanding of the nature of test performance, its interpretation, and interactions in language assessment processes (e.g., O'Sullivan, Weir, & Saville 2002). There is also a well trodden path for tests in the measurement of language learning as outcomes from instruction—as evidence for the *goodness of fit* of a language program—and, as early landmarks, the program evaluations of the 1960s and 1970s are obvious examples. There are also much more recent examples such as the *school effectiveness movement* in the UK and other English-speaking countries where, in response to increasing concerns about accountability in education, the testing of school children is used as a means for making decisions about the effectiveness of schools (e.g., Scheerens, Glas, & Thomas, 2003).

Much of this research, focused on language proficiency or achievement testing, is referenced to a context that Shohamy (1994) identifies as *external*, defined as a context "in which standardized tests are used for making decisions about individuals and programs regarding, for instance, certificates, diplomas, acceptance, rejection and placement" (p. 133). There is, however, increasing recognition of the significant limitations of an exclusive focus on learning outcomes as a measure of learner performance and the importance of capturing relevant data within the lived curriculum not only as evidence of quality of the program but also, and importantly, of the language learning process itself. Shohamy (1994) identifies two other contexts in addition to the external context for language testing: "the classroom context, where tests are used as part of the teaching and learning process," and "the SLA research context, where language tests are used as tools for collecting language data in order to answer and test SLA research questions and hypotheses" (p. 133). This chapter explores in some detail the second of Shohamy's contexts, that of the language classroom (whether as a foreign or second/additional language); it also touches upon the relationships between assessment and SLA research, with specific reference to formative language assessment.

The next section introduces a number of different facets of classroom-based assessment and relates these to recent research and writing in the field. This is followed by a summary of current debates and concerns that, in turn, feed into the identification of a range of potential pitfalls in and inhibitors to the implementation of effective classroom assessment. The chapter concludes with an outline of future directions important in researching and implementing quality classroom assessment. In order to avoid connotations of *testing* with standardized measures of achievement or proficiency, and to attempt to situate the discussion within the socio-cultural context of the classroom, the term *assessment* is used to refer to approaches to the elicitation of learner language in the classroom.

FACETS OF CLASSROOM-BASED ASSESSMENT

The analysis of major aspects of classroom-based assessment that follows is organized around a number of key themes: meanings of classroom-based assessment; purposes for classroom-based assessment; assessment approaches, frameworks, and implementation; and paradigm-appropriate orientations.

Meanings of Classroom-based Assessment

In the same way that there is inconsistency in both the use and interpretation of the terms testing and assessment, there is also considerable variation surrounding the meanings of *classroom-based assessment*. For example, Valette (1994) distinguishes between assessment that is associated with school-based tests and large-scale proficiency tests. In contrast, Huerta-Macias (1995, p. 9) emphasizes that "there is little or no change required in classroom routines and activities in order to implement alternative assessment" (p. 9), which she sees as significantly different from standardized measures and pencil and paper test formats. Such assessment embedded within instruction claims validity in relation to both curricula and instructional relevance, and authenticity in terms of classroom teaching activities and processes. Huerta-Macias (1995, p. 9) draws parallels between alternative assessment and qualitative research (e.g., Guba & Lincoln, 1994), suggesting

trustworthiness and triangulation of data are more relevant in determining quality in alternative assessment than the criteria associated with the psychometric testing tradition. However, she also cites Wilde, Del Vecchio and Gustke (1995) who suggest that to ensure reliability in alternative assessments, "use trained judges, working with clear criteria, from specific anchor papers or performance behaviours," and "monitor periodically to ensure that raters use criteria and standards in a consistent manner" (Huerta-Marcias, 1995, p. 9). Huerta-Marcias recognizes the tension that exists between "teacher as supporting language development" and "teacher as examiner and rater," both roles for which teachers need to adapt, as appropriate, within the classroom. However, Brown and Hudson (1998, p. 655, 656) criticize Huerta-Macias's approach to reliability and validity as if these alternative procedures were of the add-on proficiency type measure:

> These statements [referring to the comments on trustworthiness and triangulation of data] are too general and short sighted to fit with our experiences as decision makers who...rely on the guidelines set forth in the *Standards for Educational and Psychological Testing* (American Psychological Association, 1985, 1986) for designing measures that will be used to make responsible decisions about students' lives ... As in all other forms of assessment, the designers and users of alternative assessments must make every effort to structure the way they design, pilot, analyse, and revise the procedures so the reliability and validity of the procedures can be studied, demonstrated, and improved. The resulting decision-making process should also take into account what testers know about the standard error of measurement and standards setting.

Clapham (2000, p. 152) has also applied traditional test criteria to alternative assessment:

> A problem with methods of alternative assessment, however, lies with their validity and reliability: Tasks are often not tried out to see whether they produce the desired linguistic information; marking criteria are not investigated to see whether they 'work'; and raters are often not trained to give consistent marks.

Both Brown and Hudson (1998) and Clapham (2000) are referring primarily to formal assessment procedures—albeit administered and implemented within classes—which have a high stakes purpose of some kind. These procedures are very different from those in which classroom assessment is used to inform language learning and teaching, and where assessment is seamlessly integrated into teaching and learning. As McNamara (2001, p. 343, 344) comments, when teachers and learners "engage in systematic reflection on the characteristics of an individual performance as an aid to the formulation of learning goals in a variety of contexts":

> This then means that the kinds of difficulties with subjective assessment that are exposed through careful validation research are not really an issue with this approach. From a certain perspective, each instance of this kind of assessment is unique; it does not always have to be fitted into a larger framework of comparison across individuals or across occasions ... Nor does this kind of assessment activity necessarily involve record keeping and reporting to fulfill managerialist agendas.

This "emergent" view of classroom-based assessment where learner performance is analyzed in terms of learning goals and instructional processes rather than a finished product introduces an important interactional perspective into assessment, critical to effective formative classroom language assessment (see also Rea-Dickins, 2001, Rea-Dickins, 2006 and Gardner, 2000). As Harlen and James (1997) comment:

> The kind of information that is gathered by teachers is not tidy, complete and self-consistent, but fragmentary and often contradictory (p. 376) ... However, where the purpose is to inform teaching and help learning, the fact that a pupil can do something in one context but apparently not in another is a positive advantage, since it gives clues to the conditions which seem to favor better performance and thus can be the basis for taking action. In this way, the validity, and usefulness of formative assessment is demonstrated and enhanced ... Through this rapid loop of feedback and adjustment between teacher and learner, the informational inevitably acquires greater reliability. (p. 371)

This analysis of the meanings of classroom-based assessment reveals different understandings of assessment derived from the different *purposes* for which learners are assessed, and the selection of an appropriate paradigm by which the goodness of fit to assessment purpose is established. The next section examines the different purposes of assessment in instructional contexts to provide a firmer framework for deconstructing the different meanings and potential roles for classroom assessment.

Purposes for Classroom Assessment

Purposes for classroom assessment are diverse, ranging from meeting the bureaucratic demands placed on teachers for data on learner achievement levels to assessment that has a primarily supportive function in the formative assessment of language learners and is firmly embedded within routine instructional contexts. These purposes, in turn, also give rise to different teacher and learner positioning in assessment (see Arkoudis & O'Loughlin, 2004).

The distinction conventionally drawn has been between summative and formative purposes for assessment, invariably contrasting one with the other, with much oversimplification of both of these constructs and the relationships between them. For example, *summative assessment* has been defined as assessment that takes place at the end of a school year for administrative purposes "in order to assign grades for purposes of certification or promoting students to the next level" (Genesee & Upshur, 1996, p. 49) or to "provide useful information ... of students' achievement or progress at the end of a course of study" (Bachman & Palmer, 1996, p. 98). In contrast, *formative assessment* is presented as helping "students guide their own subsequent learning, or for helping teachers modify their teaching methods and materials so as to make them more appropriate for students' needs, interests, and capabilities" (Bachman & Palmer, 1996, p. 98). Much of the discussion on formative assessment, however, has been couched in terms of formal achievement tests, thus, the focus on accurate and comprehensive profiling of language achievement (e.g., Brown & Hudson, 1998) is unsurprising. More recently, with the pervasive concern for national school league tables (e.g., in the UK) and for accountability to government and other agencies, there is increasing reference to the managerialist and summative purposes for assessment (see for example, Brindley, 2001; South, Leung, Rea-Dickins, Scott, Erduran, in progress[1]), which for the schools or programs concerned is high stakes.

In fact, there is relatively little empirical work on assessment purposes. An early study into the functions of teacher assessment was conducted by Brindley (1989) who asked teachers to rank the importance of a list of assessment functions. In terms of perceived importance to the teachers, it is interesting to note that they ranked lowest "providing information to funding authorities for accountability purposes," whereas "placing learners in class" and "providing information on learners'

strengths and weaknesses for course planning" were ranked 1 and 2 respectively (p. 25). With reference to teacher classroom assessment, Rea-Dickins and Gardner (2000) found a striking variety in classroom assessments implemented on a regular basis for the assessment of English language learners. From teacher self-reports they identified five main purposes for assessment: assessment used formatively to inform the management and planning of teaching to assessments used summatively to review learners' developing linguistic competence and skills, to provide feedback for bureaucratic purposes, to assess an individual's readiness to access the mainstream curriculum and to provide feedback on teaching. The idea that assessment might also be formative for the learners themselves did not emerge clearly as a major purpose in this analysis.

The blurring of the boundaries between formative and summative assessment is not as clear cut as usually represented. Teachers may use the same data obtained from assessments for different purposes at different time intervals, formative in one context (e.g., a child's language sample used to inform discussion at a teachers' planning meeting where action is agreed for language support for that individual learner) and summative in another (i.e. where that same language sample is used as part of a child's school Language Achievement Record). An analysis of ESL frameworks (South, Leung, Rea-Dickins, Scott, & Erduran, in progress[2]) also reveals the multi-purpose nature of teacher assessment as operationalized through assessment frameworks and standards (see also McKay, 2000). As Black (1998) comments: "The formative and summative labels describe two ends of a spectrum in school-based assessment rather than two isolated and completely different functions" (p. 35). In general, these purposes for classroom-based assessment remain largely unproblematized and unresearched.

ASSESSMENT APPROACHES, FRAMEWORKS, AND IMPLEMENTATION

Brown and Hudson (1998) provide a useful listing of assessment procedures beyond the familiar pencil and paper tests, including checklists, journals, video-tapes, portfolios, self- and peer-assessment. They also provide a synthesis of characteristics (drawing from Aschbacher, 1991; Herman, Aschbacher & Winter, 1992; Huerta-Macias, 1995) associated with these alternative assessments, e.g., "tap into higher level thinking and problem solving skills" (p. 654). As Shohamy (1998) comments: "Each procedure is aimed at capturing different aspects and domains of language knowledge, as it is assumed that language knowledge is exemplified differently in different contexts and situations" (p. 109). This perspective is also reinforced by a teacher, talking about her use of language sampling as an assessment tool (Gardner & Rea-Dickins, 2002):

> Once I sat down and the children were having dinner [midday meal], with a shy one at the beginning of the year, because she wouldn't speak. She didn't speak for weeks. And I caught her talking to a friend after a few weeks and I sat there in my lunch break and copied down two pages. It was just social chat. It wasn't academic type language, but it was the fact that she could talk at length if she was given the opportunity so, it's just—my system is ad hoc...it's just as and when I pick things (p. 6).

As Gardner and Rea-Dickins (2002) observe, it is not surprising that teachers rated language sampling as the least stressful form of assessment for learners, as it is

usually fully contextualized in day-to-day work, "an example of continuous, naturalistic performance testing, *par excellence*" (p. 6).

Assessment innovations of a different order that have impacted significantly on modes of teacher assessment are those associated with the development of language assessment frameworks and standards. These are used in various part of the world primarily for the assessment of school-age children using English as an additional language (see McKay and Nunan, this volume).[3] Although, there is considerable variation across these frameworks, several of them incorporate detailed guidance for the teacher in important areas of classroom assessment. The ESL Bandscales (National Languages and Literacy Institute of Australia, 1994) and the TESOL Standards (TESOL, 1997), in particular, go well beyond a discussion of assessment tools and the interpretation of learner language to demonstrate ways in which learner assessment may be integrated within teaching and learning processes and embedded within instruction (see Short, 2003; South, Rea-Dickins, Scott, & Erduran, in progress).

There is a growing literature about different assessment approaches and procedures, but although issues of classroom assessment are not new (most notably, see Brindley, 1989, 1995, 2000), relatively little has been written about the actual engagement of teachers and their learners—as evidenced by research studies—in the implementation of specific approaches and assessment activities. A number of Australian research studies have examined how teachers work with assessment frameworks and how they develop an understanding of assessment issues. Breen et al., 1997) investigated the implementation of assessment in primary schools, focusing on the relationship between assessment frameworks and teachers' pedagogic practice in making judgments about the English language development of their learners. In a three year longitudinal study Davison and Williams (2002) compared teachers' use of different assessment frameworks, including the ESL Bandscales (NLLIA, 1994) and the *English Curriculum Standards Frameworks* (Board of Studies, 2000). Arkoudis and O'Loughlin (2004) have investigated teachers' understandings of reliability and validity through using assessment frameworks to produce a meaningful assessment of their students' progress. Through an analysis of teachers' stories, they illustrate how state mandated assessment policies are translated into teacher assessment practices and how the teachers develop an awareness of the limitations of such frameworks, in this case *English as a Second Language (ESL) Companion to the English Curriculum Standards Frameworks* (Board of Studies, 2000). Such research also raises quite poignantly the broader issue of teacher/examiner role conflict, which is a particular challenge where integration and embeddedness of assessment are viewed as necessary. A comparative study of Hong Kong and Australia (Victoria) by Davison and Tang (2001) investigated ESL teacher assessment practices (e.g., choice of assessment tasks, criteria, teacher feedback to students) and their beliefs about language, language development, and assessment. This research revealed a high level of teacher awareness and acceptance of the need for accountability, particularly in high stakes assessment contexts as well as a need for more opportunities for teacher interaction about assessment issues. In a later study, Davison (2004) explores the tensions faced by teachers and the types of decisions they make when assessing student work and suggests that traditional norms of validity may need to be re-conceptualized in high stakes teacher-based assessment. Cheng, Rogers, and Hu (2004) have also identified the complex and multifaceted roles that assessment

plays in different language learning contexts based on a comparative survey of ESL/EFL instructors in Canadian, Hong Kong, and Chinese tertiary settings. Hamp-Lyons & Condon (2000) have researched the use of portfolios in helping "teachers help learners assume more responsibility for their own learning" and in providing "a rich source of information to teachers as they continually reconsider their theory and practice" (p. xv).

Rea-Dickins (2002) identified various influences on teacher assessment activities, revealing that English language teachers draw upon the mainstream curriculum (i.e. subject knowledge, learning objectives and outcomes), high stakes national tests, and psychometric notions of reliability and norming to inform their assessment activities. Some are also aware of an interactional perspective on classroom formative assessment and the importance of creating opportunities for sustained talk in the classroom.

Fewer studies still have adopted a learner and learning focus in instruction-embedded assessment. Within general educational assessment, there are notable exceptions. Tunstall and Gipps (1996), for example, have developed a typology to account for different types of teacher feedback to their learners that might lead to the promotion of curricula learning (i.e. not specifically language learning). This research is noteworthy as their feedback typology is grounded in the discourse of the classroom. A similar approach to researching classroom assessment was taken by Torrance and Pryor (1998) who investigated the impact of formative assessment on pupil learning.

In the field of language education, Rea-Dickins (2002) reports on the various ways in which learners may be scaffolded in their language and content learning, as they progress through their assessment activities. Drawing on Tunstall and Gipps (1996), a range of teacher feedback strategies were demonstrated: when teachers "specify" for the learner what needs to be worked on in order to improve their use of language, when they "encourage learner self-assessment," or when they are in dialogue with a child in "constructing next steps" within the learning activity. In addition, teachers were observed providing feedback of other kinds: encouraging children to elaborate and/or explain their utterances by use of questions or echoing strategies; and assisting language performance through teacher "recasts" (Nicholas, Lightbown, & Spada, 2001) as feedback on both the content and form of the children's utterances. The recasts observed included teacher "correction of errors," "recasting of a child's utterance," "expanding on learner contributions," "offering a target like model," and inviting the learner "to fill the gap." These scaffolding strategies support learners so that their awareness of language use across the curriculum is enhanced and their language and content learning further developed and enriched.

Recent and interesting work arising from early years' research in Holland and Germany has also drawn attention to teacher feedback and to the concept of a teacher's *diagnostic competence,* which the researchers define as "the ability to interpret foreign language growth in individual children" (Edelenbos & Kubanek-German, 2004). Data from both systematic observation and ethnographic classroom studies are used to identify and illustrate teachers' diagnostic activities and processes on the basis of which these researchers offer a preliminary description of levels of diagnostic competence and associated features.

In a university level language course, Spence-Brown (2001) investigated the construct of authenticity in an assessment activity designed "to optimise

authenticity" (p. 463). Through interviews with students she identified a range of factors that compromised the authenticity of a learning task when used for purposes of formal assessment, leading her to the conclusion that authenticity must be viewed in terms of the implementation of an activity as well as a function of its design. This relationship between the purpose(s) and design of an assessment and features of its actual implementation is highly important in classroom assessment research.

In terms of washback effects on classroom assessment processes from national assessment policies and associated frameworks and standards, there is overall relatively little research in spite of the growing number and use of assessment frameworks and standards (cf., Breen et al., 1997; Scott, 2005; Scott & Erduran, 2004).

PARADIGM-APPROPRIATE ORIENTATIONS

Implicit in the various understandings of classroom-based assessment and linked to the different purposes for assessment is the way in which classroom-based teacher assessment is conceptualized. The traditional positivist position on language testing, with the tendency to map the standard psychometric criteria of reliability and validity on to the classroom assessment procedures, has been called into question, and the scope of validity has been significantly broadened (e.g., Chapelle, 1999; Lynch, 2001, 2003; McNamara, 2001) and taken further by a number of researchers. Teasdale & Leung (2000), for example, highlighted the need to clarify the epistemological bases of different types of assessment within the context of the assessment of spoken English in mainstream classrooms. Drawing on both the TESOL and general educational assessment literature, Leung (2005) problematizes some of the "constitutive issues concerning pedagogically oriented classroom-based teacher assessment" (p. 869) and the tensions that exist for teachers in their dual roles in assessing and scaffolding learning. Through an analysis of classroom episodes and teacher interview data, he argues, "attending to teachers' professional knowledge and practice ... would contribute towards understanding the 'construct' in construct-referenced assessment" (p. 884). It is this kind of understanding, requiring the critical engagement of researchers with teachers, which Leung argues is critical for the development of a grounded, dynamic and contextually sensitive research agenda and, furthermore, that the evaluation criteria traditionally associated with psychometric testing such as reliability and validity are not necessarily relevant, "especially when the outcomes of teacher assessment are not used for public comparison and reporting purposes" (p. 885).

Appropriate paradigm orientation directly links to purposes for assessment. For purposes of accountability and normative and comparative rankings across or within schools, or when important decisions about individual learners are being made, a conceptualization of assessment as standardized measurement and the role of the teacher as rater/examiner has relevance (but see Davison, 2004). However, where the teacher's main role is to support learning and to provide opportunities in which learners feel able to use and stretch their linguistic resources in an attempt to convey their meanings to others in class, the priorities are different and other criteria have resonance. Thus classroom-based assessment represents an epistemological departure from the practice of framing research within established paradigms and theoretical models in the psychometric tradition.

CLASSROOM-BASED ASSESSMENT: POTENTIAL PITFALLS

There are a number of pitfalls in the implementation of effective classroom-based assessment and a number of potential inhibitors to quality assessment practices. One rather obvious one is that classroom assessment may be operationalized as the testing of linguistic knowledge that achieves little more than presenting learners with a series of summative mini-achievement tests. There are several points to consider here in relation to the potential mismatch between teaching and learning goals and classroom assessment practices. The first has to do with the motivation and rationale for teaching a foreign language (see Karavas-Doukas & Rea-Dickins, 1997). Within the primary language curriculum, in particular, there is a range of reasons for introducing a foreign or additional language. The reasons span the acquisition of structures and lexis or of communicative language ability, goals linked to developing language awareness and intra- and inter-cultural awareness (Kubanek-German, 1997), or the need to access subject knowledge through the medium of an additional language. The question is to what extent assessment activities mirror these diverse purposes and achieve an appropriate matching and balance in terms of "content" with reference to stated curriculum goals.

A second point has to do with the pedagogic approach and language skills actually assessed. Evidence from a small-scale case study (Rea-Dickins & Rixon, 1999) suggested that even though teachers recognized the need to assess both speaking and listening skills, this did not always happen. Where speaking skills were assessed, there was evidence that this was realized through rehearsed dialogues with little or no opportunity for spontaneous language use (cf., Gardner & Rea-Dickins, 2002), and the tendency was for teachers to rely on the tried and tested written assessment of structure and lexis and writing skills. Although there is a range of elicitation tools and frameworks described in the literature and in research studies (e.g., Genessee & Upshur, 1996; Hamp-Lyons & Condon, 2000; NLLIA Bandscales, National Languages and Literacy Institute of Australia, 1994), there seems to be a continued over-reliance on the paper and pencil format in preference to more observation-driven approaches to assessment. Even assessment frameworks, such as the ESL Standards (TESOL 1997), may be used normatively, with an over-reliance on summative tests that might result in limited opportunities for teachers to provide their learners with the necessary linguistic and cognitive structuring within instructional sequences.

More generally, there is evidence of a tendency in both handbooks for teachers (e.g., Hughes, 1990; Weir, 1993) and amongst teachers themselves (e.g., Rea-Dickins, 2003) for classroom assessment practices to be referenced to criteria associated with a psychometric approach to test validation and to normative standards, which in most circumstances have little or no relevance for the bulk of classroom-based formative assessment (for examples of paradigm confusion, see Teasdale and Leung, 2000; McNamara, 2001; Lynch, 2001). Given the mixed discourses of assessment prevalent in curriculum policy documents (e.g., QCA, 2000) and the emphasis in some countries on outcomes-based assessment of performance (e.g., the National Curriculum in England) (see Brindley, 2001), it is thus unsurprising that teachers also fail to grasp some of the nuances of classroom-based assessment. The problems include teachers employing limited means to capture knowledge and develop understandings of their learners' language abilities and failing to grasp the potential of collaborative dialogue for formative assessment.

Spence-Brown (2001), summarized earlier, draws our attention to the distinction between assessment plans and specifications and actual implementation. Classroom assessments may be developed according to a specific blueprint but may be implemented by students in ways that fundamentally compromise the intended design and characteristics. This is something that has become forcibly apparent in my own research where six teachers implemented the same assessment activities in very different ways that, in turn, provided the learners with opportunities for different kinds of engagement within the activity and use of different linguistic resources, some much more formative than others (Rea-Dickins, 2003). In fact, an activity or elicitation procedure in itself is neutral. It is only in its implementation and the use to which the data that emerges from a given activity is put that then develops its formative or summative potential, a point that is rarely given enough emphasis in the language testing literature.

A final pitfall is the dual roles of teachers – as assessor/tester vs. facilitator of language support – and how learners come to understand and perceive these dual functions within instruction. A distinction has been made between high and low-stakes assessment contexts, which together with the language attainment levels/development dimensions impact on the role of the teacher, whether as rater and examiner versus language teacher and facilitator (Arkoudis & O'Loughlin, 2004; Leung, 2004). They also affect the inherent trustworthiness and comprehensiveness of assessment activities that are embedded within routine instruction developing over time for individual learners in the classroom (Davison, 2004) and the criteria evoked, whether that be psychometric criteria drawn from standardized measurement or notions of *construct-referenced assessment* and *communities of practice* (Davison, 2004; Leung, 2005).

FUTURE DIRECTIONS

A number of central directions for classroom-based assessment practice and research can be identified, although obviously these reflect my own particular orientation to researching assessment in the language classroom and what I consider to be important to the development of greater understandings of assessment processes and their effects in relation to both teachers and learners.

"Researching" Classroom-based Assessment

Given the desired embeddedness of assessment within classroom processes, "in-flight" vs. "add on" assessment, it is suggested that the most appropriate way of investigating assessment in action is situated within a broad socio-cultural approach. This would facilitate an understanding of assessment practices and the language learning potential of these practices within the social and cultural context in which they take place. This theoretical positioning implies a methodology in which assessment is studied in depth within the ecology of the classroom, and one in which multi-layering techniques combining ethnography, discourse analysis, and linguistic description are appropriate. A layered approach was adopted by Rea-Dickins (2003) in which learner engagement in assessment activities analyzed from an interactional perspective proved particularly revealing.

The Centrality of the Learner

In language proficiency testing, we may observe over the last decade increased attention to examination processes and, in particular, on the test taker in, for example, the oral interview (O'Sullivan, 2002). By the same token, I believe that the way forward in classroom-based assessment is not only on elaborating teacher assessment processes but also, and importantly, developing greater understanding of the facets of classroom-based assessment through the lens of the learners. The research of Spence-Brown (2002) and Rea-Dickins (2002) are examples of this orientation in the area of language assessment. Two further examples from research in educational assessment have positioned the learner at the center of the assessment process: the LEARN Project (Weeden et al., 1999) focused their research around learners' views of assessment; and the Effective Lifelong Learning Inventory project (Deakin-Crick, Broadfoot & Claxton, 2004[4]) is investigating empirically the concept of students' learning power and potential.

Embedding Assessment within Classroom Learning and Learning

McNamara (1998, p. 311) draws his readers' attention to Spolsky (1995) who stressed "that tests and examinations have historically been a means of control and power ever since the original shibboleth test in the Bible" (see also Shohamy, 2001a, 2001b). If classroom assessment is interpreted as a series of summative tests in the classroom, disembedded from the flow of teaching and learning, this can be criticized as being unfair and a denial of formative language learning opportunities. It represents further evidence of the stranglehold that prevails in the form of language testing practices associated with external measures of language performance. There is, thus, a need to be alert to a change in emphasis from what learners have achieved—this becomes less of a priority for most of the time—to how learners can be supported in their language learning in different classroom situations through varied activities. Inherent in much current classroom assessment discourse is a view of assessment as a technicist endeavor, very probably linked to a policy context that prioritizes the use of assessment data for bureaucratic purposes of accountability and standard setting across schools, as well as the operationalization of learner language performance as *achievement*. This contrasts with a view of assessment as embedded within the socio-cultural practices of the classroom (see McNamara, 2001), one that also supports emergent language development. In the words of Gipps (1994):

> Assessment is an interactive, dynamic and collaborative activity. Rather than being external and formal in its implementation, assessment is integral to the teaching process and is embedded in the social and cultural life of the classroom. Such an approach can be seen as constructive and enabling because of its focus on assessing the process of learning, its attempt to elicit elaborated performance, and its emphasis on collaborative activity, whether the collaboration is with the teacher or a group of peers. (p. 158)

Research that attends to the relationships between assessment and instruction will be an important future focus.

Exploring Relationships between Formative Assessment and Second Language Acquisition

A limited number of researchers has examined the interfaces between language testing and assessment as highlighted by Shohamy and described in the introduction of this chapter (for exceptions, see the summary by Bachman & Cohen, 1998; Brindley, 1998; Shohamy, 2000). In the case of classroom embedded assessment and, in particular, assessment that is intended to promote the development of language learning and learner language, there is a direct and explicit link to be made with processes of second language acquisition. As Shohamy (2000) argues:

> The disciplines of language testing (LT) and second language acquisition (SLA) belong to the same field, that of language learning. They share similar goals of understanding the process of language learning, assessing it and looking for ways to improve it. It is, therefore, expected that the two disciplines would interact, share and contribute to one another. (p. 542)

The quality of teacher feedback and the impact of this feedback on student uptake and output become important in this respect. However, few classroom assessment studies have explored the interaction between the two disciplines (for some examples see Edelenbos & Kubanek-German, 2004; Leung & Mohan, 2004; and Rea-Dickins, 2002). Thus, there remains a need for increased understanding and collaborative work between the two fields: The impact of formative classroom assessment on acquisition needs to be explored and tracked. SLA studies that have teacher feedback as their focus are particularly useful starting points (e.g., Doughty & Williams, 1998; Ellis, Basturkmen, & Loewen, 2001).

Quality in Classroom-based Assessment

What constitutes quality in classroom-based assessment is a key question and an area for further research. As observed earlier, the means by which to achieve consistency in making judgments about language samples are well rehearsed, but much uncharted territory remains in the development of quality formative language assessment. The types of criteria that become important in classroom assessment include "resonance with curricula goals and instructional processes" and the provision of a "rich variety of opportunities" for learners to use and stretch their linguistic resources, using language appropriate to different contexts. The Assessment Reform Group (1994; see also Clarke, 1998, 2001; Wiliam, 2001), drawing on research in educational assessment, has developed principles for good practice in recognition that assessment for learning has the following characteristics:

- It is part of effective planning.
- It reflects how students learn.
- It is central to classroom practice.
- It is a key professional skill.
- It has an emotional impact.
- It affects learner motivation.
- It promotes commitment to learning goals and assessment criteria.
- It helps learners know how to improve.
- It encourages self- and peer-assessment.
- It recognizes all achievements.

Given that preservice training and professional development in the area of language testing and assessment may be rather "hit and miss," and many teachers are unfamiliar with the intricate relationships between formative and summative assessment, there are significant implications for teacher education as well.

CONCLUSIONS

A number of the tensions surrounding classroom-based assessment have been raised in this chapter, including the question of what is actually meant by classroom-based assessment, and the tendency in the assessment discourse for sharp distinctions to be made between a summative assessment activity and a formative one, with most research focusing on the former rather than the latter. This may be to the detriment of assessment opportunities that support student language learning. Good teaching—where teachers respond to learners' language learning and needs, with different types of feedback of an appropriate kind, of learner involvement through collaborative learning activities and self- and peer-assessment, with ample opportunities for language practice—implies good formative assessment practice. The next decade should see an increase in research on classroom-based assessment and a closer investigation of the linkages between formative classroom language assessment and second language acquisition. This is not, however, proposing an either/or situation, and it will be interesting to explore ways in which there might be a greater integration between the areas of language testing, classroom language assessment and second language acquisition in Applied Linguistics.

NOTES

[1] A review and critical evaluation of different assessment frameworks and standards, funded by the Paul Hamlyn Foundation and the National Association for Language Development in the Curriculum (NALDIC), 2002-2003. The research team comprises Hugh South, Constant Leung, Pauline Rea-Dickins, Catriona Scott, and Sibel Erduran.

[2] This research—Classroom Assessment of English as an Additional Language: Key Stage 1 Contexts—was funded by the Economic and Social Research Council (ESRC Major Research Grant R000238196, 1999 – 2002). Further details from: P.Rea-Dickins@bristol.ac.uk.

[3] These are *Canadian Language Benchmarks 2000: English as a second language – for adults* (Centre for Canadian Language Benchmarks, 2000); *South Australian Curriculum, Standards and Accountability Framework: English as a second language* (Department of Education, Training and Employment (DETE), South Australia, 2002); *A Language in Common* (QCA, 2000); ESL Development: Language and Literacy in Schools (National Languages and Literacy Institute of Australia (NLLIA), 1994); ESL Standards for Pre-K-12 Students—TESOL (TESOL Task Force, 1997), *ESL Companion to the English CSF: Curriculum and Standards Framework II* (Board of Studies, Victoria, 2000).

[4] The Effective Lifelong Learning Project is based in the Graduate School of Education, University of Bristol.

REFERENCES

Arkoudis, S., & O'Loughlin, K. (2004). Tensions between validity and outcomes: Teachers' assessment of written work of recently arrived immigrant ESL students. *Language Testing, 20*(3), 284–304.

Assessment Reform Group. (1999). *Assessment for learning: Beyond the Black Box*. Cambridge: University of Cambridge School of Education. See also: arg.educ.cam.ac.uk.

Bachman, L. F. 2004. *Statistical analysis for language assessment*. Cambridge: Cambridge University Press.

Bachman, L., & Cohen, A. (1998). *Interfaces between second language acquisition and language testing research*. Cambridge: Cambridge University Press.

Bachman, L., & Palmer A. (1996). *Language testing in practice*. Oxford: Oxford University Press.

Banerjee, J., & Luoma, S. (1997). Qualitative approaches to test validation. In C. Clapham, & D. Corson (Eds.), *Encyclopedia of Language and Education: Vol. 7. Language testing and assessment* (pp. 275–287). Dordecht: Kluwer.

Black, P. (1998). *Testing: Friend or foe?* London: The Falmer Press.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education, 5*(1), 7–74.

Board of Studies. (2000). *ESL companion to the English CSF*. Victoria.

Breen, M., Barratt-Pugh, C., Derewianka, B., House, H., Hudson, C., Lumley, T., & Rohl, M. (1997). *How teachers interpret and use national and state assessment frameworks.* (Vols. 1–3). Canberra, Department of Employment, Education, Training and Youth Affairs.

Brindley, G. (1989). *Assessing achievement in the learner-centred curriculum*. NCELTR Research Series. Macquarie University Sydney: National Centre for English Language Teaching and Research.

Brindley, G. (1995). (Ed.). *Language assessment in action*. Research Series. Macquarie University Sydney: National Centre for English Language Teaching and Research.

Brindley, G. (1998). Describing language development: Rating scales and SLA. In L. Bachman, & A. Cohen (Eds.), *Interfaces between second language acquisition and language testing research* (pp. 112–140). Cambridge: Cambridge University Press.

Brindley, G. (2001). Outcomes-based assessment in practice: Some examples and emerging insights. *Language Testing, 18*(4), 393–408.

Brown, J., & Hudson, T. (1998). The alternatives in language assessment. *TESOL Quarterly, 32,* 653–675.

Canadian Language Benchmarks. (2000). *Canadian language benchmarks 2000: English as a second language for adults*. Ottawa, ON: Centre for Canadian Language Benchmarks.

Chapelle, C. (1999). Validity in language assessment. *Annual Review of Applied Linguistics, 19,* 254–272.

Cheng, L., Rogers, T., & Hu, H. (2004). ESL/EFL instructors' classroom assessment practices: Purposes, methods and procedures. *Language Testing, 21*(3), 360–389.

Clapham, C. (2000). Assessment and testing. Annual Review of Applied Linguistics, 20, 147–161.

Clarke, S. (1998). *Targeting assessment in the primary classroom.* Bristol: Hodder & Stoughton.

Clarke, S. (2001). *Unlocking formative assessment.* London: Hodder & Stoughton.

Davies, A. (1997). Introduction: The limits of ethics in language testing. *Language Testing, 14*(3), 235–241.

Davison, C. (2004). The contradictory culture of teacher-based assessment: ESL teacher assessment practices in Australian and Hong Kong secondary schools. *Language Testing, 20*(3), 304–334.

Davison, C., & Tang, R. (2001, February). *The contradictory culture of school-based assessment: Teacher assessment practices in Hong Kong and Australia.* Paper presented at the American Association for Applied Linguistics, St Louis, USA.

Davison, C., & Williams, A. (Eds.). (2001). *Learning from each other: Literacy, labels and limitations.* Studies of child English language and literacy development K-12, Vol. 2. Melbourne: Language Australia.

Deakin-Crick, R., Broadfoot, P., & Claxton, G. (2004). Developing an effective lifelong learning inventory: The ELLI project. *Assessment in Education, 11*(3), 247–272.

Doughty, C., & Williams, J. (1998). *Focus on form in classroom second language* acquisition. Cambridge: Cambridge University Press.

Edelenbos, P., & Kubanek-German, A. (2004). Teacher assessment: The concept of 'diagnostic competence'. *Language Testing, 21*(3), 259–283.

Ellis, R., Basturkmen, H., & Loewen, S. (2001). Learner uptake in communicative ESL lessons. *Language Learning, 51*, 281–318.

Gardner, S., & Rea-Dickins, P. (2002). *Focus on language sampling: A key issue in EAL assessment.* London: National Association for Language Development in the Curriculum (NALDIC).

Genesee, F., & Upshur, J. (1996). *Classroom-based evaluation in second language education.* New York: Cambridge University Press.

Gipps, C. (1994). *Beyond testing: Towards a theory of educational assessment.* London: The Falmer Press.

Green, A. (1998). *Verbal protocol analysis in language testing research* (Vol. 5). Cambridge: Cambridge University Press.

Guba, E., & Lincoln, Y. (1994). *Handbook of qualitative research.* Thousand Oaks, CA.: Sage.

Gunn, M. (1995). Classroom-based assessment in intensive English centres. In G. Brindley (Ed.), *Language assessment in action* (pp. 239–270). Macquarie University Sydney: National Centre for English Language Teaching and Research.

Hamp-Lyons, L. (1997). Washback, impact and validity: Ethical concerns. *Language Testing, 14*(3), 295–303.

Hamp-Lyons, L., & Condon, W. (2000). *Assessing the portfolio: Principles for practice, theory, and research.* Cresskill, New Jersey: Hampton Press, Inc.

Harlen, W., & James, M. (1997). Assessment and learning: Differences and relationships between formative and summative assessment. *Assessment in Education, 4*(3), 365–379.

Huerta-Macías, A. (1995). Alternative assessment: Responses to commonly asked questions. *TESOL Journal, 5,* 8–11.

Hughes, A. (1990). *Testing for language teachers.* Cambridge: Cambridge University Press.

Karavas-Doukas, K., & Rea-Dickins, P. (1997). (Compilers.). *Evaluating innovation and establishing research priorities.* Euroconference Proceedings. University of Warwick: Centre for English Language Teacher Education. (Contact: P.Rea-Dickins@bristol.ac.uk).

Kubanek-German, A. (1997). Modern foreign languages in German primary schools: a case study. In K. Karavas & P. Rea-Dickins (Compilers.), 217–225.

Kunnan, A. (1998). (Ed.). *Validation in language assessment.* Mahwah, N.J.: Lawrence Earlbaum Associates, Inc. Publishers.

Leung, C. (2001b, February). *Working with pupil talk: Assumptions and stances in teacher assessment of spoken interaction in the mainstream classroom.* Paper presented at the American Association for Applied Linguistics, St Louis, USA.

Leung, C. (2002, April). *Validity as social construction—Stability and flux.* Paper presented at the American Association for Applied Linguistics, Arlington, VA, USA.

Leung, C. (2005). Classroom teacher assessment of second language development: Construct as practice. In E. Hinkel (Ed.), *Handbook of research in second language learning and teaching.* Mahwah, N.J.: Lawrence Earlbaum Associates, Inc. Publishers.

Leung, C., & Mohan, B. (2004). Teacher formative assessment and talk in classroom contexts: Assessment *as* discourse and assessment *of* discourse. *Language Testing, 21*(3), 335–359.

Lynch, B. (2001). Rethinking assessment from a critical perspective. *Language Testing, 18*(4), 351–372.

Lynch, B. (2003). *Language assessment and programme evaluation.* Edinburgh: Edinburgh University Press.

McKay, P. (2000). On ESL standards for school-age learners. *Language Testing, 17*(2), 185–214.

McNamara, T. (1996). *Measuring second language performance.* Harlow, Essex: Addison Wesley Longman Ltd.

McNamara, T. (1998). Policy and social considerations in language assessment. *Annual Review of Applied Linguistics, 18,* 304–319.

McNamara, T. (2001). Language assessment as social practice: Challenges for research. *Language Testing, 18*(4), 333–350.

Messick, S. (1989). Validity. In R. Linn (Ed.), *Educational measurement* (3rd ed.) (pp. 13–103). New York: Macmillan.

Mislevy, R. J., Steinberg, L. S., & Almond, R. G. 2002. Design and analysis in task-based language assessment. *Language Testing,* 19(4), 477–496.

National Languages and Literacy Institute of Australia. (1994). *ESL development: Language and literacy in schools* (Vol. 1). Canberra: The National Languages and Literacy Institute of Australia.

Nicholas, H., Lightbown, P., & Spada, N. (2001). Recasts as feedback to language learners. *Language Learning, 51,* 719–758.

O'Loughlin, K. (2001). *The equivalence of direct and semi-direct speaking tests.* Cambridge: Cambridge University Press.

O'Sullivan, B. (2002). Learner acquaintanceship and oral proficiency test pair-task performance. *Language Testing, 19*(3), 277–295.

O'Sullivan, B., Weir, C., & Saville, N. (2002). Using observation checklists to validate speaking-test tasks. *Language Testing, 19*(1), 33–56.

Purpura, J. (1999). Learner strategy use and performance on language tests: A structural equation modelling approach (Vol. 8). Cambridge: Cambridge University Press.

Qualifications and Curriculum Authority (QCA). (2000). *A language in common.*

Read, J., & Chapelle, C. (2001). A framework for second language vocabulary assessment. *Language Testing, 18*(1), 1–32.

Rea-Dickins, P., & Gardner, S. (2000). Snares and silver bullets: Disentangling the construct of formative assessment. *Language Testing, 17*(2), 215–243.

Rea-Dickins, P. (2001). Mirror, mirror on the wall: Identifying processes of classroom assessment. *Language Testing, 18*(4), 429–462.

Rea-Dickins, P. (2002). Exploring the educational potential of assessment with reference to learners with English as an additional language. In C. Leung (Ed.), *Language and additional/second language issues for school education: A reader for teachers* (pp. 81–93). York: York Publishing Services.

Rea-Dickins, P. (2003a). *Classroom Assessment of English as an Additional Language: Key Stage 1 Contexts.* End of Project Report for ESRC Major Research Grant R000238196, 1999–2002.

Rea-Dickins, P. (2003b, September). *Framing assessment: Breaking the mould.* Plenary Paper presented at TESOL Symposium on ESL/EFL Standards for Young Learners, University of Rome. Alexandria, Virginia: TESOL, Inc.

Rea-Dickins, P. (2006). Currents and eddies in the discourse of assessment: A learning-focused interpretation. *International Journal of Applied Linguistics, 16*(2), 163–188.

Rea-Dickins, P., & Rixon, S. (1999). Assessment of young learners' English—Reasons and means. In S. Rixon (Ed.), *Young learners of English: Some research perspectives* (pp. 89–101). Addison Wesley Longman/The British Council.

Scheerens, J., Glas, C., & Thomas, S. (2003). *Educational evaluation, assessment and monitoring: A systemic approach.* Lisse, Abingdon, Exton (PA), Tokyo: Swets & Zeitlinger.

Scott, C. (2005). *Washback in the UK primary context with EAL learners: Exploratory case studies.* Unpublished PhD Dissertation. Bristol: Graduate School of Education, University of Bristol.

Scott, K., & Erduran, S. (2004). Learning from international frameworks for assessment: EAL descriptors in Australia and the USA. *Language Testing, 21*(3), 409–431.

Shohamy, E. (1994). The role of language tests in the construction of second-language acquisition theories. In E. Tarone, S. Gass, & A. Cohen (Eds.), *Research methodology in second language acquisition* (pp. 133–42). Hillsdale NJ: Lawrence Erlbaum.

Shohamy, E. (1998). Applying a multiplism approach. In E. Li & G. James (Eds.), *Testing and evaluation in second language education.* Hong Kong: Language Centre, The University of Science and Technology.

Shohamy, E. (2000). The relationship between language testing and second language acquisition revisited. *System, 28*, 541–553.

Shohamy, E. (2001a). Democratic assessment as an alternative. *Language Testing, 18*(4), 373–392.

Shohamy, E. (2001b). *The power of tests: A critical perspective on the uses of language tests.* London: Pearson.

Spence-Brown, R. (2001). The eye of the beholder: Authenticity in an embedded assessment task. *Language Testing, 18*(4), 463–481.

Spolsky, B. (1995). *Measured words.* Oxford: Oxford University Press.

Teasdale, A., & Leung, C. (2000). Teacher assessment and psychometric theory: A case of paradigm crossing. *Language Testing, 17*(2), 163–184.

TESOL. (1997). *ESL standards for pre-K-12.* Alexandria, VA: Teachers of English to Speakers of Other Languages.

Torrance, H., & Pryor, J. (1998). *Investigating formative assessment.* Buckingham: Open University Press.

Tunstall, P., & Gipps, C. (1996). Teacher feedback to young children in formative assessment: A typology. *British Educational Research Journal, 22*(4), 389–416.

Valette, R. (1994). Teaching, testing and assessment: Conceptualizing the relationship. In C. Hancock (Ed.), *Teaching, testing and assessment: Making the connection* (pp. 1–42). Lincolnwood, IL: National Textbook Company.

Weeden, P., & Winter, J. with Broadfoot, P., Hinett, K., McNess, E., Tidmarsh, C., Triggs, P., & Wilmut (September 1999). *The LEARN Project: Learners' Expectations of Assessment for Learning Nationally.* Report for the Qualifications & Curriculum Authority. Bristol: Graduate School of Education.

Weir, C. (1993). *Understanding and developing language tests.* Hemel Hempstead, Hertfordshire: Prentice Hall International (UK) Ltd.

Wiliam, D. (2001). An overview of the relationship between assessment and the curriculum. Scott, D. (Eds.). *Curriculum and assessment.* (pp. 165–181). Westport, CT: Ablex Publishing.