Introduction to Language Classroom Assessment

After reading this chapter, you should be able to answer the following questions.

- What is language classroom assessment?
- What are the four major aspects of language classroom assessment?
- What research evidence do you need to know about language assessment?

Definition

Language testing and assessment is a relatively new field within the broad context of applied linguistics. This field is rooted in applied linguistics because language testing and assessment primarily deal with English language learners (ELLs) and test-takers, and consists of test designers, publishers, teachers, and researchers who have a strong interest in and influence on the teaching and learning of English around the world. The field also includes testing of languages other than English (e.g., the testing of Chinese as a second or foreign language). The major theoretical framework that guides the field of language testing and assessment is derived from educational measurement based on classical and modern test theory (e.g., Bachman, 1990) and theories from applied linguistics (e.g., Canale & Swain, 1980).

The nature of the field, as explained here, has reached two important landmarks: (1) what is at stake and important to the field (in this case, learners' learning of the English language and test-takers' test

performance of the English language) and (2) the shift from a focus on testing to a focus on assessment. This shift synchronizes with the worldwide movement in general education to combine assessment *of* learning and assessment *for* learning. *Assessment of learning* refers to those assessments that happen after learning has occurred to determine whether learning has happened. They are used to make statements of student learning status at a point in time. *Assessment for learning* refers to the process of seeking and interpreting evidence for use by students and their teachers to decide where the students are in their learning, where they need to go, and how best to get there.

The terms *testing* and *assessment* are both used in this book to show the historical development of the field. First, the focus was on testing, and then the focus moved to assessment, which includes testing. This second, broad definition includes large-scale testing such as the Test of English as a Foreign Language (TOEFL) and the International English Language Testing System (IELTS) and the classroom tests and assessments that teachers design and use in their own classrooms. This book focuses on the classroom tests, assessments, theories, and practices that guide language classroom assessment.

The definition of assessment can include events, tools, processes, and decision*s* (Taylor & Nolen, 2008). These four aspects, which are also fundamental to language classroom assessment, are discussed in subsequent chapters:

- *Assessment events* can support students when the events occur with enough frequency that the teacher knows whether instruction is successful and which student or group of students may need additional support.

- *Assessment tools* can support student learning when the tools give students clear ideas about what is important to learn and the criteria or expectations for good work, and when assessment matches with instruction.

- *Assessment processes* can support students, in that students see teachers as allies to their education; feedback can help students focus and better understand requirements.

- *Assessment decisions* can support students when grades accurately reflect what students learn.

Testing and Assessment Experience

Teachers make decisions about assessment practices on a day-to-day basis. Their decisions are influenced to a large extent by individual past testing and assessment experiences. For example, research has demonstrated that learning from past experience changed the circuitry in individuals' brains; thus individuals can categorize what they have observed, make a decision, and carry out appropriate actions. When something positive results from a decision, individuals are more prone to decide in a similar way, given a similar situation. On the other hand, individuals tend to avoid repeating past mistakes. Therefore, it is important for teachers to pause and reflect on their own past testing and assessment to guide their current and future assessment practices. The following activity is designed to guide you in this reflection.

Activity

1. Reflect on one experience in your life that made you feel good about being assessed or taking a test. Alternatively, you may reflect on one experience in your life that made you feel terrible as a result of being assessed or taking a test.

   - What was the experience?

   - Which aspects of testing and assessment were related to that experience?

   - Was the experience related to test content, test tasks, test types, or how the test was administered or marked?

2. Write down the factors that you feel have contributed to your positive experience in taking a language test, for example,

   - clear test directions,

   - familiar test formats, and

   - other factors.

   and the factors that you feel that have contributed to your negative experiences, for example,

   - noisy testing environment,

   - too little time, and

   - other factors.

3. How do these positive or negative factors contribute to the validity and reliability of assessment?

Validity refers to the accuracy and appropriateness of the test score and use. *Reliability* refers to the consistency of testing process in relation to test administration and scoring.

4. In Excerpts 1 and 2 below, read the test-takers' test-taking experiences. How do these experiences enhance or decrease the validity and reliability of assessment?

Excerpt 1. Negative Test-Taking Experience
(Cheng & Deluca, 2011, p. 110)

> I couldn't hear the tape clearly so I immediately told the invigilator and was sent to the special room (an alternate listening room). Already anxious about missing some listening materials, I was assigned to sit at the back of the room in front of windows that faced a basketball court. The door to my right was open, and I could hear the noise of students playing downstairs and of those walking past. A portable stereo player was placed on a chair at the front of the room. The volume was really low and I found it extremely difficult to follow the content.

Excerpt 2. Positive Test-Taking Experience
(Cheng & Deluca, 2011, p. 111)

> Usually when taking listening tests, the time limit is so tight it makes me nervous. Some tasks or questions appear after a long audio text such that test-takers have to memorize what they have just heard; the tests require test-takers to have good memory skills. In contrast, this listening test was chunked into four parts, and the first three parts were divided into further subparts. Before and after each part and subpart, sufficient time was given to test-takers to familiarize themselves with the questions or check their answers. These practices allow test-takers to feel at ease and enabled assessment of test-takers' listening ability—which should be the major criterion.

Excerpt 1 illustrates how test environmental factors contributed *negatively* to test-takers' perceptions of assessment fairness. Excerpt 2, on the other hand, describes a testing experience wherein time was managed so that this test-taker could effectively complete tasks and engage in deeper cognitive processes. Further, increasing timing and allowing breaks throughout the test was linked to reduced anxiety levels and better performance for this test-taker.

Understanding such testing and assessment experiences is essential for continued validation inquiry. Such inquiry can support teachers as they learn experientially and reflect critically on core aspects of assessment, such as construct representation (e.g., test administration and conditions, timing, test structure and content, and scoring) and construct-irrelevant variance (e.g., aspects of test use and consequences such as test coaching or preparation, emotions and self-efficacy, and misuses of test scores versus test purpose). Furthermore, understanding test-takers' experiences can illustrate the impact of test consequences on perceptions of test validity and help teachers to understand the interconnectedness of testing constructs, processes, and uses.

Language Testing and Assessment Research

Early developments in language testing and assessment were signified by the work of Oller (1979) on the nature of language ability as a single unitary construct (in which all four language skills are seen as a whole) and by the seminal work of Canale and Swain (1980) on the teaching and testing of language communicative competence. These works forced language testers to consider the sociolinguistic aspects of language use and the contexts in which language testing takes place. Canale and Swain's theoretical framework of communicative competence strongly influenced the working model of communicative language ability in language testing used by teachers and researchers. Bachman and Palmer (1996), language teachers in their early careers, pointed out that the working model of communicative language ability "provides a valuable framework for guiding the definition of constructs for any language testing development situation" (p. 67). Bachman (2000) further points out that language testing practice was informed "by a theoretical view of language ability as consisting of skills (listening, speaking, reading and writing) and components (e.g., grammar,

vocabulary, pronunciation) and an approach to test design that focused on testing isolated 'discrete points' of language, while the primary concern was with psychometric validity and reliability" (pp. 2–3).

The current debate remains whether teachers should assess language skills and components separately or integrated in large-scale testing and in classroom assessment contexts. This debate has implications for teachers because teachers design and conduct their own classroom assessments. Following is an exploration of how language testing and assessment have developed as an interdisciplinary field of applied linguistics.

Three very important aspects in language testing and assessment represent the major developments in the field over the past 50 years: test validation, washback research, and emerging research into teachers' classroom assessment practices, including recent developments in diagnostic assessment. It is essential for teachers who are interested in refining the classroom practices they engage in on a day-to-day basis to understand all three aspects of development.

At the core of any testing inquiry is *validity*. Like teachers, test designers and test users want to be confident in the meaning that is attached to testing results and the consequences of those results. More specifically, establishing the validity of a test corresponds to the accuracy and appropriateness associated with assessing student ability (Messick, 1989). Although some of these frameworks focus on establishing internal validity through an examination of psychometric processes within testing programs, which deals with the design, administration, and interpretation of tests (Bachman, 2000), others maintain a broader scope, considering contextual factors and social consequences of test validity (McNamara, 2007). Across these frameworks, there is a growing emphasis on collecting validity evidence from multiple stakeholders (e.g., teachers and students who use the tests) and using multiple research methods (e.g., instructional data from your own classroom).

Alderson and Wall (1993) proposed 15 hypotheses regarding the potential influence of language testing on various aspects of language teaching and learning. Cheng, Watanabe, and Curtis (2004) made the first systematic attempt to capture the essence of the washback phenomenon. *Washback* refers to the relationship between testing, teach-

ing, and learning. Through their collection of washback studies from around the world, they responded to the question *What does washback look like?* within the teaching and learning context. Since then, many large-scale empirical studies have been conducted by language teachers and researchers, and these studies have demonstrated the influence of testing on major aspects of teaching and learning. The results of these studies (e.g., Cheng, 2005) point to the urgent need for teachers to understand the nature of testing, to refine their own classroom assessment without unquestioningly accepting or following large-scale testing. Several other consequences have also been researched, including the cognitive aspects of testing (i.e., the cognitive processing of test tasks and formats), the use of test-taking strategies, the interaction of motivation and anxiety with test-takers' test performance, and the social aspects of testing (e.g., test-takers' perceptions and potential test uses or misuses within a specific context; see Cheng, 2008).

Only recently have studies of the assessment practices used by teachers of English started to emerge (Rea-Dickins, 2004; Brindley, 2007). Studies of classroom assessment practices used by teachers of English have been relatively limited compared with the fairly large body of studies of large-scale language testing practices in the field. What has been done so far has shown that teachers' assessment practices are often influenced by external testing, and these studies further identified a lack of assessment training in teacher education (Cheng, 2005). Breen et al. (1997) and Cumming (2001) found clear links between the instructional and assessment purposes held by English language teachers and the assessment practices they used.

Assessment *for* learning is currently one of the most talked about pedagogical approaches for enhancing student achievement. Teachers, as the agents of assessment, need to ensure the reliability and validity of their classroom assessment practices and use these practices in ways that support their students' learning (Black & Wiliam, 1998a, 1998b). Consequently, an increasing number of attempts have been made to adapt large-scale language testing to serve the purpose of supporting students' learning. However, the key to such classroom practice success in supporting student learning resides in *you*, as teachers, on using quality classroom assessment practices.

Based on the research evidence and developments discussed above, we know the field of language testing and assessment continue to work on assessment of language skills and components, to focus on the accuracy and appropriateness of test score interpretation, and, increasingly, to focus on the sociocultural context where testing and assessment take place. This concept of teaching and assessing language ability within context enables us to understand the interaction between the test, test-takers, and interlocutors (raters, markers, and teachers) within context. This concept also supports us in designing our own assessment, taking consideration of aspects of an assessment task. This particular focus has a great deal to offer in relation to how assessing the integrated skills should be conducted and marked as well as how and how fast feedback of test performance should be communicated to test-takers. This focus also has significance to language classroom assessment that teachers conduct on a day-to-day basis.

To emphasize, three aspects of research studies—test validation, washback research, and emerging research into teachers' classroom assessment practices (which is the focus of this book)—all have implications on how teachers design and conduct their day-to-day classroom assessment.