Diagnostic Second Language Assessment in the Classroom

Ian Blood
Teachers College, Columbia University

The term *diagnosis* is familiar, and refers to the identification of disease or disorder in an individual. In the broadest sense, then, diagnostic second language (L2) *assessment* refers to any L2 assessment practice, whether in the form of a formal written test or informal teacher questioning, that yields diagnostic feedback—information on learner strengths and weaknesses.

In low-stakes classroom contexts, where psychometric rigor is sacrificed for the attention and rich intuitions of teachers, informal diagnostic assessment occurs on a regular basis in the form of student questioning, explanation, and the provision of written feedback on quizzes, tests, and written work. Indeed, Huff and Goodman (2007) showed that K-12 science and language arts teachers highly value diagnostic feedback and prefer assessments which yield detailed information that can be used to identify the instructional needs of individual learners. Despite this apparent interest on the part of teachers in the diagnostic function of assessment, Alderson (2005, 2007) points out that diagnostic testing has been largely neglected in the L2 literature. Fundamental questions regarding the proper domain and application of diagnostic testing are unresolved: Must a diagnostic L2 test measure proficiency and be based on a theoretical model of L2 ability, or can diagnostic assessment be equally applied to achievement in a curriculum? Where does the boundary between formative assessment and diagnostic assessment lie? What kinds of feedback, and at what level of detail, are most beneficial to L2 learners?

Shohamy (1992) sees both proficiency and diagnostic testing as within the purview of diagnostic testing, arguing that, in classroom contexts, the usefulness of tests should be evaluated in terms of the degree to which they provide teachers and students with information on learning which can then be used to inform decisions on the focus of instruction and personal study. She proposes a testing model for the school context in which the generation of diagnostic feedback is a central feature. Such feedback should be "ample, detailed and innovative," and should involve "a variety of language dimensions within the four language skills" (Shohamy, 1992, p. 516). A diagnostic language test, therefore, should go beyond the rank ordering of students with respect to familiar, broad L2 constructs such as "speaking" or "reading" and should seek to give information, in as rich and useful a form as possible, on the test-taker's strengths and weaknesses with respect to the component knowledge, skills, and abilities of broader constructs.

While Shohamy's call for attention to be focused on designing classroom assessments that will provide rich diagnostic feedback is uncontroversial, it does raise a crucial question: Where should these assessments come from? Will they be held to the same measurement standards as higher-stakes tests? If so, do classroom teachers have access to the knowledge and resources necessary to create reliable and valid diagnostic tests? ETS has already begun development on a suite of computer-based diagnostic literacy assessments for classroom contexts (O'Reilly & Sheehan, 2009). Such assessments may prove quite valuable in helping teachers provide diagnostic feedback; however, the inherent weakness of such a test is that it can never be fully representative of the diverse language curricula in classrooms across the globe. A fundamental question, then, is how to resolve the tension between quality of measurement and contextual relevance. Large-scale tests provide more reliable measurement but can only give

feedback based on a theoretical model of literacy or a standardized literacy curriculum. Teacher-made tests, for their part, may lack the rigorous measurement standards of professionally developed tests, but they would be expected to be more grounded in the curriculum and potentially more relevant to the immediate needs of teachers and learners.

In view of the demand for diagnostic feedback on the part of teachers and the prominent role that feedback on language performance has been given in L2 acquisition theory (e.g., Gass, 2006), research into what assessment practices can best yield rich, detailed, and valid feedback must remain a goal of the field.

REFERENCES

Alderson, J. C. (2005). *Diagnosing foreign language proficiency: The interface between learning and assessment.* London, UK: Continuum International Publishing.

Alderson, J. C. (2007). The challenge of (diagnostic) testing: Do we know what we are measuring? In J. Fox, M. Wesche, D. Bayliss, L. Cheng, C. Turner, & C. Doe (Eds.), *Language testing reconsidered* (pp. 21-39). Ottawa: University of Ottawa Press.

Gass, S. M., & Mackey, A. (2006). Input, interaction and output: An overview. *AILA Review, 19*(1), 3-17.

Huff, K., & Goodman, D. (2007). The demand for cognitive diagnostic assessment. In J. P. Leighton & M. J. Gierl (Eds.), *Cognitive assessment for education: Theory and applications* (pp. 19-60). New York: Cambridge University Press.

O'Reilly, T., & Sheehan, K. (2009). *Cognitively Based Assessment of, for, and as Learning: A Framework for Assessing Reading Competency*. ETS Research Report RR-09-26. Princeton, NJ: Educational Testing Service. Retrieved from http://www.ets.org/Media/Research/pdf/RR-09-26.pdf

Shohamy, E. (1992). Beyond proficiency testing: A diagnostic feedback testing model for assessing foreign language learning. *The Modern Language Journal, 76*, 513-521.

Ian Blood is currently an Ed.M. student at Teachers College, Columbia University, studying second language assessment. His research interests include second language literacy assessment and learning-oriented assessment.