

An Overview of Studies on Diagnostic Testing and its Implications for the Development of Diagnostic Speaking Test

Zhongbao Zhao¹

¹ Foreign Languages College, Zhejiang Gongshang University, Hangzhou, China

Correspondence: Zhongbao Zhao, Foreign Languages College, Zhejiang Gongshang University, Hangzhou, No.18, Xuezheng Str., China. E-mail: Michaelzhao998@hotmail.com

Received: November 16, 2012 Accepted: December 14, 2012 Online Published: January 25, 2013

doi:10.5539/ijel.v3n1p41

URL: <http://dx.doi.org/10.5539/ijel.v3n1p41>

The research is financed by National Research Center for Foreign Language Education in China (Project ZGWYJYJJ2012A25)

Abstract

This paper aims at identifying the distinctive features of diagnostic testing and providing theoretical framework for the development of a diagnostic speaking test through a comprehensive literature review. First, the researcher analyzes the distinctive features and thus differentiates assessment for learning from assessment of learning. Then, the researcher does an overall review on relevant studies on the definition, characteristics and empirical studies of diagnostic testing with a view to shedding some light on the development of the diagnostic speaking test.

Keywords: assessment for learning, formative assessment, diagnostic testing, diagnostic speaking test

1. Introduction

Testing second language speaking is one of the youngest fields of research and practice in language testing. And it was only in the last 30 years that many studies of speaking tests were conducted on rating scales, task types, elicitation techniques, test reliability and validity (e.g., Fulcher, 1997, 2003; Gong, 2012; Luoma, 2004; O'Sullivan, Saville & Weir, 2002; Zhao, 2010; Zhao and Fan, 2012).

Studies in language testing have proved that a well-developed test can impact positive washback on both teaching and learning. With regard to the teaching and learning of college English in China, the development of a diagnostic English speaking test with high reliability and validity will enable teachers to diagnose students' English speaking ability, which can help them to determine the proper kinds and levels of difficulties of teaching and learning activities and improve teaching quality.

In order to have a better understanding of the distinctive features of diagnostic testing, the researcher first reviews studies on classroom assessment and assessment for learning, distinctions between formative and summative assessment, then pays special attention to the review on the definitions, characteristics and empirical studies about diagnostic testing for the purpose of shedding light on the development and validation of the diagnostic speaking test.

2. Overview of Studies on Diagnostic Testing

2.1 Assessment for Learning and Assessment of Learning

Classroom assessment provides a crucial theoretical and practical platform for teachers to carry out studies on teaching and learning. It serves as meaningful sources of information for teachers to know the quality of their teaching and the kind of activity they need to work on. It can also inform learners of their progress in learning.

The traditional concept of classroom assessment appears to be confined to a theoretical framework based on behaviorist views of cognition and development (Linn & Gronlund, 2000). Classroom assessment of the traditional type is mostly administered at the end of an instructional period with the purpose of assessing learning products, and it is always separated from teaching. According to Rennie (1998), the focus of second language education has switched from the mastery of discrete language skills, such as grammar and vocabulary to the development of communicative ability since the early 1980's. The shift from behaviorist psychology to cognitive

and constructive views of learning leads to the realization that classroom assessment may be a component of the teaching—learning process, and learners are considered as active participants rather than passive test takers. In this view, classroom assessment has become an important medium that can support and enhance learning (Shepard, 2000). In a similar way, Linn and Gronlund (2000) argue that classroom assessment can address both the learning process and learning outcomes: these measurements can provide ways for teachers to know both the effectiveness of the process used to perform a given task and the outcome of the performance. They also remark that classroom assessment plays a vital role in the evaluation of students' performance.

Brookhart (1999) conducted a study in two third-grade classrooms to investigate the importance classroom assessment in students' learning and achievements. The researcher found that informational feedback, which gave students information they could use to see what they knew and how they could do better next time, was critical to student effort and future success.

It is widely accepted nowadays that classroom assessment based on cognitive and constructive views of learning can provide both students and teachers with understandable information in a form they can use to improve performance. This is what Black and Wiliam (1998, p. 2) termed as assessment *for* learning:

“... all those activities undertaken by teachers, and by their students in assessing themselves, which provide information to be used as feedback to modify the teaching and learning activities in which they are engaged.”

The Assessment Reform Group (1999) defines assessment *for* learning as a kind of assessment whose major purpose is to enhance learning, instead of certifying and selecting pupils or making schools accountable.

Assessment *for* learning involves using strategies to guarantee that students understand their learning purposes and the criteria that will be used to evaluate their performance. The Assessment Reform Group (2002) suggests that the strategies should concentrate on students learning activities, recognize the role of the student's self esteem, and provide students with constructive guidance about how to improve, develop student's capacity of self assessment and recognize very kind of achievement of all students. Assessment *for* learning helps students to understand their own learning. It gives them immediate descriptive feedback about their own work to help them think about how to improve the quality of their learning. It can also help teachers to figure out students' needs and make preparations for their teaching activities.

Research shows that impressive progress can be made in students' achievement through consistently applying the principles of assessment *for* learning, especially for struggling learners (Black & Wiliam, 1998).

It is noted that assessment *for* learning is quite different from assessment *of* learning. The major purpose of Assessment *of* learning is to judge a student's level of performance on a specific task or at a certain period of teaching. Such type of assessment will provide information for schools' reporting. According to Department for Education and Skills (2004), assessment *for* learning is categorized as formative assessment, while assessment *of* learning belongs to summative assessment.

The impact of assessment for learning on students' performance has been explored by Black and Wiliam (1998). Their empirical study confirmed that assessment for learning can encourage students to take active parts in learning activities.

2.2 Formative Assessment and Summative Assessment

The differences between formative and summative assessment have been discussed by quite a number of researchers, who have identified the following major differences.

First, formative assessment focuses on the process while summative assessment pays attention to the product. Formative assessment is carried out in the process of teaching to evaluate students' learning and it involves getting the evidence of and adjusting teaching and learning.

Second, students get different types of feedback from the two kinds of assessment. Summative tests provide students with a score but they don't offer any diagnostic information about how to improve their study. On the other hand, students receive informative feedback from a formative test which can help them to improve their learning.

Third, formative assessment enables teachers to figure out the strengths and weaknesses of their students in a certain subject and thus helps them to select and structure teaching materials. Due to its purpose of facilitating teaching and learning, formative assessment is categorized as a low-stakes test.

Despite its obvious advantages over summative assessment, formative assessment still proves to be difficult to implement in classroom teaching. In this study, therefore, measures should be taken to solve these problems in

the development of the diagnostic speaking test in order to facilitate the implementation of diagnostic testing in classroom teaching.

2.3 Characteristics of Diagnostic Tests

It is clear that a diagnostic test is primarily used to discover learners' strengths and weaknesses and provide detailed feedback for both teachers and learners to make decisions.

For a diagnostic test, how to provide diagnostic feedback to test takers and teachers is of crucial importance. So the rating scale and the score reporting method should be given due attention. The essence of diagnostic testing lies in a comprehensive and creative feedback system. Alderson *et al* (1995) discuss extensively about a variety of problems on how to make score reports. The provision of feedback to students after they take a test is one of the unique features of diagnostic tests (Alderson, 2005).

In addition, the usefulness of feedback has also been investigated in educational assessment literature (e.g., Black & Wiliam, 1998). Black & Wiliam (1998) argue that descriptive diagnostic feedback will play an important role in the improvement of teaching and learning. They also carried out an extensive review about the impacts of feedback on students' performance and found out that feedback could improve learning outcomes. Wiggins (1998) argues that feedback should concentrate on weaknesses more than strengths, but it would be more helpful when coupled with concrete descriptions about better performance. This type of descriptive feedback makes the test results more understandable to examinees. McMillan and Workman (1998) assert that meaningful feedback specific to the assessment task provides necessary conditions for students to assess and adjust their work, thereby increasing their motivation, confidence and self-worth.

Feedback given to students can be more direct when it comes in the form of guidance and consists of advice on next steps or what should be done in order to improve or succeed (Black & Wiliam, 1998). Feedback that is used only to point out students' particular areas of strength or weakness can therefore be considered as indirect (e.g., Earl, 2003). Brown and Hudson (2002) point out that feedback can be provided to different users, such as students, teachers, administrators, etc. Useful feedback can help its users to make different instructional or learning decisions.

Feedback as diagnostic information is, firstly, of great importance to students. The diagnostic feedback can help students in understanding what learning objectives have not been achieved and need additional work. Such types of feedback enable students to set specific learning goals and improve their learner autonomy. Secondly, feedback can help teachers better attain teaching objectives and make relevant instructional adjustments. Thirdly, feedback can provide the administrators with valuable information to support their decision-making concerning language learning and teaching objectives and arrangements.

To sum up, in order to develop diagnostic feedback for the diagnostic speaking test, we should bear in mind that the feedback mechanisms must be considered in terms of who should receive the feedback, what should be reported, and how the report form should look like (e.g., Brown & Hudson, 2002; Council of Europe, 2001).

Since proficiency tests and placement tests can also serve for diagnostic purposes, it is very hard to distinguish diagnostic tests from other kinds of assessments through a simple comparison of their definitions. And until now few systematic studies have been conducted on the unique features of diagnostic tests. For the purpose of distinguishing diagnostic tests from other types of tests, Alderson (2005) put forward some distinctive characteristics of diagnostic testing through a synthesis of available literature. These hypothetical features may inspire language testers to make further efforts to improve the design of diagnostic tests. The hypothetical features of diagnostic tests cover the major aspects of language test development, including for example, test purpose, test construct, test content, test feedback, etc. These features may also serve as guidelines for the development of diagnostic tests. Among these features, the following ones offer some inspirations for the development of the DCEST in this study:

- “Diagnostic tests are designed to identify strengths and weaknesses in a learner’s knowledge and use of language.
- Diagnostic tests are more likely to focus on weaknesses than on strengths.
- Diagnostic tests should lead to remediation in further instruction.
- Diagnostic tests thus give detailed feedback which can be acted upon.
- Diagnostic tests provide immediate results, or results as little delayed as possible after test-taking.
- Diagnostic tests are based on some theory of language development, preferably a detailed theory rather a global theory”. (Alderson, 2005, p. 11)

2.4 Empirical Studies on Diagnostic Language Tests

When it comes to diagnostic tests in second or foreign language learning and development, Alderson (2005) argues that the currently available language testing literature offers very little guidance on how diagnosis should be appropriately conducted what content diagnostic tests should have, what theoretical basis they should rest on, and how their use should be validated. But in recent years, this situation has been improved and efforts have been made to the development and validation of diagnostic tests in foreign language testing.

With regard to the development of diagnostic tests, Sewell (2004) developed and administered a diagnostic test for the purpose of providing feedback on skills and abilities to adult learners in the UK. A variety of test tasks were employed including both multiple choices and open writing tasks. And research results demonstrated that the diagnostic test could provide a comprehensive diagnostic feedback on learners' strengths and weaknesses.

In addition, Jang (2005) investigated the validation process of the effects of the diagnostic cognitive reading skills assessment on teaching and learning. Empirical evidence showed that the diagnostic test was quite valid and could provide informative feedback to teachers and students. The researcher also proposed a customized diagnostic assessment framework as an alternative approach to formative diagnostic assessment. This system allowed teachers and other users to plan, customize, and implement a diagnostic test at any particular time of an instructional program.

Furthermore, another study on the development and validation of a diagnostic test of German was conducted by Richards (2008). The research results showed that the test could provide useful feedback about some language points. The researcher also stressed that the importance of self-assessment in the implementation of the diagnostic test. A problem with this research was that the effect of the test on students' learning was not explored.

With regard to the validation studies, Jang (2009) investigated the validity of the application of the Cognitive Diagnostic Assessment (CDA) and concentrated on the importance of the diagnostic feedback. The results indicated that the CDA approach was more effective in providing diagnostic feedback about the level of proficiency in reading skills than that of the traditional ways.

The above review shows that until now only a few studies on the development and validation of diagnostic tests have been carried out and very few of the studies were focused on the component skill of speaking in English. And further research on the consequential validity of diagnostic tests has been called for by a number of researchers (e.g., Alderson, 2005; Jang, 2008; 2009; Richards, 2008).

3. Conclusion

Through the comprehensive overview, it is concluded that assessment *for* learning and formative assessment can provide both theoretical and practical guidance for the development of diagnostic tests. It is suggested that feedback provided by diagnostic tests can be used to improve language teaching and learning. And it is argued that diagnostic testing is basically criterion-referenced because of its close relationship with classroom teaching and learning. Diagnostic testing is also considered as a typical kind of assessment *for* learning, which is regarded as an essential component of language teaching and learning.

The review indicates that a variety of studies have been carried out on the development of diagnostic tests, but few are focused on the effects of diagnostic feedback on teaching and learning, and on the validation studies of diagnostic tests. Therefore, it is hoped that the diagnostic speaking test to be developed can elicit and validate a procedural framework for college English teachers to produce their own diagnostic oral tests and to fill in the gap in the research of diagnostic testing.

References

- Alderson, J. C. (2005). *Diagnosing Foreign Language Proficiency: The Interface between Learning and Assessment*. London: Continuum.
- Assessment Reform Group. (1999). *Assessment for Learning: Beyond the Black Box*. Cambridge: University of Cambridge School of Education.
- Assessment Reform Group. (2002). *Assessment for Learning: 10 Principles*. Cambridge: University of Cambridge School of Education.
- Black, P., & Wiliam, D. (1998). Assessment and Classroom Learning. *Assessment in Education*, 5(1), 7-74. <http://dx.doi.org/10.1080/0969595980050102>
- Brookhart, S. M. (1999). The Art and Science of Classroom Assessment: The Missing Part of Pedagogy.

- ASHE-ERIC Higher Education Report*, 27(1). Washington, DC: The George Washington University, Graduate School of Education and Human Development.
- Brown, J. D., & Hudson, T. (2002). *Criterion-referenced Language Testing*. Cambridge: Cambridge University Press.
- Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge: Cambridge University Press.
- Earl, L. M. (2003). *Assessment as Learning: Using Classroom Assessment to Maximize Student Learning*. Thousand Oaks, CA: Corwin Press, Inc.
- Fulcher, G. (1997). The Testing of Speaking in a Second Language. In C. Clapham & D. Corson (Eds.), *Language Testing and Assessment, Vol. 7 of the Encyclopedia of Language Education* (pp. 75-85). Dordrecht: Kluwer Academic Publishers.
- Fulcher, G. (2003). *Testing Second Language Speaking*. London: Longman/Pearson Education.
- Gong, Z. (2012). Research on Foreign Language Classroom: Status quo and Implications—An Empirical Study on the Certain Papers Published from 1996 to 2011. *Journal of Zhejiang Gongshang University*, 2012(3), 82-88.
- Jang, E. E. (2005). *A Validity Narrative: Effects of Reading Skills Diagnosis on Teaching and Learning in the Context of NG TOEFL* (Unpublished doctoral dissertation). University of Illinois at Urbana-Champaign, IL.
- Jang, E. E. (2008). A Framework for Cognitive Diagnostic Assessment. In C. A. Chapelle, Y.-R. Chung, & J. Xu (Eds.), *Towards Adaptive CALL: Natural Language Processing for Diagnostic Language Assessment* (pp. 117-131). Amer, IA: Iowa State University.
- Jang, E. E. (2009). Cognitive Diagnostic Assessment of L2 Reading Comprehension Ability: Validity Arguments for Fusion Model Application to *LanguEdge* Assessment. *Language Testing*, 26, 31-73. <http://dx.doi.org/10.1177/0265532208097336>
- Linn, R. L., & Gronlund, N. E. (2000). *Measurement and Assessment in Teaching* (8th ed.). Upper Saddle River, NJ: Prentice-Hall.
- Luoma, S. (2004). *Assessing Speaking*. Cambridge: Cambridge University Press.
- McMillan, J. H., & Workman, D. J. (1998). *Classroom Assessment and Grading Practices: A Review of the Literature*. Richmond, VA: Metropolitan Educational Research Consortium.
- O'Sullivan, B., Saville, N., & Weir, C. (2002). Using Observation Checklists to Validate Speaking-test Tasks. *Language Testing*, 19(1), 33-56. <http://dx.doi.org/10.1191/0265532202lt219oa>
- Rennie, J. (1998). Current Trends in Foreign Language Assessment. *The ERIC Review*, 6(1), 27-31.
- Richards, B. J. (2008). Formative Assessment in Teacher Education: The Development of a Diagnostic Language Test for Trainee Teachers of German. *British Journal of Educational Studies*, 56(2), 184-204. <http://dx.doi.org/10.1111/j.1467-8527.2008.00403.x>
- Sewell, J. (2004). *Diagnostic Assessment within the Skills for Life Strategy*. Paper presented at the Conference of the International Association for Educational Assessment, Philadelphia, June 2004.
- Shepard, L. A. (2000). The Role of Assessment in a Learning Culture. *Educational Researcher*, 29(7), 4-14. <http://dx.doi.org/10.3102/0013189X029007004>
- Wiggins, G. (1998). *Educative Assessment: Designing Assessments to Inform and Improve Student Performance*. San Francisco, CA: Jossey-Bass Publishers.
- Zhao, Z. B. (2010). Review of Diagnostic Foreign Language Proficiency: The Interface between Learning and Assessment. *Foreign Languages World*, 2010(4), 91-94.
- Zhao, Z. B., & Fan, J. S. (2012). Review of Language Assessment in Practice: Developing Language Assessment and Justifying their use in the Real World, *Modern Foreign Languages*, 2012(1), 105-107.