

*The Journal of*  
*Nagasaki University of Foreign Studies*  
*No.23 2019*

**Classroom-based Language Testing and Assessment:  
Teaching Practices, Issues and Future Pathways**

Nyree JACOBS

教室内の言語テストと評価：教育実践および課題と今後の展望

ナイリー ジェーコブス

長崎外大論叢

第23号  
(別冊)

長崎外国語大学  
2019年12月

## Classroom-based Language Testing and Assessment: Teaching Practices, Issues and Future Pathways

Nyree JACOBS

### 教室内の言語テストと評価：教育実践および課題と今後の展望

ナイリー ジェーコブス

#### Abstract

Traditionally, classroom-based language assessment has taken the form of standardized paper and pencil tests. However, due to concerns about the effectiveness of assessing learners' knowledge based on limited scores and a narrow focus on standards and accountability, classroom-based assessment has adapted to reflect the interactive nature of the classroom and facilitate meaningful learning. This report looks at the theory behind classroom-based language assessment, provides a comparison of classroom-based assessment and large-scale tests, discusses alternative assessment tasks utilized for summative and formative purposes and considers reliability and validity when assessing in a classroom-based context. Future pathways will be explored and the report will conclude with a brief summary of implications for educators when implementing classroom-based assessment.

これまで、教室での語学学習の評価は、定式化された紙と鉛筆による形を取ってきた。しかしながら、限られたスコアに基づいて学習者の知識を評価することの有効性、基準となるものの焦点幅が狭いこと、さらに説明可能性に懸念があるため、教室での評価は教室での双方向性を反映して有意義な学習を促すように適合されてきた。本論文では、教室での語学学習の評価に関する理論を概観し、教室での評価と大規模なテストとの比較結果を提供し、総合的評価と形成的評価に用いられる代替的な評価について議論した上で、個別の教室での文脈において評価する際の信頼性と有効性を検討する。また、教育者が教室での評価を行うための意義と今後の展望を簡潔にまとめ、結論とする。

**Keywords:** Classroom-based assessment strategies, language, education

#### Introduction

Classroom-based assessment has been defined by Hill and McNamara (2012) as “any reflection by teachers (and/or learners) on the qualities of a learner’s (or group of learners’) work and the use of that information by teachers (and/or learners) for teaching, learning (feedback), reporting, management or socialization purposes” (pp. 396). Traditionally, assessment has taken the form of standardized paper and pencil tests; however, concerns have been raised about the effectiveness of assessing learners’ knowledge and abilities based on one or two single scores, where test anxiety and stress can impact on test performance (Charvade, Jahander & Khodabandehlou, 2012). Learning is driven by what teachers and students do in classrooms (Black & Wiliam, 2010; Tuinamuna, 2011). There are concerns that if teachers focus on standards and accountability that ignores classroom teaching and learning processes, teaching quality will become static, without room for improvement (Black & Wiliam,

2010). Classroom-based language assessment should reflect the interactive nature of the classroom, where teachers can adapt their teaching practices to meet students' needs (Black & Wiliam, 2010). Classroom-based assessment can take many forms including quizzes, portfolios, observation, and group discussion, and is primarily focused on formative assessment to modify teaching and learning activities, provide feedback, and facilitate meaningful learning.

This report will first look at the theory behind classroom-based assessment, followed by a comparison of the dimensions associated with classroom-based assessment and large-scale tests. Next, there will be a discussion of alternative assessment tasks used for summative and formative purposes, and an evaluation of their effectiveness. The appropriateness of reliability and validity will be considered in relation to classroom-based assessment, future pathways will be explored, and the report will conclude with a brief summary of implications for teachers when implementing classroom-based assessment.

### **Theory behind classroom-based assessment**

Over the years there has been a growing shift in attention towards the relationship between assessment and classroom learning, and away from restricted forms of tests that have been shown to have limited impact on students' learning experiences (Black & Wiliam, 1998). The underlying motive for this shift is that improved classroom-based assessment will result in enhanced learning experiences (Black & Wiliam, 1998). It has been argued by Black & Wiliam (2010) that "teaching and learning must be interactive" (pp. 82), and that teachers need to be aware of their students' progress, preferred learning styles, and any difficulties with learning to inform teaching methods and provide effective feedback to learners. Teachers can use a variety of methods to gather this information, including observation, reading students' written work (ie. portfolios), and group discussion (Black & Wiliam, 2010).

Classroom-based assessment incorporates both *summative assessment* (assessment of learning) and *formative assessment* (assessment for/as learning); however, there is a greater focus on the latter (These terms will be discussed further in the "Formative and Summative Assessment" section) (Hill & McNamara, 2012). The term *assessment* can refer to any activity undertaken by teachers, or by students in the form of self, or peer-assessment, that can provide information to be used as feedback to inform teaching and modify learning activities (Black & Wiliam, 2010). Such assessment is considered formative assessment when its use is seen to adapt teaching methods to meet the needs of students (Black & Wiliam, 2010). Questions have been raised regarding ways to improve formative assessment, and the relationship between improved formative assessment and raised standards within the classroom (Black & Wiliam, 2010). In addition, given the often intuitive nature of decision-making used by teachers during classroom-based assessment, there are concerns from researchers about how to identify *when* classroom-based assessment is taking place (Rea-Dickins, 2006 in Hill & McNamara, 2012). To improve teacher assessment practices, Hall et.al (1997 in Rea-Dickins, 2001) suggested a developmental model of assessment where "teachers progress through clearly defined stages in their planning for and implementation of assessment" (pp. 430). The first stage (Assessment planning stage) looks at developing assessment tasks based on targets specified in the curriculum (Rea-Dickins, 2001). Stage 2 (Observation stage) is concerned with getting to know the class as a whole and as individual learners in real class time in terms of their stage of development

and levels of attainment. Stage 2 is mainly focused on development and therefore most likely to be formative in purpose (Rea-Dickins, 2001). Stage 3 (Specific task stage) is closely linked to the curriculum with assessment focused on progress corresponding to curriculum criteria (Rea-Dickins, 2001). Stages 4 (Continuous review stage) and 5 (Levelling stage) are both summative in function. Stage 4 is concerned with creating student profiles to record progress and achievement, followed by Stage 5 where the information is used for placement purposes (Rea-Dickins, 2001). Originally suggested by Rea-Dickins (2001) and modified by Hill and McNamara (2012), the notion of *assessment opportunity* was proposed to allow for more intuitive forms of assessment. Assessment opportunity includes: “any actions, interactions or artifacts (planned or unplanned, deliberate or unconscious, explicit or embedded) which have the *potential* to provide information on the qualities of a learner’s (or group of learners’) performance” (Hill & McNamara, 2012, pp. 397). The definition was developed to allow observers to recognize any incidental forms of assessment that may take place in the classroom, so their effects can be examined (Hill & McNamara, 2012).

Within the scope of classroom-based assessment, recent studies have looked at purposes for using assessment data. According to Hill & McNamara (2012, pp. 405), assessment data has several functions, including “teaching”, “learning”, “reporting”, “management” and “socialization” respectively. Assessment may inform the pace of teaching, determine the type of feedback offered to students as individuals or as a group, be used to inform decisions about learners’ end-of-year reports, control and reinforce student behaviour, and create or maintain a positive classroom atmosphere (Hill & McNamara, 2012). *Learner understandings* is an area that has been highlighted as requiring further research. It has been suggested that teacher-to-learner interactions (i.e. concept checks) and learner-to-learner interactions (i.e. discussion of task requirements) can reveal learners’ existing understanding of task requirements and assessment procedures, and bridge any gaps between teacher intentions and learner understandings (Hill & McNamara, 2012).

### **Classroom-based assessment and large-scale tests**

There are several factors that separate classroom-based assessment from large-scale tests. In their classroom, teachers usually have a good understanding about the abilities and skills of the learners they are assessing, and have several opportunities to observe them participate in a wide range of tasks and activities, individually and in groups. Through their participation and interactions with the teacher and classmates in familiar surroundings, learners develop their skills and acquire new cultural knowledge (Fulcher & Davidson, 2007). In contrast, the assessors in large-scale tests are usually unfamiliar to test-takers, and are only able to assess them on one occasion. Moreover, interactions between assessors and test-takers are usually confined to instructions and requests. Factors that are assumed to be irrelevant to the test itself, such as room temperature or uncomfortable seats can impact on learners’ test performance (Fulcher & Davidson, 2007). In classroom-based assessment, the inclusion of a range of tasks that are interlinked is often considered a sign of good task design (Fulcher & Davidson, 2007). However, in large-scale tests, diversity and authenticity of test items is required to promote reliability and validity (Fulcher & Davidson, 2007). It is preferable for large-scale test designers to create tasks that can be scored by machines (multiple-choice test items for example), to ensure time and cost efficiency. These tests aim to control any factors that could appear to make assessments subjective as a matter of test fairness (Fulcher & Davidson, 2007).

Conversely, as teachers are familiar with the learners in their classroom, they can base their judgments about learners' ability on a wide range of accumulated evidence. Classroom-based assessments are often formative as opposed to summative, where the primary objective is to facilitate a positive washback effect on classroom activities rather than exclusively measuring students' success (Fulcher and Davidson, 2007).

### **Formative and Summative assessment**

Classroom-based assessment may be formative or summative in nature. *Summative assessment* is concerned with assessment of learning, and is usually conducted towards the end of a program of study to measure the success of groups or individuals, the program itself, and for placement or certification purposes (Fulcher & Davidson, 2007; Hill & McNamara, 2012). There is yet to be a consensus on a clear definition for the term *formative assessment*; however, for the purposes of this report, formative assessment will refer to any activities undertaken by educators and/or by their students, that can be used to modify teaching and learning activities, and provide information that can be used as feedback (Black & Wiliam, 1998). Rea-Dickins (2001) examined the different identities assessment can assume within the classroom. Assessment from a bureaucratic perspective reflects external expectations for teachers to provide information that can be used to track standards and in effect uphold school reputations. Assessments associated with a bureaucratic identity share features with formal language tests, and could be expected to be summative in purpose (Rea-Dickins, 2001). According to Rea-Dickins (2001), assessment also has an internal identity associated with pedagogic and learning functions. From this perspective, assessment "contributes to knowledge about groups of individuals or learners and influences instructional decisions" while "contributing to learning as opposed to measuring learning" (pp. 452). Assessment within this identity could be closely linked to formative assessment.

Recent studies surrounding classroom-based assessment have raised concerns about using one or two single scores to evaluate students' knowledge. Innovations in assessment procedures have seen a shift from summative to formative assessment, which have become widely used to promote assessment as a tool to scaffold learning, rather than utilising habitual testing methods (Rea-Dickins, 2001; Charvade, Jahander & Khodabandehlou, 2012). Classroom assessment can take many forms with conceptualizations that often overlap, including *criterion-referenced assessment* where learners are required to meet specific criteria, and *alternative assessment*, which encompasses any assessment that provides alternatives to traditional testing methods (Leung, 2005; Fulcher & Davidson, 2007). The following sub-sections will look at examples of alternative assessment within a classroom context for formative and summative purposes, and evaluate their effectiveness.

### **Self-assessment - Portfolios**

It has been suggested that self-assessment effectively promotes autonomous language learning because it encourages learners to stay focused on their development and assess their learning progress. Through self-assessment, learners can gain insight into their strengths and weaknesses and establish attainable goals (Charvade, Jahander & Khodabandehlou, 2012). One instrument commonly used for self-assessment is the portfolio. Previously, portfolios were used mainly as a tool for performance-based assessment. However, recently, portfolios have been used as an instrument of assessment as well as an opportunity for learners to take responsibility for

monitoring their progress and reaching their goals (Charvade, Jahander & Khodabandehlou, 2012). Recent studies have looked at the effects of portfolio assessment on learners' reading comprehension ability. Traditionally, reading assessment involves the use of standardized tests with time limits, which can impact on a learner's performance. Another limitation of traditional reading assessment tools is that they focus more on discrete skills rather than a culmination of the learner's problem-solving and cognitive abilities (Charvade, Jahander & Khodabandehlou, 2012). Traditional assessment assumes that reading skills can be measured objectively, which would be beneficial if the learner's goal was to gain employment in a context such as a factory or hospital where proficient reading comprehension would be considered a requirement (Charvade, Jahander & Khodabandehlou, 2012). However, for an EFL classroom context, the use of a portfolio for formative purposes could assist students to develop real-world skills such as negotiation of meaning and decision-making skills. A study conducted by Charvade, Jahander & Khodabandehlou (2012) supports the claim that portfolio assessment has a positive effect on EFL learners' reading comprehension ability, and backs previous findings concerning the benefits of using portfolio assessment to provide ongoing feedback within instruction, and involve students in the process of their own assessment.

### **Observation – Role-play**

Observation is another form of classroom-based language assessment where it would be beneficial for students to be involved in the assessment process. Particularly with the assessment of speaking, using a student-involved classroom-assessment process can be both a motivation tool, and a way to increase learners' confidence (Stiggins & Chappuis, 2005). A speaking assessment in a standardized test only offers learners one opportunity to perform to the best of their ability while faced with factors such as test anxiety and unfamiliar materials, which can impact on their performance.

There are also issues surrounding subjective assessments of a learner's speaking ability, which can influence the grade they achieve. Ongoing classroom assessments of speaking through the observation of role-play activities can be used to encourage students' confidence, and help them identify their own strengths and weaknesses and those of their peers (Stiggins & Chappuis, 2005). In most ESL/EFL contexts, a formative approach to the assessment of speaking helps students gain a sense of control over their development, and reduces achievement gaps (Stiggins & Chappuis, 2005). However, in some cases observation of speaking skills for summative purposes within a classroom context would be appropriate. Universities in Japan commonly use role-play activities to assess learners' spoken communication skills, with students' scores contributing to the successful completion of their university course. These scores are also often of great interest to potential employers, and can determine graduates' success in securing employment. Classroom-based assessments of this nature are becoming increasingly high-stakes, not only for the learners being assessed, but also for institutions who are required to report students' progress to meet national standards, and maintain their reputations in a competitive industry (Brindley, 1998). As a side effect, teachers are finding themselves under pressure to carry out detailed assessments at an individual level, while at the same time report outcomes to external audiences in order to meet accountability requirements (Brindley, 1998).

## E-learning Assessment

More recently, some teachers are opting to implement e-learning assessment tools based on materials learnt in classroom contexts. Although e-learning assessment, more often than not, takes place outside of the classroom, many teachers utilise results to address learner deficits, to inform classroom-based teaching pedagogy, and as a contribution towards final grades. However, concerns have been raised over the effectiveness of e-learning assessment in relation to individual learning styles, technical implications, and learning and knowledge management (Barbosa & Garcia, 2005). In addition, institutions and educators need to be mindful when implementing standardised e-learning assessment programs in terms of relevance to content taught in the classroom and validity and reliability of assessment.

## Validity and Reliability

Classroom-based assessment is generally accepted as a useful form to measure language skills. However, questions have been raised about its reliability and validity. Huerta-Macias (1995 in Leung, 2005) argues that these notions, which are typically associated with standardised tests, should be viewed in terms of *trustworthiness*, or in other words, how well the assessment depicts real-life classroom activities, and *auditability*, which assumes that an assessment will produce the same results if replicated or audited, when applied to classroom-based assessment. Gipps (1994 in Leung, 2005) suggests replacing traditional reliability measures with *comparability*, which is based on consistency. According to Gipps, performance assessment needs to carefully consider comparability if used for accountability purposes, however, for classroom-based formative assessment purposes, comparability is less of a concern (Leung, 2005).

Cohen (1994 in Leung, 2005) recommends that teachers evaluate their own assessment tools, and that the reliability of assessment instruments incorporates three factors, "...test factors, situation factors, and individual factors" (Leung, 2005 pp. 872). From this perspective, it appears classroom-based assessment is grouped within the more general concept of performance assessment (Leung, 2005). In a previous section, the dimensions of classroom-based assessment and large-scale tests were compared revealing several distinct features. Psychometric tests are based around an individual and are expected to be quantifiable, but have a limited washback effect on classroom practices. In contrast, classroom-based assessment is more interactive and often informal in nature. Information gained through informal assessment is used to identify students' needs and inform teaching and learning practices, so many language teachers view their role as an assessor as contributing to *consequential validity* (Leung, 2005; Fulcher & Davidson, 2007). Classroom-based assessment tasks are often unique, and as assessment conditions are not initially controlled, they are almost impossible to replicate (Leung, 2005). Large-scale tests are usually high-stakes leading to important outcomes for test-takers. By comparison, the relaxation of reliability seems appropriate for classroom-based assessment on account of the low-stakes nature (Leung, 2005). Based on the differences between large-scale testing and classroom-based assessment, it can be argued that the evaluation criteria regarding reliability and validity that are traditionally associated with large-scale tests are not necessarily relevant to classroom-based assessment, particularly when the outcomes are commonly used to modify teaching practices as opposed to being used for external reporting purposes (Leung, 2005).



### Future Pathways for Classroom-based Assessment

In his book *Measuring Second Language Performance*, McNamara (1996, in Harding, 2014) addressed the factors involved in communicative competence in an additional language and noted the absence of “ability for use” in existing models of communicative language ability (CLA). *Ability for use* takes into account not only an individual’s ability to communicate, but also their capacity to communicate across various language settings (Harding, 2014). In classroom-based language testing, it has been suggested that, to address this deficit and increase validity and reliability of assessment, where possible, students are paired with, or put in group-speaking assessment tasks with test-takers from other L1 backgrounds (Harding, 2014). Particularly in higher education, as classrooms continue to become more internationalised, opportunities for this type of assessment is likely to increase creating more valid and reliable methods of testing with a wider scope.

*Feed forward*, a term that has been popularized recently, is focused on future outcomes. Although there appears to be no set definition, feed forward refers to either feedback given by educators after an assignment that includes specific direction for future assignments, or feedback that can be applied to an upcoming assessment (Hine & Northeast, 2016). When feed forward strategies are employed by educators within assignments (i.e. when feedback is given before an assignment’s deadline, so it can have a direct impact on that assignment) it can assume a formative assessment role, creating opportunities for teachers and students to engage in meaningful dialogue, and offering learners more autonomy over their learning outcomes (Dulama & Ilovan, 2016; Hine & Northeast, 2016). Studies have shown problems with feedback and learner dissatisfaction with the content and timing of feedback and ambiguity surrounding assessment requirements (Hine & Northeast, 2016). The adoption of feed forward strategies has proven to improve student performance and motivation, and modify teaching and learning activities to create a positive learning environment.

### Conclusion

Studies have shown that traditional testing methods are limited in their measurement of student learning, and provide a narrow washback effect on classroom practices, which has contributed to a shift to classroom-based assessment (Clark, 2008). It is widely accepted that classroom-based assessment is a useful tool for the assessment of learning and assessment for learning opportunities. However, achieving a balance between summative and formative assessment has proven challenging for educators. Ambiguity surrounds definitions of terms such as ‘formative assessment’, and even a definition for the term ‘classroom-based assessment’ has not reached a consensus (Hill & McNamara, 2012).

Teachers are involved in a juggling act of undertaking some summative assessment to satisfy institutional requirements and meet the expectations of stakeholders including parents and external agencies, while using formative assessment to inform teaching practices, and scaffold learning through feedback (Black & Wiliam, 2010). Although frameworks are being developed to provide a more structured approach to assessment for formative purposes, teachers are still at risk of being held accountable for using more intuitive forms of assessment such as those pertaining to the concept of ‘assessment opportunity’ that are not easily identifiable (Hill & McNamara, 2012). In addition, it’s possible that teachers require more professional development and support to implement more modern assessment forms such as self-assessment, to help them achieve the outcomes for which



they are responsible (Brindley, 1998). The notions of reliability and validity that are applied to traditional large-scale tests need to be reconsidered in relation to classroom-based assessment in recognition of the variations that exist between these two contexts (Leung, 2005). In a classroom situation, adhering to strict measures of reliability and validity could restrict innovation in pedagogy, so instead, teachers need to focus on the needs of learners, and use assessment as a tool to motivate, engage, and develop the confidence of their students (Rea-Dickins, 2001).

## References

- Barbosa, H., & Garcia, F. (2005). Importance of online assessment in the e-learning process. In *2005 6th International Conference on Information Technology Based Higher Education and Training* (pp. F3B-1). IEEE.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*, 5(1), pp. 7-74.
- Black, P., & Wiliam, D. (2010). Kappan Classic: Inside the black box: Raising Standards Through Classroom Assessment: Formative assessment is an essential Component of classroom work and can raise student achievement. *Phi Delta Kappan*, 92(1), pp. 81-90.
- Brindley, G. (1998). Outcomes-based assessment and reporting in language learning programs: a review of the issues. *Language Testing* 15(1), pp. 45-85.
- Charvade, K.R., Jahandar, S., & Khodabandehlou, M. (2012). The Impact of Portfolio Assessment of EFL Learners' Reading Comprehension Ability. *English Language Teaching*, 5(7), pp. 129-139.
- Clark, I. (2008). Assessment is for learning: Formative Assessment and Positive Learning Interactions. *Florida Journal of Educational Administration and Policy*, 2(1), pp. 1-16.
- Dulama, M. E., & Ilovan, O. R. (2016). How Powerful Is Feedforward in University Education? A Case Study in Romanian Geography Education on Increasing Learning Efficiency. *Educational Sciences: Theory and Practice*, 16(3), 827-848.
- Fulcher, G., & Davidson, F. (2007). *Language testing and assessment: An advanced resource book*. Abingdon, UK: Routledge.
- Harding, L. (2014). Communicative language testing: Current issues and future research. *Language Assessment Quarterly*, 11(2), 186-197.
- Hill, K., & McNamara, T. (2012). Developing a comprehensive, empirically based research framework for classroom-based assessment. *Language Testing*, 29(3), pp. 395-420.
- Hine, B., & Northeast, T. (2016). Using feed-forward strategies in higher education. The terrifying novel assignment: using feed-forward to improve students' ability and confidence on assignments that test new skills. *New Vistas*, 2(1), 28-33.
- Leung, C. (2005). Classroom teacher assessment of second language development: Construct as practice. In E Hinkel (Ed.), *Handbook of research in second language teaching and learning*. Mahwah, New Jersey: Lawrence Erlbaum Associates, pp. 869-888.
- Rea-Dickins, P. (2001). Mirror, mirror on the wall: Identifying processes of classroom assessment. *Language Testing*, 18(4), pp. 429-462.
- Stiggins, R., & Chappuis, J. (2005). Using Student-Involved Classroom Assessment to Close Achievement Gaps.

*Theory Into Practice*, 44(1), pp. 11-18.

Tuinamuana, K. (2011). Teacher professional standards, accountability, and ideology: Alternative discourses.  
*Australian Journal of Teacher Education (Online)*, 36(12), 72.

