

# Designing Assessment Tools: The Principles of Language Assessment

Mitaka Yoneda

*Department of English Communication, Junior College Division  
Mukogawa Women's University, Nishinomiya, 663-8558, Japan*

## Abstract

This article looks at the issue of assessment in language classes, presenting the theory behind it along with three models of classroom assessment. Because assessment is not synonymous with testing, it should be continuous throughout the teaching and learning process in the classroom. Teachers can assess students in many ways other than administering tests. It is important, for example, to apply triangulation, wherein at least three performances or factors are used to formulate a grade. When teachers design assessment tools for class, they need to consider the five principles of language assessment: practicality, reliability, validity, authenticity, and washback. By applying triangulation and meeting the standards of the five principles of language assessment, I create a model to assess my students' achievement in a university Business English II class, a fourth-year course designed to enhance students' business presentation skills. The planned assessment for this class includes vocabulary quizzes, paper-and-pencil tests, and presentation of a project. Students will obtain feedback, both written and verbal, from the instructor, their peers, and themselves. Through this model, I hope to deepen the understanding of language assessment design and its application in real-life teaching situations.

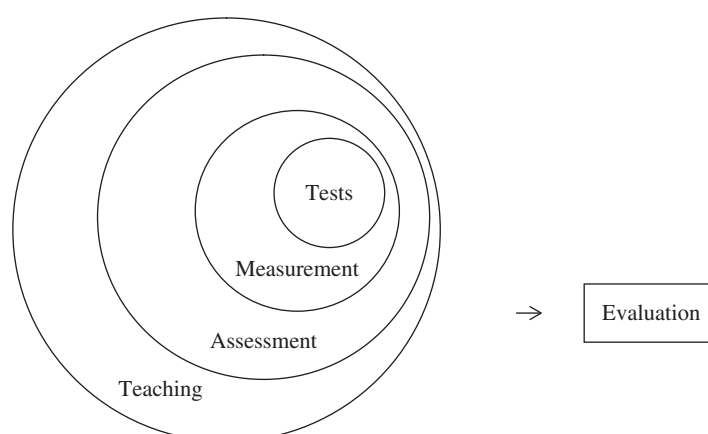
## Introduction

The method of assessing student progress is one of the vital decisions a teacher must take. In order to achieve optimal assessment results, it is important to design the assessment instrument within a theoretical framework. In this paper, I provide an overview of the theories of English language assessment presented by Brown and Abeywickrama (2010) and then apply the theories to my current teaching situation. The assessment tools used here are just one sample of such tools, but through them, I hope to present the theory of assessment and thereby enhance readers' understanding of language assessment in the classroom.

## Definition and Categories of Assessment

Assessment can be confusing and can mean different things, depending upon individual perceptions: some people think that assessment is the same as testing; some may think that grading is synonymous with assessment; and others may think that assessment equals measurement. Brown and Abeywickrama (2010, p. 6) attempted to clarify these concepts by highlighting the relationships between tests, measurement, assessment, and teaching (see Fig. 1).

Drawing from Brown and Abeywickrama (2010), we can see that assessment in educational practice is not merely testing or measurement. Rather, tests are "a subset of assessment, but they are certainly not the only form of assessment that a teacher can make" (Brown & Abeywickrama, 2010, p. 5). That is, teachers can assess students without using tests. Evaluation takes place when a value—such as that derived from grading and



*Fig. 1. Tests, measurement, assessment, teaching, and evaluation (Brown & Abeywickrama, 2010, p. 6)*

scoring—is associated with assessment (Linville, 2011, Unit 2, p. 5). It is designed to “motivate learners by giving them feedback on their progress” (Linville, 2011, Unit 2, p. 4). Cameron (2001) also explained that assessments can motivate learners, help teachers plan more effective lessons, inform the improvement of instructional programs, and support further learning with feedback (pp. 214-218). Cameron also notes that assessment must be a collaborative act between teacher and student, and that it must support the aims of the lesson. Hence, teachers should keep in mind that if students are not given appropriate feedback or if the evaluation focuses only “on what they don’t know or... [is] punitive” (Linville, 2011, Unit 2, p. 5), assessment can be discouraging. Thus, assessment, although necessary, useful, and instructive, must be well designed and implemented to be effective.

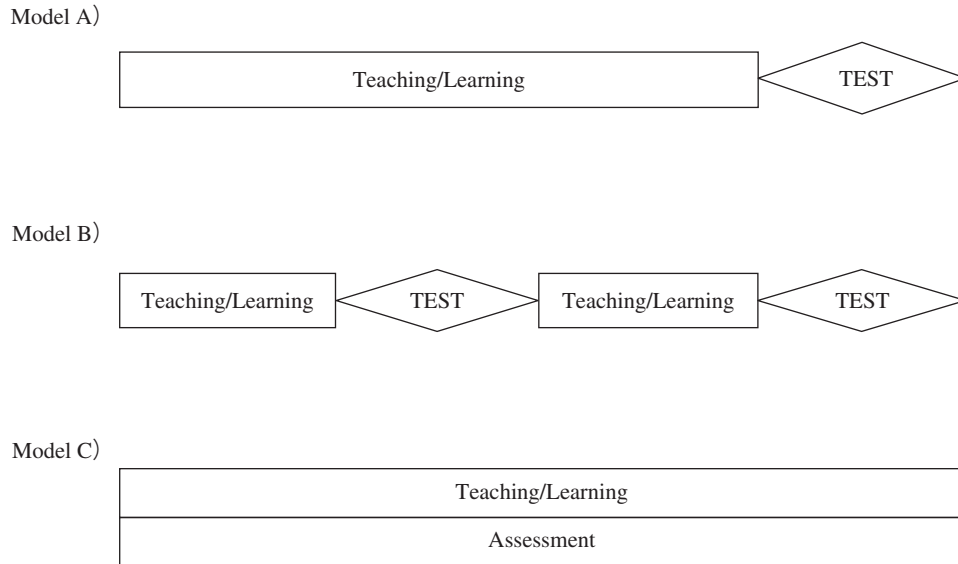
Researchers have also identified different types of assessments; Brown and Abeywickrama (2010) for example, have distinguished between informal and formal assessments (p. 6). Informal assessment can take place spontaneously in the form of “incidental, unplanned comments and responses” (p. 6) such as “Very good!” “Great!” and “Well done!” Formal assessment, on the other hand, is more planned, systematic, and content oriented. It is “designed to appraise achievement. Tests are always formal assessments” (Linville, 2011, Unit 1, p. 22). Another example is teachers’ comments by way of critiquing students’ writing.

Another way to distinguish assessment is by function, whether formative or summative. Formative assessment means “evaluating students in the process of ‘forming’ their competencies and skills with the goal of helping them to continue that growth process” (Brown & Abeywickrama, 2010, p. 7). It “helps students see their weak areas so they can learn more” (Linville, 2011, Unit 1, p. 22). It should be designed to provide learners with feedback about their progress while they are still studying a unit (Linville, 2011). A weekly quiz is a typical example of formative assessment. On the other hand, a summative assessment is conducted to “measure, or summarize, what a student has grasped” (Brown & Abeywickrama, 2010, p. 7). Thus, it usually takes place at the end of a course, where it is graded and taken into account in the overall evaluation.

## **Models of Assessments**

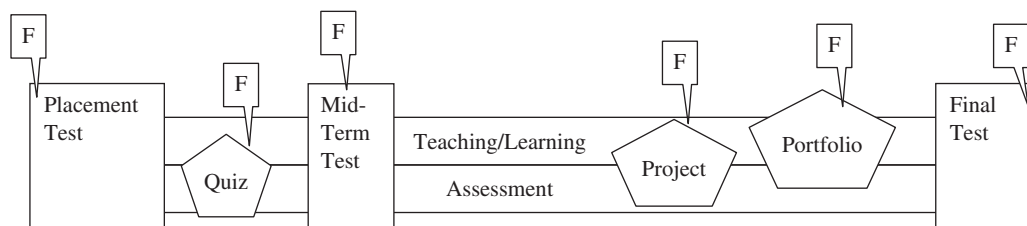
A further aspect to consider is the best time to conduct assessment during the course. Linville (2011 Unit 1 p. 15) has presented three assessment models and asks which of these models is the best. Model A shows students taking one test at the end of the course. The content of the test would cover what the students have learned during the course. Model B shows students taking two tests: mid-term and final. Each test covers the content that the

students will have learned preceding each test. Finally, Model C shows assessment as concurrent with the teaching and learning process; that is, it is conducted throughout the course. The three models are illustrated below as follows:



**Fig. 2.** Assessment Models (Linville, 2011, Unit 1 pp. 15-18)

The best method is Model C because assessment is most effective if conducted throughout the teaching and learning process. In addition, Linville (2011) went on to explain that many opportunities for feedback must be offered and that “teachers assess students’ progress and results at different times and in different ways” (Unit 1, p. 19). Linville has offered a sample assessment cycle seen in the following diagram (F in the following diagram represents feedback).



**Fig. 3.** A Sample Assessment Cycle (Linville, 2011, Unit 1 p. 20)

Thus, assessment has to continue while teaching and learning is going on, and feedback on the assessment must be given to students after each assessment is conducted.

One important facet of feedback timing is a practice known as “triangulation” (Brown & Abeywickrama, 2010, p. 321). Triangulation means using multiple measures for assessment. Brown and Abeywickrama state that teachers should assign “at least two (or more) performances and /or context before drawing a conclusion” (p. 159), and Linville recommends obtaining “at least three types of information” (2011, Unit 9, p. 5). Triangulation is recommended because teachers cannot judge the abilities of students by means of only one tool, as the likelihood is too high that some abilities may remain hidden (Brown & Abeywickrama, 2010, p. 321). Thus, according to Brown and Abeywickrama, one or more of the following designs should be included in the triangulation:

- several tests that are combined to form an assessment

- a single test with multiple test tasks to account for learning styles and performance variables
- in-class and extra-class graded work
- alternative forms of assessment (e.g., journal, portfolio, conference, observation, self-assessment, peer-assessment) (p. 159).

The measuring of multiple factors also increases reliability and validity, which will be discussed in the next section.

### **The Five Principles of Assessment**

Teachers need to consider five principles of language assessment when they create assessments (Brown & Abeywickrama, 2010, p. 25):

- Practicality
- Reliability
- Validity
- Authenticity
- Washback

These principles, which are all of equal importance, may be used to evaluate a designed assessment.

Practicality refers to evaluating the assessment according to cost, time needed, and usefulness. This principle is important for classroom teachers. Brown and Abeywickrama (2010) have explained the attributes of practical tests as follows: a practical test

- stays within budgetary limits
- can be completed by the test-taker within appropriate time constraints
- has clear directions for administration
- appropriately utilizes available human resources
- does not exceed available material resources
- considers the time and effort involved for both design and scoring

(Brown & Abeywickrama, 2010, p. 26).

Reliability means that the assessment is consistent and dependable (Brown & Abeywickrama, 2010, p. 19), which means that the same score will be achieved from the same type of students no matter when it is scored or who scores it. Brown and Abeywickrama (2010) have summarized the feature of this principle as follows: a reliable test

- is consistent in its conditions across two or more administrations
- gives clear directions for scoring/evaluation
- has uniform rubrics for scoring/evaluation
- lends itself to consistent application of those rubrics by the scorer
- contains items/tasks that are unambiguous to the test-taker (Brown & Abeywickrama, 2010, p. 27).

To make the test reliable, especially for subjective and open-ended assessments, it is important to write scoring procedures clearly and to train teachers to be able to score the assessment correctly (Linville, 2011, Unit 2, p. 11).

Validity has been explained by Brown and Abeywickrama (2010) as follows: a valid test

- measures exactly what it proposes to measure
- does not measure irrelevant or “contaminating” variables
- relies as much as possible on empirical evidence (performance)

- involves performance that samples the test's criterion (objective)
- offers useful, meaningful information about a test-taker's ability
- is supported by a theoretical rationale or argument (Brown & Abeywickrama, 2010, p. 30).

"If the assessment is supposed to measure students' speaking skill, for example, it should be designed to assess only speaking" (Linville, 2011, Unit 2, p. 12).

Authenticity is defined as "the degree of correspondence of the characteristics of a given language test task to the features of a target language task" (Bachman & Palmer, 1996, p. 23). An authentic test has the following attributes:

- contains language that is as natural as possible
- has items that are contextualized rather than isolated
- includes meaningful, relevant, interesting topics
- provides some thematic organization to items, such as through a story line or episode
- offers tasks that replicate real-world tasks (Brown & Abeywickrama, 2010, p. 37).

The final principle, washback, "refers to the effects the tests have on instruction in terms of how students prepare for the test" (Brown & Abeywickrama, 2010, p. 38). It refers to the outcomes for the learner, the teacher, and the teaching context (Linville, 2011, Unit 2, p. 15). Washback can be positive or negative. Thus, the concept of washback should comprise the following features: a test that provides beneficial washback

- positively influences what and how teachers teach
- positively influences what and how learners learn
- offers learners a chance to adequately prepare
- gives learners feedback that enhances their language development
- is more formative in nature than summative
- provides conditions for peak performance by the learner (Brown & Abeywickrama, 2010, p. 38).

As mentioned above, positive washback can motivate students to learn more; this positively influences teachers, thus further improving the classroom environment (Linville, 2011, Unit 2, p. 15). Teachers should ask themselves, after designing an assessment, if the assessment is practical, reliable, valid, authentic, and evocative of positive washback (Linville, 2011, Unit 2, p. 19).

## **Application of Principles to a Real Teaching Environment**

I will use these five principles to design an assessment for a course, Business English II, taught at the university level and in a real-life situation.

### **Course**

Business English II is aimed at helping fourth-year students acquire presentation skills in English for use in the international business scene. More specifically, in this course, the students are expected to achieve the following, according to the course syllabus:

- To obtain the vocabulary and phrases required for a presentation in English in business conferences
- To brush up the skill of conveying messages effectively to an audience
- To broaden knowledge related to business by learning presentations on different kinds of matters
- To acquire presentation skills in English to express ideas logically.

Since most of the learners had taken Business English I, as well as some other courses related to business English the previous year, they already had some knowledge of business English. Therefore, Business English II is meant to build up students' skills and to deepen their understanding of English for business purposes so that they can use these skills and knowledge immediately after starting work the following year. In addition, these skills and knowledge will be an advantage when searching for jobs.

This course is one of three from which students may choose to fulfill a requirement. Thus, enrollments are usually large—that is, about 160 students are enrolled in it. Four classes of this course are taught each semester, and each class is taught by a different instructor. I am teaching one of these classes this academic year.

#### Learners

The learners in this course are fourth-year EFL students majoring in English. Most have nine years of academic background in the English language. Moreover, most have spent one semester in college in the US to take EFL programs. On the basis of these experiences, the students' English proficiency is considered about intermediate, from low intermediate to high intermediate. Because learners are not placed into the class according to their achievements, slight differences exist in their English proficiency levels.

#### Teaching context

Although this course is taught in four separate classes by four different instructors, the same course syllabus is used in each class. This is a one-year course that extends over two semesters. Since students earn credit each semester, assessment is administered in each unit. In each semester, the students meet 15 times. On the 16th day, a final examination is administered.

#### Textbooks

The following two textbooks are used in this course: *English for Presentations* (Grussendorf, M) and *Biz Eitan 600 [600 Business English Words]* (Tsuji, K and Tsuji, S).

*English for Presentation* has six units; three units are covered each semester with one unit covered over about four classes. There are 26 units with about 20 sentences each in *Biz Eitan 600*, so 10 to 13 units are covered each semester with one unit per class.

#### The purpose of the assessment

The purpose of the assessment in this class is to monitor the skills achieved in terms of business presentations, as well as achievement in vocabulary. Because “assessment is an ongoing process that encompasses a wide range of methodological techniques” (Brown & Abeywickrama, 2010. p. 3), several approaches can be employed for assessment in this course. Instead of making students take a test at the end of the course, it is the most desirable to aim for “assessment [that] continues throughout the teaching and learning process, offering many opportunities for feedback” (Linville, 2011, Unit 1, p. 19), as mentioned earlier, and to “assess student progress and results at different times and in different ways” (Linville, 2011, Unit 1, p. 19). To continue assessment throughout the course, and also to apply triangulation, the following assessment tools are suggested.

1. Vocabulary quizzes
2. Project for research and presentation
3. Paper-and-pencil tests

For the vocabulary test, I will assign memorization of vocabulary and phrases from one unit in *Biz Eitan 600* per a week. At the very beginning of each class, I will hold a quiz on the vocabulary and phrases in the unit, and then

the students will exchange answer sheets with a partner for pair work to mark the answers. In this way, the students can check the words or phrases that are wrong, thus obtaining washback, one of the five language assessment principles.

For their project, the students will give presentations, possibly during the last few class periods, so that the last few days can be used for washback, which will consist of written instructor feedback, self-feedback, and peer feedback. As this course is for students seeking employment, they will be assigned to research a company in which they are interested, summarize their research results, make PowerPoint slides, and give a presentation to the class. They will present their research on the company as if it were a real-life situation. This assignment can measure not only speaking but also reading and writing skills. When classmates listen to a student's speech, they contribute written peer feedback, which also enables the assessment of the students' listening skills. Finally, they will be required to watch their own presentations, which will be recorded, thus having the opportunity to assess their own presentations objectively.

The tests are designed to assess the reading and listening skills students learn from the textbook, *English for Presentations*. Because assessment should be done throughout the teaching and learning process, the tests will be administered at least twice. Since this course covers three units in one semester, the best and most logical time to give a test is after each unit as a summative cum achievement test to see how much students have learned in each unit. The assessment tools for this class are presented in the following diagram (F represents feedback and Q, quiz).

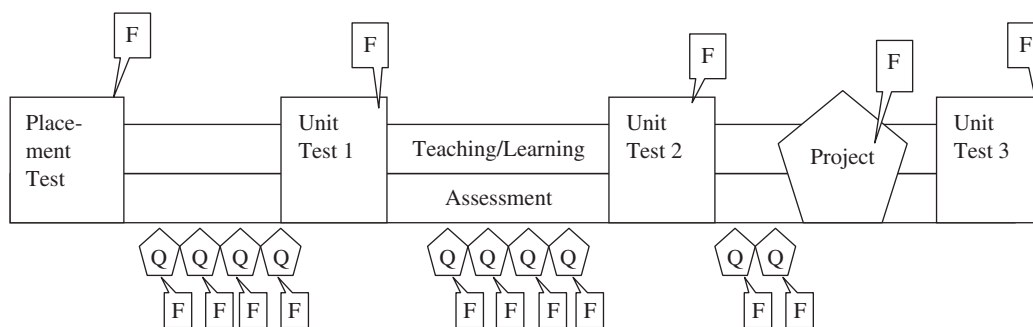


Fig. 4. Suggested Assessment Cycle

Students will be informed of the details of assessment content in writing at the beginning of the course by means of the syllabus. In particular, the students will need time to prepare for the presentation, so the project assignment will be given at the beginning with an instruction to start the research soon. Provision of a long lead time will verify that the assessment is reliable. The quizzes and unit tests are conventional in style and mostly objective. By way of contrast, the presentation includes subjective items. These make a mixture of the different assessment items.

### Summary and Conclusion

As this article has shown, teachers need to evaluate their assessment tools according to the five cardinal criteria for judging a test: practicality, reliability, validity, authenticity, and washback (Brown & Abeywickrama, 2010, p. 25). The tests, quizzes, and presentations proposed here are believed to be practical, considering the cost and administrative handling, as they are prepared by the instructor or the students. Though it takes time to design, prepare, and score the tests, the time requirement is manageable. To ensure reliability, the instructors have provided

clear administration and scoring instructions. Most of the test and quiz items will be objective, written so that they will not be ambiguous or confusing to the students. It is also relatively easy for the instructors and administration to provide a noise-free classroom with sufficient audio equipment. In terms of validity, the tests, quizzes, and presentation are valid because they are designed to measure what they need to: that is, how well the students listen to, read, and learn from the textbook. In addition, instructors will need to verify that all the questions on the tests have been introduced as activities in the classroom, that students know how they should tackle the questions, and that all questions are straightforward, not including any tricky ones. Once these items are verified, the tests and quizzes will be judged as fully valid.

The instructors need to verify whether the tests, presentations, and quizzes obviously incorporate authenticity, too; if they reflect natural use of the English language, especially with regard to business scenarios; and whether they are contextualized. The students will be able to use what they have learned in the workplace immediately after they begin their jobs in the following year. Finally, in terms of positive washback, quizzes will be checked during students' pair work, so they can verify the correct answers on the spot. For the presentations, feedback will be given by the instructor, by the students as peer feedback, and by the students themselves when they watch their own recorded presentations. Using these means, students can compare their own feedback to assessment from others. The instructor will also provide comments and feedback after each presentation. Each of these types of feedback will ensure and improve washback. As demonstrated, the assessment tools for this class will be designed according to the five language assessment principles.

The five principles of language assessment provide important guidance for designing and choosing assessment tools. Students should be clearly informed about which assessment tools will be used in class and should be assessed only on what they have learned in class. After teachers have used their assessment tools, it is important for them to evaluate the results so that they can improve their assessment tools for future use (Linville, 2011, Unit 9, p. 12). Linville also recommends that teachers "reflect on the assessment process" (p. 12) and have the students reflect as well. If employed in this way, assessment becomes a process parallel to teaching and learning, and it will surely improve the teaching environment. This paper aims to help readers deepen their understanding of the process of designing language assessments and applying them to real-life teaching situations.

## Acknowledgement

This paper is based on the EFL Assessment of the E-Teacher Scholarship Program, taught by the University of Maryland, Baltimore County. I would like to acknowledge the US Department of State, Bureau of Educational and Cultural Affairs, and the American Embassy in Japan for providing me the opportunity to study under this e-learning program and to share the teaching experience and pedagogical ideas with dedicated teachers all over the world.

## References

- Bachman, L. F. & Palmer, A. S. (1996). *Language testing in practice*. New York: Oxford University Press.
- Brown, H. D. & Abeywickrama, P. (2010). *Language assessment, principles and classroom practices* (2nd ed.). White Plains, NY: Pearson Education, Inc.
- Cameron, L. (2001). *Teaching languages to young learners*. Cambridge, UK: Cambridge University Press.
- Grussendorf, M. (2007). *English for presentations*. Oxford, England: Oxford University Press.
- Linville, H. (2011). *Assessment Principles II and Techniques for Assessment and Testing*. [PowerPoint slides]. Retrieved



## Designing Assessment Tools

from <http://blackboard.umbc.edu>

Tsuji, K. & Tsuji, S. (2012). *Biz Eitan 600 [600 Business English Words]*. Tokyo: MacMillan Languagehouse.