

# Assessment Literacy: Building a Base for Better Teaching and Learning

**A**re you assessment literate? What does that mean—to be assessment literate? Assessment is something that we as teachers must do all the time, but many of us feel unprepared or uncomfortable when it comes to testing our students. Teachers often reuse tests without analyzing or revising them and seldom use statistical procedures to see how a test—or a test item—is actually performing. Assessing students often means reaching for a test or quiz that is already prepared, whether it be a test included with a textbook, something another teacher prepared, or a standardized test produced by a major testing organization or our institution. These aren't necessarily bad choices (and sometimes it may not be our choice at all), but to make sure they are good choices, we must be knowledgeable about the principles and practices of assessment.

In order for assessment to be effective, classroom teachers need to be assessment literate—knowledge-

able about the key concepts of testing and how they can inform the design of assessments and decisions surrounding their usage. This article will start you on the path to being assessment literate. You will learn about the terms that make up the cornerstones of testing, how to plan your courses with assessments in mind, and how to make a test blueprint. Knowing more about assessment will not only help you to assess your students more effectively, but it will also provide you with a means of evaluating your own teaching and help you to produce tests that will actually motivate your students to learn. Let's begin by learning more about the words *testing* and *assessment*.

## Testing vs. assessing

The word *test* can make people nervous. It has semantic qualities that make us think of being judged or measured by someone or something. Many people have an emotive reaction to testing and associate it with negative experiences that they may have had as

students. In an educational context, the terms *testing* and *assessment* are often used interchangeably to indicate the measurement of student learning. However, although a test is a type of assessment—usually thought of in the traditional sense of an exam or quiz—assessment is a more comprehensive term. It often indicates the collection of information about student learning that might include not only tests but also a variety of techniques such as performance tasks, portfolios, and observation.

While tests are thought of as a means to give grades to students, assessments offer diagnostic information for both students and teachers. The ultimate purpose of assessment is to improve student learning, as opposed to just being able to give a mark for the amount of course content a student has mastered. Today teachers tend to talk about assessing (rather than testing) their students because we see the ongoing evaluation of student learning as more than just testing knowledge and skills in a particular area at one point in time for grading purposes. Thus, throughout this article, references to tests will be made with the ultimate goal of using them as assessment tools and not purely as testing instruments.

### **Importance of assessment**

As we all know, assessment plays an important role in teaching and learning. It affects decisions related to instruction, determines the extent to which instructional objectives are met, and provides information for administrative decisions. It has been estimated that teachers spend as much as 50 percent of their time in assessment-related activities (Stiggins 1991), and that when assessment is implemented effectively, student achievement is improved (Campbell and Collins 2007).

Yet many teachers feel assessment and testing are not relevant to their classroom practice and report that they feel unprepared to undertake assessment-related activities. Popham (2004) reports that most public school educators in the United States tend to think of assessment as “a complex, quantitative arena well beyond the comprehension of mere mortals” (82). Some of these feelings may come from the anxiety that teachers felt when they were students taking tests, especially if they didn’t understand how the tests were graded or if the objectives of the tests weren’t clear.

Teacher-education programs are also at fault for not making sure teachers are adequately trained before entering the classroom (Mertler 2004). As Taylor (2009) points out,

language education programs at graduate level typically devote little time or attention to assessment theory and practice, perhaps just a short (often optional) module; and although there is no shortage of books on language testing and assessment available today, many of these are perceived to be (and often are) highly technical or too specialized for language educators seeking to understand basic principles and practice in assessment. (23)

During our time in school and teacher-training courses, we take many tests, but how often are we actually given practice creating them, marking them, and interpreting the results? Developing these skills is part of becoming assessment literate.

### **Assessment literacy**

An essential element of assessment literacy is the ability to connect student assessment to the learning and teaching process. Teachers can make this link by first matching test items to instructional objectives, then using the test results to provide feedback on both student performance and how well the instructional objectives were met. An assessment-literate teacher is able to interpret data generated from a test to make useful modifications to teaching and to use assessments as a tool to improve student learning. Assessment-literate teachers are also able to discuss assessments with others in terms of key concepts in testing. With this in mind, we can explore common terms associated with tests, along with their practical application.

### **Key concepts and considerations**

Seven key concepts—usefulness, reliability, validity, practicality, washback, authenticity, and transparency—are cornerstones in testing that help to ensure that a test is solid (i.e., that it will consistently measure what you want it to measure in an efficient manner, and that both teacher and student will see it as a valuable source of information regarding learning). Understanding these concepts and being able to improve practices related to them are

important in developing assessment literacy. Each is discussed separately below, but as you will notice, they are connected to and support one another; together, they form the basis for building solid assessments.

**Usefulness and purpose**

According to Bachman and Palmer (1996), usefulness is the most important consideration when choosing or designing a test. Teachers must consider what the purpose of a particular assessment is and whether this purpose is congruent with the students they are testing and the course they are teaching. All language tests must be developed with a specific purpose, a particular group of test takers, and a specific language use in mind. Even tests with the general purpose of testing English language ability (proficiency) are designed with a specific group of test takers in mind. Take, for example, three standardized tests used globally for the purpose of measuring language ability: the Test of English as a Foreign Language (TOEFL), the International English Language Testing System (IELTS), and the Michigan English Test (MET). Each of these has been developed with very specific audiences and purposes (see Figure 1).

The examples in Figure 1 illustrate that tests are designed with very specific audiences and purposes in mind. This specificity is what allows them to effectively measure what they are designed to measure and makes them useful for a specific purpose. You must carefully consider the purpose of a test before

administering it. If you choose a pre-made test and it does not match your students’ needs or your purpose, then it will not be an adequate assessment of your students and will not provide the information that you need in order to make informed decisions about the teaching and learning taking place in the classroom.

For example, if you wanted to measure the reading ability of your students to see if they would be able to order from a menu when visiting the United States on an exchange trip, you couldn’t just use any reading test you find in a textbook or online. You would need to find one (or better yet, make one) that is specific to the skills taught in class, that meets the vocabulary needs of the situation the students would be immersed in, and that uses an appropriate text style that matches what you expect the students to encounter. Having them read a passage from a newspaper or a short story and then answer questions would not adequately measure their ability to read and order from a menu at a restaurant. So when you choose or design a test, consider the purpose of the test, the group of test takers it is designed for, and the specific language use you want to evaluate.

**Reliability**

Your assessments not only need to be useful for the intended purpose, they also need to be reliable. Reliability refers to the consistency of test scores. If you were to test a student more than once using the same test, the results should be the same, assuming that nothing

Test	Information about the Purpose
TOEFL	Measures the ability “to use and understand English at the university level,” and evaluates how well the test taker can “combine listening, reading, speaking and writing skills to perform academic tasks” (Educational Testing Service 2014).
IELTS	Has an academic and a general-training version. The academic version is for those who want to study in an English-speaking university; the general version focuses on basic “survival skills in broad social and workplace contexts” (IELTS 2013).
MET	“Intended for adults and adolescents at or above a secondary level of education who want to evaluate their general English language proficiency” in social, educational, and workplace contexts. It is “not an admissions test for students applying to universities and colleges in the United States, Canada, and the United Kingdom” (Modern Language Center 2010).

**Figure 1. Examples of standardized proficiency tests and their purposes**

else had changed. Reliability can be threatened by fluctuations in the learner, in scoring, or in test administration. Fluctuations in the learner are out of the testing administrator's control; we cannot control whether a student is sick, tired, or under emotional stress at the time of a test. But we as teachers can limit the fluctuations in scoring and test administration. The guidelines for how a test is administered, the length of time allotted to complete the test, and the conditions for testing should be established in advance and written in a test-specifications document. (See the Validity section for an example of test specifications.) As much as possible, there should be consistency in testing conditions and in how a test is administered each time it is given. Teachers can minimize fluctuations in score by preparing answer keys and scoring rubrics, and by holding norming sessions with those who will be scoring the test.

You can take steps to improve the reliability of your tests. You need to make sure that the test is long enough to sample the content that students are being tested on and that there is enough time for most of the students to finish taking the test. The items should not be too easy or too difficult, the questions should not be tricky or ambiguous, the directions should be clear, and the score range should be wide. Before you administer the test, you might want to have someone else take it to see whether he or she encounters problems with directions or content. Use that person's feedback to see where the test might need to be improved.

### Validity

One thing to keep in mind is that a test may be highly reliable, but not valid. That is, it might produce similar scores consistently, but that does not mean it is measuring what you would like it to. A test has validity when it measures what you want it to measure. The most important aspect of validity is the appropriateness for the context and the audience of the test. Think about what is to be gained by administering a test and how the information will be used. Suppose your goal is to measure students' listening ability, and you give a test in which students answer questions in written format about a lecture they hear. In that case, you need to make sure that the vocabulary, sentence structure, and grammar usage in the

written questions are not beyond the level of the students. Otherwise, you will be testing them on more than just their listening comprehension skills and thus decreasing the validity of the test as a measure for listening ability.

A number of factors can have an adverse effect on validity, including the following:

- unclear directions
- test items that ask students to perform at a skill level that is not part of the course objectives
- test items that are poorly written
- test length that doesn't allow for adequate sampling or coverage of content
- complexity and subjectivity of scoring that may inaccurately rank some students

The best way to ensure validity and reliability is to create test specifications and exam blueprints. These will help ensure that tests created and used match what is intended for the course and the students. Figure 2 shows an example of general information for the test specifications of a final exam for a higher-education pre-academic English-language program course. For each of the subtests (listening, reading, and writing), specifications would also be written and would include the type of skills being assessed, level of vocabulary, grammar structure, and length of text to be used.

### Practicality

Practicality refers to how "teacher friendly" a given test is. Practicality issues include the cost of test development and maintenance, time needed to administer and mark the test, ease of marking, availability of suitably trained markers, and administration logistics. If the test you want to give requires computers, and these are not available or connectivity is unreliable, there will obviously be a practicality issue with the delivery of the test. For many teachers, the amount of time required to mark a test is an important practicality issue. You can overcome this issue by weighing how important a particular assessment is in terms of overall course mark and determining how much time you want to spend marking it. For example, if a vocabulary quiz will not be worth much in the overall course mark, you might consider having students exchange papers and mark them instead of marking

<b>General Test Information — Final Exam</b>	
<b>Purpose</b> (Why are you testing?)	To test student mastery of listening, reading, and writing curricular objectives for Level 1
<b>Intended population</b> (Who are you testing?)	Students in university pre-academic intensive English program – Level 1
<b>Intended decisions/stakes</b> (How important is the test for the course grade?)	High stakes – weighted as 40% of the final grade
<b>Response format</b> (What type of questions will you use? How will the test taker show mastery of the objective?)	Listening: multiple choice, short answer, matching, gap fill, and information transfer
	Reading: multiple choice, short answer, matching, gap fill, and information transfer
	Writing: one-paragraph response to a prompt (input is a picture or personal knowledge)
<b>Number of examiners</b> (How many people are needed to administer the test? Are there any restrictions for test supervisors?)	One test supervisor per 20 students; two markers per exam (cannot be the class teacher)
<b>Number and weighting of items/tasks</b> (How many questions will there be on each part? How much will each part be worth for the overall grade of the test?)	Listening: approximately 20 items (33%) Reading: approximately 20 items (33%) Writing: 1 task (34 %)
<b>Examination length</b> (How much time will the assessment take overall? Is there a time length per section?)	Maximum of 2 hours Listening: 30 minutes Reading: 40 minutes Writing: 40 minutes
<b>Order of tasks</b> (In what order will the sections be tested?)	1. Listening 2. Reading 3. Writing
<b>Rating scale type</b> (Conditions necessary for marking the exam)	Reading and Listening: Answer key agreed to before the test; Writing: Two markers, analytical criteria, third marker if necessary
<b>Reporting type</b> (How will the score be reported? As a whole score, or per section? What is the passing grade?)	Single score (maximum 100%; pass mark 70%)

**Figure 2. Example of general test specifications**

each one yourself. This arrangement also allows students to review the materials at the same time. For marking writing, it might be more practical to have students review each other's work and peer edit the first draft than to have the teacher make comments on each initial draft.

### Washback

Washback refers to the effects of testing on students, teachers, and the overall program. It can be positive or negative. Positive washback occurs most often when testing and curriculum design are based on clear course outcomes that are known to all students and teachers. On



the other hand, exams that require extensive preparation can have negative washback and be harmful to the teaching and learning process; if instruction solely focuses on helping students pass the test, other learning activities may be neglected. To make sure washback is positive, teachers should link teaching and testing to instructional objectives. Tests should reflect the goals and objectives of the course along with the types of activities used to teach the content. That underscores the importance of planning assessments at the same time you plan the course.

Another way to bring about positive washback is through feedback. Providing feedback in a timely manner is important if you want students to learn and benefit from the assessment process. In the above example of practicality, having students mark their classmates' papers provides timely feedback to the students and helps them understand where they might need further practice or review. Using short quizzes that are graded immediately by the students throughout the course may let students know where they need to study more; it may also redirect teacher energy toward the areas that need more instruction time.

Involving students in the marking process is one way to create positive washback from testing. Other ways are to use authentic testing materials and to make the assessment procedures transparent—the topics of the next two sections.

### **Authenticity**

Tasks that reflect real-world situations and contexts in which the language will be used provide motivation for learners to perform well. Assessment tasks should be relevant to real-life contexts in which the language will be used. For example, if a course is designed for students who will be answering phones in English in a call center, an oral exam that mimics a telephone-call format would be more authentic than a test in which students listen to an academic lecture and respond to questions related to the lecture, or one where the students write the correct forms of verbs in sentence blanks. The assessments should relate to the purpose of the course, which in turn relates to course objectives, which are then tested on the assessments.

### **Transparency**

Transparency refers to the availability of information to students. Students should be

aware of the skills, vocabulary, and grammar that they will be expected to learn, and they should receive a clear explanation of how these will be assessed. Transparency makes students part of the testing process by ensuring that they understand what the course objectives are and what will be tested, as well as the format of tests and how they will be used and graded. Students should have the chance beforehand to practice question types that will be used in a test. Using a new test format, one that students are unfamiliar with, could affect the test's reliability. When students do not perform well on a test, it should be because they have not learned the material, not because they didn't understand the directions to complete a task.

Increasing transparency will also reduce students' test anxiety and allow them the chance to perform better. To increase transparency, many schools and educational institutions publish their test specifications. For example, the Oregon Department of Education publishes test specifications for the English Language Proficiency Assessments by grade level on its website (Oregon Department of Education 2014). These documents list not only content to be tested, but also in what ratio, along with appropriate test-item types.

### **Planning your assessments**

Now that you are familiar with the seven cornerstones of assessment, let's examine how you would go about planning an assessment that is useful, valid, reliable, practical, authentic, and transparent, and that has positive washback.

Planning your assessments goes hand in hand with developing your course learning objectives and should start when you begin planning the course. How you will assess student learning will affect how you present materials and teach the course. There are several phases in the assessment process. One of the most important is the initial planning stage. When you plan an exam, begin by describing your assessment context. Think about what the purpose of the course is, which resources you have available, and how the instructional setting and larger educational context influence the course. This is the information that you will put in the test specifications, discussed above in the Validity section, in the categories for purpose and intended population.

The next step is to identify students' needs and develop course learning objectives. Learning objectives are determined by what you want your students to know—and may be mandated by institutional or national priorities for education within your context. You should specify what you want your students to learn or be able to do after taking the course. This will guide you in developing not only lessons and curriculum, but also in deciding how you will assess whether students have learned what you want them to. Identifying course learning objectives will give you and the students goals to work toward during the course. Each of these objectives can then be divided into the skills needed to accomplish the objectives, whether they relate to vocabulary, structure, or fluency skills. With these learning objectives in hand, you will be able to design a test and check that the test you hope to use will accurately measure these objectives.

The best way to do this is to create a blueprint of the assessment, matching course objectives to the test questions. By using the course learning objectives to guide the content and the purpose of your exam, you can make sure that your assessments serve both as a tool for providing information about student

learning and as a means of assessing the course materials and instructional practices.

### Creating an exam blueprint

Having an exam blueprint increases the likelihood that you will actually test what you set out to test (i.e., the test will have validity). Test blueprints help you avoid overemphasizing one area or completely missing another area that needs to be tested. A blueprint is a tool to determine what is important for the students to know and the relative weight of each area in relation to other areas or skills being tested; at the same time, a blueprint ensures that the content being taught is properly represented on the test. The blueprint can also help a teacher see that the method used for assessing matches the cognitive demand that is intended.

Begin creating a blueprint by listing the learning objectives you want to measure, the way they will be tested, and how much of the total exam will cover each area. There may be several items on the exam related to each objective, but by first mapping out what you hope to test, you can be sure to include questions that assess all your objectives. An example of a simplified blueprint is given in Figure 3. It lists the skill area (in this case, the skills to be tested are reading, listening, writ-

Skill	Learning Objective	How Objective(s) Will Be Tested	Total
Reading	<ul style="list-style-type: none"> <li>• Can scan to find specific information</li> <li>• Can recognize main idea of a paragraph</li> <li>• Can understand pronoun references</li> </ul>	Read a paragraph and answer questions related to a reading passage using multiple-choice, short-answer, matching, gap-fill, and information-transfer items	30%
Listening	<ul style="list-style-type: none"> <li>• Can recognize main idea of a section</li> <li>• Can listen for specific information</li> <li>• Can listen for numbers</li> </ul>	Listen and answer questions using multiple-choice, short-answer, matching, gap-fill, and information-transfer items	30%
Writing	Can use pronouns to show cohesion between sentences	Write sentences related to a personal topic	15%
Grammar	<ul style="list-style-type: none"> <li>• Present simple</li> <li>• Adjectives</li> <li>• Subject/object pronouns</li> </ul>	Multiple-choice and fill-in-the-blank items	25%
<b>Total</b>			<b>100%</b>

Figure 3. Simplified exam blueprint

ing, and grammar), the learning objective, the item/question type, and overall percentage of importance in the context of this assessment.

You can also select a test that is already made and map it backward to see if it will fit your purposes or if items need to be added, adjusted, or replaced. To map backward, you would list each question, what it tests, and the number of points it is worth. At the end of this exercise, you should be able to see

what content is being tested and whether it is tested in the correct proportion to what you hope the students are learning. Figure 4 shows sample questions that might be on a reading test (in this case, the topic of the passage was dogs); the questions have been analyzed to determine the objectives being assessed and the mix of item types being used.

One way to develop items for a test is to write them on notecards (or if you have a

Item	Learning Objective	Item Type
What is the main idea of the reading? A. How to care for a dog B. The many different breeds of dogs C. The many ways that dogs are important to people	Reading: Can understand main idea	Multiple-choice question
In Paragraph 1, what does <i>they</i> refer to? A. veterinarians B. dog trainers C. dogs	Reading: Can understand pronoun references	Multiple-choice question
What is the meaning of the word <i>flush</i> in Paragraph 3? A. raise B. remove C. even with	Reading: Can understand vocabulary in context	Multiple-choice question
What is the main idea of Paragraph 3? A. Dog showing is a popular sport. B. There are several hundred breeds of dogs. C. Obedience training for dogs is important.	Reading: Can recognize main idea of a paragraph	Multiple-choice question
According to the reading, which is considered a sporting dog? A. collie B. fox terrier C. pointer	Reading: Can scan to find specific information	Multiple-choice question
Match the type of dog with its description, according to the information in the reading.  basset hound _____ poodle _____ Chihuahua _____ terrier _____  A. is trained to pull sleds B. has long ears C. is the smallest pure-bred dog D. sheds very little hair E. is used to herd animals	Reading: Can scan to find specific information	Matching

Figure 4. Analysis of exam questions relative to learning objectives



computer, add them to a spreadsheet) so that you can then sort them according to skill, question type, or objective covered. This system will let you know whether you have too many questions of the same type or need to add more in a certain category. Writing questions at the end of each lesson you teach is also an effective way to reflect on what you taught during the lesson and how you could assess that content in a manner the students will be familiar with. Working with a group of teachers to make new test items, or even working with your students, is a way to add variety to your item bank (your collection of possible test questions) and to get other opinions on what might be appropriate.

Having students become part of the assessment process can increase transparency, help them recognize what is important to know from the content being taught, and motivate them to review and study together. If students have a clear understanding of what they are being tested on and how they will be tested, positive washback is promoted. If you are working at an institution that uses very high-stakes testing and standardized exams for all students, you might consider having the students create assessments for review or quizzes. For example, you could ask each student to write questions for a particular grammar item or vocabulary list. This method will also help you informally assess student knowledge of the topics being taught as you review students' questions and answers to decide whether they are suitable for a small-scale assessment.

After questions are written, label each with the learning objective that it covers and the item type. You can then look at the questions, make selections, and put them together for the quiz. You can look at item cards yourself or with other teachers to determine the strength of each question for testing a given objective. Looking at questions with others is a useful practice that can also generate more items or improve the ones you have. You can make stacks of questions relating to the same curriculum objective and determine how many of each you will need for the appropriate weighting on the test or quiz. Once the assessment has been administered, you will also want to add information to each question about how it performed (see Basic Statistics for Testing, below). These testing or item statistics will be useful in determining whether

to use the questions again and which learning objectives might need to be reviewed or taught in a different manner in future lessons. In this way, an assessment can provide both teacher and student feedback.

### **Providing feedback**

An important process in assessment development and revision is feedback. By providing feedback on an actual test after administering it, teachers ensure constant improvement with regard to the reliability and validity of all assessments administered within a program. Information gathered from the test can be useful to determine final marks of an exam and to decide whether any questions should not be counted in the grading process. All information that is generated through testing should be kept with that test and reviewed before the test is administered again. Teachers who use an assessment should be asked to complete a feedback form. Sample questions and the areas you might consider for feedback are given in the following section.

### **Creating an exam file**

Reflecting on the exam is an important part of the test development process. After you administer each test that might be reused, create an exam file. The file should contain information about the actual tasks in the test, its administration, and its level of difficulty. You could create a feedback form that has questions related to these topics, then fill it out when developing and reviewing the test, and reflect on it after the administration. Questions that might be useful in gathering feedback on an exam are listed below.

#### **Setting: Task and Administration**

- Is it clear how the student is expected to respond to each item?
- Does the exam test the curriculum content?
- Does the exam contain formats familiar to the students?
- Are the exam tasks authentic and meaningful?
- Is there more than one logical option/answer possible for any blank/question?
- Is the layout user-friendly? Does the formatting match specifications?
- Are there any typos, misprints, or other errors in the production of the test?

- Are the listening and reading texts at an appropriate level?
- Are the texts clear in their organization and content?
- Is the quality (e.g., pace, sound clarity, voice quality) of the recording good?
- Is the time allotment appropriate for the length of the exam?

**Demands: Task**

- Do the test tasks adequately sample the skills and strategies needed for [listening/reading/writing]?
- Are any questions answerable by students who haven't heard the lecture or read the text?
- Do the task types reflect student classroom experience?
- Are the task types adequate for measuring the course objectives?
- Is this test an accurate measuring tool for progress for this course?

Along with the feedback form, an exam file would also include the test-specifications document indicating the purpose and general audience for the test, the test blueprint indicating how each question matches a learning objective, and basic statistics. Changes to an exam or scoring process should be based on the analyses and feedback. Making notes about the test immediately after grading it offers an opportunity to update an exam while it is fresh in your mind. You might even think about having your students answer a few questions about the test as well. These could include how difficult they thought it was, whether there were questions they found confusing, and whether the content covered was what they expected would be on the test. Valuable information about your assessment can also be gathered through basic statistical analysis.

**Basic statistics for testing**

After you administer an assessment, basic statistics give you an idea of how your students are performing and information about the test in general. Two useful statistics to measure are the mean score and the pass/fail rate. The mean is the mathematical average of all the test takers' scores; you can compute the mean by adding all the test takers' scores and dividing the total by the number of students

who took the test. You can compute a mean score for the whole test and for each section. The mean score provides information about the reliability of a test over time, especially if the delivery of the course remains the same. If a test is reliable, you would expect a similar result among students at different times or between groups of similar students. If you have a test in which various skills are tested, such as the one in the example of the test specifications with a listening, reading, and writing section (Figure 2), computing the mean for each section would give you an idea of which skill areas the students have more difficulty with, as indicated by a lower mean score.

To determine the pass/fail rate, meanwhile, you must first decide what a passing score is on the exam. I have been in places where 60 percent is passing at the university level for some courses, depending on their difficulty, and other places where 75 percent or even higher represents the passing score. It depends on the level at which students are expected to master the course objectives, and on whether the purpose of testing is just to measure learning objectives or to move only the top performers on to the next course or level of study. Once you have determined a passing score, add the number of students who reached the passing score or higher and divide by the total number of test takers. This number represents the pass rate. The same can be done for the fail rate, counting instead the number of students who did not receive at least the passing mark and dividing by the total number of test takers.

The pass/fail rate can be calculated for the test as a whole, for each section or skill area, and even for individual questions. Looking at individual-item pass rates gives you an idea of the difficulty of each question and can help you find items that may need to be rewritten or areas that will need further instruction or review. If you add this information to the questions that you created on notecards or a computer spreadsheet, in the future you can also put together tests with questions that have a variety of difficulty levels. This information will also tell you which skills you might not need to spend so much time reviewing in the future and which ones may need more instruction time.

### Practice makes perfect

Just as learning a language takes practice, so does developing assessment skills. By thinking about the assessment process and analyzing the tests we use, we can all become better teachers and test developers. Teachers often report that assessment literacy is not something that they acquired by taking a course, but instead is something they learned on the job through the ongoing experience of working with and developing assessments. When I asked teachers how they had gained the knowledge that they have about assessment, almost all mentioned post-formal-education experience (“through discussion of the effectiveness of assessments with colleagues,” “trial and error,” “being part of a group making tests,” “on-the-job experience”). Teachers seem to feel that assessment literacy is not something you are taught in a course, though you can get the fundamentals and theories there; it is basically something that must be developed through ongoing experience and practice. Assessment literacy is a skill that you have to learn by doing, not just by reading about it. Real-life teaching in the classroom provides the perfect opportunity to apply assessment fundamentals and to develop assessments that will motivate your students to learn—and that will motivate you to develop lessons that will facilitate learning.

A lot of research has focused on what information about assessment should be taught in education courses, but maybe instead we should be looking at the opportunities for development that present themselves throughout our teaching careers. As Boyles (2006) suggests, teachers can use the information from assessment to adjust teaching practices, to provide evidence of student performance, and to guide the curriculum-review process. The information we gather from assessments should provide feedback to students, but it should also be used as a means for teachers and program administrators to review instructional practices and curricular objectives. Brookhart (2001) notes that in order for teachers to “own the goal of high-quality assessment, they must be convinced that assessment is just as important to student learning as ... effective classroom management or lesson design.” Research indicates

that “teacher professional development is more effective when it is school embedded, cooperative and sustained over time ... organized within and/or across schools and focuses on improving practice over time through the sharing of knowledge, experience and expertise” (O’Leary 2008, 112).

So, with that in mind, make sure to share with your colleagues what you have learned from this article (or share this article with them), talk about your assessments, work together to improve them, and remember that assessment plays an important role in the teaching and learning processes. By reflecting on the tests used in your classroom, how they are developed, and the results obtained from them, you will become more assessment literate and a better teacher. Testing will not seem so complex or beyond comprehension, as many teachers feel it is. You, in turn, will provide students with a better experience, empowering them as learners instead of instilling fear of tests. With assessment literacy comes the understanding of the valuable role assessment plays in the classroom for both teachers and students.

### References

- Bachman, L. F., and A. S. Palmer. 1996. *Language testing in practice*. Oxford: Oxford University Press.
- Boyles, P. 2006. Assessment literacy. In *New visions in action: National assessment summit papers*, ed. M. Rosenbusch, 18–23. Ames: Iowa State University.
- Brookhart, S. M. 2001. The standards and classroom assessment research. Paper presented at the Annual Meeting of the American Association of Colleges for Teacher Education.
- Campbell, C., and V. L. Collins. 2007. Identifying essential topics in general and special education introductory assessment textbooks. *Educational Measurement: Issues and Practice* 26 (1): 9–18.
- Educational Testing Service. 2014. “About the TOEFL iBT® Test.” [www.ets.org/toefl/ibt/about?WT.ac=toeflhome\\_ibtabout2\\_121127](http://www.ets.org/toefl/ibt/about?WT.ac=toeflhome_ibtabout2_121127)
- IELTS. 2013. “Test Takers—Academic or General Training?” [www.ielts.org/test\\_takers\\_information/what\\_is\\_ielts/academic\\_or\\_general\\_training.aspx](http://www.ielts.org/test_takers_information/what_is_ielts/academic_or_general_training.aspx)
- Mertler, C. A. 2004. Secondary teachers’ assessment literacy: Does classroom experience make a difference? *American Secondary Education* 33 (1): 49–64.
- Modern Language Center. 2010. “MET: Michigan English Test.” [www.mlc.edu/jo/met-michigan-english-test](http://www.mlc.edu/jo/met-michigan-english-test)
- O’Leary, M. 2008. Towards an agenda for profes-

- sional development in assessment. *Journal of In-service Education* 34 (1): 109–114.
- Oregon Department of Education. 2014. “Testing—Student Assessment.” [www.ode.state.or.us/search/page/?id=496](http://www.ode.state.or.us/search/page/?id=496)
- Popham, W. J. 2004. Why assessment illiteracy is professional suicide. *Educational Leadership* 62 (1): 82–83.
- Stiggins, R. J. 1991. Relevant classroom assessment training for teachers. *Educational Measurement: Issues and Practice* 10 (1): 7–12.
- Taylor, L. 2009. Developing assessment literacy. *Annual Review of Applied Linguistics* 29: 21–36.

---

**DAWN ROGIER** is the Regional English Language Officer for Brunei, Malaysia, the Pacific Islands, Papua New Guinea, and the Philippines. In her free time, she enjoys analyzing test items and discussing statistical procedures.