

# Language Assessment in Real Classrooms: Why We're Writing This Book

Assessing students' language growth and needs is challenging, and it is even more challenging for busy teachers with conflicting responsibilities. Assessment can seem like a time-wasting, pointless, and even unethical endeavor, resulting in discouraged students and teachers. There is certainly no reason to assess students if you, the teacher, do not understand the purpose of the assessment and what you are going to do with the results. Therefore, we begin this book at the start of the assessment process by asking the question: Why assess? Of equal importance is the answer to the question: What will I do with the results?

## Activity 1.1. Reasons to Assess

Working on your own or with a partner, make a list of some reasons to assess students' language skills and explain what you will do with the results.

### Reasons to assess students' language skills

---

---

---

---

### Explain what you will do with the results

---

---

---

---

Assessment, in fact, is a vital component of good teaching. We as teachers need to know how individual students are progressing on specific aspects of our curriculum. This helps us to both give specific individualized attention to struggling (or high-achieving) students and to discover how certain methods or materials are working in the classroom as a whole.

## Definition of Some Terms

**Evaluation:** The process of evaluating evidence in order to come to a conclusion. Multiple sources of evidence, including test scores, can be used in an evaluation.

**Assessment:** A more or less formal procedure for gathering evidence. A test is one type of assessment. An observation report is another.

**Test:** A type of assessment in which the procedures are carefully controlled so as to be repeatable.

This information can be at a micro level, such as when we learn which sound–spelling correspondences an individual child is having difficulty with, or at a more macro level, such as when we learn about what math-related vocabulary a group of English language learners (ELLs) will need extra help with in order to learn the math content.

This book examines classroom assessment (not standardized assessment) through the lenses of three ESL teachers from across the United States. The teachers in the book face a number of challenges that are particular to their individual situations, as well as other challenges that all public school teachers face. We examine the decisions they make about assessments in the areas of literacy, oral language, and content-specific language development. We hope this book helps you and your students as you plan and implement assessments.

## Introducing Our Teachers

Throughout the book we use three realistic (but not real) teachers' situations to illustrate the various problems and choices that language assessment provides. We have chosen to illustrate a wide range of language teaching situations, from young to high school learners, beginners, or newcomers, to more advanced ELLs. The classrooms we examine are mixed ELL and non-ELL content (science, social studies, etc.) classrooms, bilingual and stand-alone ESL classrooms. In one of our example situations (Beth's class), the students all share the same first language, and in the other two example situations, the teachers work with a mix of first languages in their classrooms. If you are currently a practicing classroom teacher, many elements of these contexts should sound quite familiar. These teachers and the assessments they design are examples of what we consider to be best practice. In other words, they are models to inspire you and to illustrate concepts, but not necessarily to be copied. We know that each situation is different and that you need to consider your own students and classes in your own context. Their challenges, regardless of the extent to which their classrooms are similar to or different from yours, likely mirror many of your own. For all teachers, time and resources are in scarce supply, and we acknowledge this through our examples.



### MEET BETH

Beth is a second-grade teacher in New York's Washington Heights area in the northernmost part of Manhattan. Reflecting the composition of the neighborhood, most of Beth's 25 students come from the Dominican Republic. Beth has been teaching for 10 years in Grades 1–4 and has been teaching second grade for the past 4 years. As in the past, this year some of her students are designated as ELLs, and some are not. Most of them, however, speak Spanish at home regardless of their official designation as ELL or non-ELL. At least two of the ELLs seem to have had no prior formal educational experiences and may not be literate at all in Spanish or English. As the school year begins, learning more about these two students' proficiencies and needs are among Beth's first priorities.

As a mainstream classroom teacher with TESOL specialty certification, Beth is expected to provide both content and second language instruction as needed to all of her students. Her most important assessment needs (among many) are tracking students' progress in their English literacy development and finding out how language-related scaffolding will aid the ELLs' progress in mathematics and other content areas.

We use examples from Beth's classroom in all of the core chapters of this book—checking for understanding, social versus academic language, assessing oral language, assessing literacy, and content-area language assessment.



### MEET TOM

Tom is an ESL specialist in a large middle school in Bellevue, Washington, a suburb of Seattle. He teaches stand-alone ESL classes (pull-out) and also supports two seventh-grade science teachers with majority ELLs (push-in). The ELLs, reflecting the diversity of the area, come from a variety of countries, including Mexico, Guatemala, Somalia, Vietnam, and South Korea. Budget cuts have increased class sizes at Tom's school, and the science classes he supports each have more than 30 students, adding to the challenge. His most critical uses for language assessment in the stand-alone classes are checking progress in both social and academic language for the beginning-level students and deciding in which aspects of the science curriculum students will need the most language support. Tom also needs to plan, teach with, and support two science teachers, one of whom has no background in teaching ELLs, language learning, or teaching diverse populations.



### MEET ALICIA

Alicia teaches high school social studies in Levelland, Texas, an agricultural center in the western part of the state. The school population is now majority ELL, with many children from migrant agricultural families. Most of her students have ties to Mexico and speak Spanish as their primary language, even though many of them were born in Texas. Consistent with No Child Left Behind (NCLB) regulations, she has been working with school administrators, staff, and fellow teachers to increase students' access to the full curriculum. She conducts some stand-alone ESL classes for newcomers and migrant students, and works especially closely with two math teachers—her mentor from when she started teaching and a colleague with whom she began her teaching career. In addition to teaching social studies, Alicia works with the staff and administration at her high school to help understand the language issues of these students and how these issues may have an impact on state-mandated standardized testing.

## What Kind of Assessment Is This Book About?

All of these teachers need the kind of rich, detailed information that well-designed **formative assessment** provides. By formative assessment, we refer to assessments that provide detailed, specific information (**informative**) about students' capabilities and struggles that will help teachers choose and modify lessons and supplemental materials (Hughes, 2003; Stoyhoff & Chappelle, 2005). Formative assessment also provides important information to parents and students themselves. It is important to keep in mind that the main purpose of assessment for a classroom teacher is information gathering—not motivating or punishing students, or practicing for standardized testing. To be effective teachers, we need to know what students already know and can do, what they need to know and how they learn best, or how to teach them in the most effective manner.

We want to know, for example, which science vocabulary words Tom (the Seattle-based seventh-grade ESL specialist) should put on a review handout next week for the ELLs for whom he provides language support. To supplement a science-specific vocabulary review, he could provide some instruction on “academic” linking words such as *furthermore*, *however*, and *therefore*. Tom could gather the information he needs to make this decision in a few ways. For example, he could do the following:

- informally but systematically check students’ knowledge of the words through questioning or conversation
- follow up if students look confused or use a word incorrectly
- administer a short quiz to students
- look for evidence of students’ understanding of these words in their recent class work

Depending on the time available, Tom’s familiarity with the students’ recent progress, and the type of work they are doing, any of these approaches could be useful and all of them can be called formative assessment. The major purpose of this book is to provide language teachers with formative assessment techniques that they can put to immediate use in the classroom.

In contrast, we use the term **summative assessment** to refer to assessments designed to provide an overall, “big picture” view of students’ progress (sum up) in order to make decisions, about issues such as exiting from ELL status, or to provide information about schoolwide progress in a core subject (Stoyhoff & Chappelle, 2005). State-mandated standardized assessments are generally administered once per year and are the summative assessment results teachers see most often. These types of tests are not designed to guide classroom instruction at a micro level, but can provide useful information about students’ general proficiency at a point in time. Most of this book focuses on designing good formative assessments and using them in real classrooms.

## The Process of Assessment or Evaluation

All educational assessment, from the most informal checking for understanding to the most high-stakes state assessments, can and probably should be designed in much the same way. In the case of large-scale assessments, these stages may be documented and detailed in technical reports, and extensive research and development work will be carried out, involving dozens of professionals. In the case of a quick quiz or an obser-

vation of student work, the same principles will apply but may never be written down or given more than a few minutes of thought. However, the process itself is fundamentally similar.

The first stage in assessment is to determine the purpose of the assessment—how it will be used and what decisions you will make based on the results (Bachman & Palmer, 2010). The most difficult part of this decision is deciding and describing exactly what you want to know about students. Simply stating “I want to know how well my students are reading” is not nearly specific enough to be useful in most cases. It is almost always more useful to link your assess-

### Jargon Sidebar

Testing professionals like to talk about **tasks** and **items**. What’s the difference, and why does it matter? Well, *item* is just another word for a single test question (e.g., multiple choice, short answer, true/false). However, for many purposes, asking several individual questions is not the best way to find out about students’ abilities. Therefore, we use test tasks—the language “performance” we ask the students to do, such as writing an essay or a lab report, or giving an oral presentation. A task often involves multiple skills, such as reading a short text and then summarizing it in writing or orally. We are used to seeing *items* in tests of listening and reading, and *tasks* in tests of speaking and writing.

ment to specific standards or curricular objectives. You can focus back on what you have been working on most recently or forward on the new material you are about to begin.

How does this work in real-life classrooms? We have already mentioned that Beth has two new students in her second-grade class this year. They both seem to lack the basic literacy background that students usually attain in first grade. These students arrived in Beth's classroom in the second week of school, without formal records of previous U.S. schooling. In accordance with her state's policy, the students were administered a standardized literacy assessment. In New York, this assessment is the LAB-R. The results of the test are not specific enough to help Beth plan her curriculum with the students' needs in mind. If they have not developed first-grade literacy skills, Beth needs to know about it right away in order to modify, or differentiate, her lessons for these two students and bring in additional support staff from the school to help out in the classroom. In this case, her **purpose** is to find out first what these students' literacy levels actually are so that she can make decisions about their instruction, such as to what extent she should differentiate what they are asked to do in the classroom. Beth defines her **construct** (see sidebar) as "these students' abilities with letter and sight word recognition, sound-spelling correspondences, and ability to form letters and common words in English." Initially, she needs only a general idea about their reading levels. Once she has established this, she will need more specific information, and the construct will then become more specific.

Now you can reflect on your own experience with assessment by completing Activity 1.2.

### Jargon Sidebar

Psychologists use the term **construct** (n.) to mean the cognitive or behavioral attribute they are interested in learning about. The meaning of a construct must be carefully described so that it is clear to teachers and students. For example, "reading skill" is not specific enough to be useful. "The ability to decode words with beginning blends" (the *bl* in *blend* is an example of a beginning blend) is a more useful construct definition and is directly relevant to what a teacher is trying to do. Writing construct definitions requires a deep knowledge of the subject matter and is similar in many ways to the creation of lesson objectives and curricular goals.

#### Activity 1.2. Reflection on Assessment Quality

Think back to a recent assessment you took as a student.

What was good or bad specifically about the assessment?

Did you think the results matched your actual abilities?

If yes, why? If no, why not?

Compare your answer with a partner.

Now that you have thought about an assessment you have experienced, let's discuss how such assessments are written. The next stage in assessment is **task design**. In this stage, we need to decide how we are going to learn about the construct Beth has stated: "these students' abilities with letter and sight word recognition, sound-spelling correspondences, and ability to form letters and common words in English." A good task is something that requires a student to show what he or she knows and can do within a given construct, along with instructions and context. The task should be designed to help Beth find out what she needs to know about students. Obviously, there is no one "right" answer for how Beth should design this

assessment; there are many ways to design an effective assessment task. Often, an effective assessment task will be very similar to an effective teaching activity. The assessment allows the teacher to find out what the students learned from the activity.

Because Beth is concerned about the students' basic reading and writing abilities, she chooses two tasks that she has used before in the classroom: a read-aloud task from a story book intended for beginning-level readers and a simple copying task whereby students copy some words from the story book. Beth decides to administer these tasks to the two newcomers together while the rest of the class is engaged in silent reading.

She chooses a story and words to copy that will take less than 10–15 minutes, which is an appropriate length of time for a task for students of this age. Beth is certain that these tasks require the use of the abilities defined in the construct definition above. The greater the students' abilities with "letter and sight word recognition," the better they will be able to perform the task of reading aloud and then copying some words from the story. The coordination between construct and task is critical to the concept of **assessment validity** (see sidebar); in order for results to be meaningful, the task should match the abilities defined in the construct (Bachman, 1990; Bachman & Palmer, 2010; Hughes, 2003).

### Jargon Sidebar

**Validity** is generally defined as the overall usefulness, or suitability, of an assessment for a particular purpose. A test is valid if it is indeed a useful tool for making some decision or gathering some information. One of the most critical aspects of validity is what is often called **construct validity**, or the degree to which a test actually measures its construct (the abilities it is intended to measure). Some test tasks do a better job than others of measuring a particular construct, and seemingly trivial attributes of tasks (e.g., whether or not they are timed) can have a big influence on what abilities they actually tap into.

### Activity 1.3. Matching Tasks and Constructs

For each construct listed in the first column draw a line to a task that could measure it.

#### Construct

Knowledge of fourth-grade science vocabulary  
 Ability to explain a process orally  
 Understanding the main point of a text  
 Understanding classroom instructions

#### Tasks

Scanning a short newspaper article for the main point  
 Match vocabulary words to pictures  
 Stand up when the teacher tells you to  
 Give directions on how to set up a science experiment

After you complete this activity, reflect on your own or with a partner about other tasks that could measure each of the four constructs.

**Extension:** Write out some constructs that you might assess as a teacher in the grade level you are interested in. Be as clear and concrete as possible.

In Activity 1.3, several possible language assessment constructs are defined. Working alone or with a partner, identify at least one task that you think would "tap into," or require the use of, those abilities.

## Defining Criteria

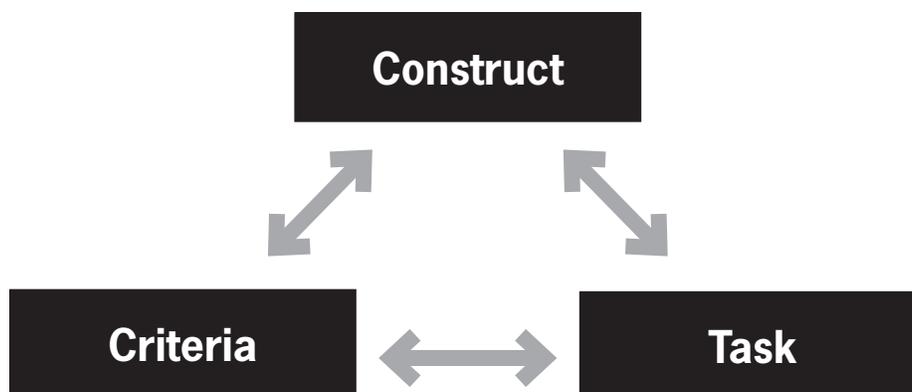
Another part of developing a task is identifying the criteria by which you will measure the task outcomes (Bachman, 1990; J. D. Brown, 1996). Sometimes this is easy, because there is only one possible correct answer, for example, in the case of a multiple-choice question. However, the more realistic a task is, the more complicated it is to develop criteria that capture what students have done. How do we decide on these criteria? This is one of the hardest aspects of assessment, because we have to be explicit about what we expect students to do. Is getting 8 out of 10 words correct on a spelling quiz excellent, adequate, or actually substandard? Of course, the answer has to take into account a large number of variables. Answering a question like this requires knowledge of students, the curriculum, and grade-level expectations, which these days are enshrined in state and national **standards**. In Beth's case, her evaluation came from her thorough understanding of grade-level literacy expectations, including knowledge of the developmental steps students go through in the early stages of learning to read and write. Because she knew little about these new students, her criteria could not be very specific.

We've been talking about task development as if it is a linear process. However, the reality, of course, is that it is more iterative than linear. As you fine-tune the construct and the task, keep these questions in mind:

- How you will know if students have been successful in completing the task?
- What information do you want to receive from student task performances?
- What information will you give students about what they have gained and how they can improve?
- What will you do with the information?

Figure 1.1 shows the relationship among construct(s), task, and criteria for assessing how students perform on the task.

As Figure 1.1 shows, the construct, criteria, and task are interrelated, and if the definition of one changes in any way, the others must also reflect this change. This means that if you amend your criteria, you also need to change the task and examine the construct to make sure that you are still measuring that construct. Now let's look at how, based on the construct she has identified, Beth develops a task and criteria for rating it.



**Figure 1.1. Relationship Between Construct, Task, and Criteria**

## Task 1: Student Read-aloud

Beth has decided to develop the read-aloud task we described earlier. She starts by taking a Level 1 reader from her bookcase. She chooses the book *Dogs and Frogs*, which begins with one word per page and builds up to short sentences by the end. All concepts in the book are illustrated. Beth decides to write down her instructions so that she can be clear and consistent with students and use the same instructions in the future.

**Construct:** Beth has identified the construct as defining students' abilities with letter and sight word recognition and sound-spelling correspondences. She decides that she will give this task to each student individually. She will open the book, point to words, and ask them to read the words. Then she will evaluate how well they completed the task.

**Evaluation:** Beth decides that there are three basic levels, which we have used to create the following rubric. We know that Beth might not write them out as specifically as we have here. For this book, in the second grade, grade-level expectations would be three smileys.

☺	☺☺	☺☺☺
Student can recognize some letters and sight-read one or two words.	Student reads half or more of the words without prompting.	Student reads all words with confidence.

**Results:** After Beth administers this task to the students, she finds that both are at the one-smiley level. Next, Beth wants to find out how well they can copy words. She chooses six words from the book and asks the students to copy these words.

## Task 2: Copying Words

**Construct:** Beth has defined the construct as “ability to form letters and common words in English.” After the students finish the read-aloud task, Beth gives them each five words to copy: two that they could read and three that they could not. These are the words she chooses: frog, dog, log, more, in.

**Evaluation:**

☺	☺☺	☺☺☺	☺☺☺☺
Student cannot hold pencil and does not know which way to hold the paper.	Student can copy a few letters of half of the words.	Student can copy most of the letters from some of the words.	Student can copy almost all words legibly.

**Results:** After Beth administers this task to the students, she finds that one is at the one-smiley level and one is at the two-smiley level.

**Implementation:** What happens after Beth administers these short tasks to the two students, and how does it make a difference? Beth concludes that these students are not yet meeting grade-level expectations.

They will need significant additional help beyond the regular second-grade curriculum. Using these results to document the students' needs, she requests an aide from the principal for these two students. Further, now that she knows more about their reading levels, she begins to differentiate her regular instructional materials for them in order to maximize their access to the classroom content.

**Reflection:** This example shows how a busy and caring teacher can incorporate a quick but well-designed task to accomplish specific goals. With a clear construct, a realistic task, and specific criteria, Beth was able to create a useful assessment that better helped her understand these students' needs.

In the remainder of the book, we describe similar assessment challenges and responses on the parts of our three teachers. We discuss different approaches to assessment and different types of tasks that the teachers develop, implement, and use to improve teaching. Chapter 3 is devoted to informal checking for understanding, which can serve as a midpoint between continued instruction and more formal assessment. Chapter 4 examines different ways to assess oral language, in terms of both comprehension and production. Chapter 5 is devoted to literacy skills, including reading and writing, at lower grades, whereas Chapter 6 examines literacy skills in students in the higher grades. Chapter 7 addresses content-specific language skills, such as the language component of math, science, and social studies coursework. Chapter 8 discusses how to describe, report, and communicate the meanings of assessment results to students, parents, administrators, and other stakeholders. Chapter 9 discusses the role of large-scale assessments in driving teaching and learning in the post-NCLB era.

Each chapter includes several activities for you to complete as well as a summary activity at the end. We designed these activities to help you reflect on assessment and think about how to implement assessment principles in your own classroom. Each chapter also draws on the experience of our fictional teachers—Beth, Tom, and Alicia—and presents concrete examples of ways in which teachers may choose to confront the realities of assessment in K–12 classrooms.

## End-of-Chapter Activity

Find a language (ESL) textbook in your school or in the library. Look for an assessment—usually they can be found at the end of the chapters. Describe the assessment in the following ways:

- What is the construct? It is probably not stated anywhere, so you will need to infer it.
- What is the task? In other words, what exactly are the students asked to do?
- What are the criteria? How will the teacher evaluate students' success in using this assessment? Again, this information may not be stated directly, so you will need to infer it.