M Y T H

2

A comprehensive final exam is the best way to evaluate students.

In the Real World . . .

In college, I took German for three years, earning very average grades. (This is Lia.) As a reward for my persistence, I studied in Austria for a semester to experience German-speaking culture and "authentic" language. During that semester, I was greatly surprised to learn that my German was more than adequate for basic communication and that I could spend hours in conversation using only German with my host mother. By the last month of my stay, shopkeepers no longer answered my questions in English and they conversed with me in German. As I was departing Austria, I was able to arrange, in German, for my suitcase to precede me on a train to Belgium. I was delighted to find it waiting at the station when I arrived.

I had more success using my German when traveling abroad than I'd expected based on my grades in college classes. This seemed like a mismatch because I felt my classes prepared me well in authentic German; I'd spent four days a week in class talking about topics such as family, school, and travel. Although the first two years of coursework followed a communicative syllabus, 75 percent of my grade was based

23

Assessment Myths
Applying Second Language Research to Classroom Teaching
Lia Plakans and Atta Gebril
http://www.press.umich.edu/5056216/assessment_myths
Michigan ELT, 2015
24   Assessment Myths

on an end-of-semester final. This exam was strictly grammar-based, administered to all sections of German at the same time, and graded uniformly. It was given in a large, crowded auditorium. I was elbow-to-elbow with other students, and all of us were filling in bubble sheets. The conclusion that I came to about my language skills was that my German was okay, but my grades reflected that I was bad at tests. In-class assessments were occasional quizzes modeled on the final exam. This disconnect was a problem for us as students in understanding the goals of the course and in interpreting our grades. Without regular assessment parallel to class activities, we lost the opportunity to gauge and guide our language learning. If I had not studied abroad, I would have simply assumed that I had never really learned German.

What the Research Says . . .

Since the 1960s, a distinction has been made between two kinds of assessment: formative and summative (Shavelson et al., 2008). **Summative assessments** evaluate cumulative knowledge or abilities usually at the end of a learning sequence, such as final comprehensive exams. **Formative assessments** are frequent checkpoints embedded in the context of the course that evaluate learning both formally and informally. Their purpose is to help learners and teachers check progress and also to steer them toward meeting course goals. Formative and summative assessments are not unrelated and exist on a continuum rather than being strictly dichotomous (Black, Harrison, Lee, Marshall, & Williams, 2003; Shavelson et al., 2008; Stiggins, 2005)—for example, a midterm exam could be used for both formative and summative purposes.

Ideally, the two kinds of assessments share an underlying model of language and learning, called a **construct** (see page 10 for a definition). Both should be related to course goals and objectives as well as the same vision of language learning. In fact, summative and formative assessments that inform each other are the most effective and expedient. Figure 2.1 lists examples of each type of assessment. While the ten-

FIGURE 2.1: Examples of Formative and Summative Assessment in Language Learning

Formative ————————➤	Summative
role plays/dialogues	final exam
journals/quickwrites	exit tests
reading group discussions	standardized proficiency interviews
comprehension checks	midterm exams
learner "can-do" checklists	portfolios
graphic organizers	
midterm exams	
portfolios	

dency is to judge one as more valuable than the other, the best practice is for both to have a place in education and to be aligned.

Formative Assessment

Formative assessment has several unique characteristics. First, it is a frequent, ongoing part of instruction, even when it takes an abbreviated form. Second, it informs both teachers and students of progress being made. This progress can be captured for individual learners as well as subgroups and whole classes. Formative assessment identifies gaps between what students know and can do and what they need to know and be able to do. Finding and filling this gap is an important step in learner-centered instruction. The goal for formative assessment is to improve learning, rather than to judge individual achievement. If students understand this characteristic of formative assessment, they may be less anxious and perform closer to their actual ability. Last, due to the possibility of individualizing formative assessment and its outcomes, this kind of assessment is adaptive and allows for differentiation, something particularly critical in classrooms where students often have a wide range of language abilities.

Shavelson et al. (2008) provide a scale for types of formative assessment. On one end are "on-the-fly" assessments, which include such teachable moments as when a teacher overhears or notices stu-

dent interactions that reveal a gap in learning. At the other end of the scale is **embedded assessment**, or structured inventories aligning specific learning goals with curricular objectives. Between these two extremes are planned-for interaction and exchanges targeting certain learning objectives, such as a series of questions posed during class discussions. Considering assessment at all these levels opens up possibilities for continually collecting information about students' learning and for addressing gaps before the end of the term/program when it is too late to remedy them.

Black et al. (2003) reviewed four recent developments in formative assessment practices that can be useful for teachers to consider. The first is attending to classroom dialogue as a means to guide assessment and give feedback. In this case, the teacher listens in on group work to observe the students' language use and to notice where and why communication breakdowns occur. Another area of increasing attention in formative assessment is peer and self-assessment, which engage students in the evaluation of their learning. A third area, specifically in writing, is the use of "comment-only" marking or writing conferences (one-on-one discussions of writing), which emphasize feedback in the absence of a grade. It removes judgment and focuses on the students' improvement. The fourth development described by Black et al. (2003) is the use of summative tests in formative ways. This approach entails including some aspects of final or standardized tests in classroom assessment but with lower stakes and a different emphasis. Teachers can provide learners with test items or tasks and then use them to discuss model performances or engage in other guided activities.

Classroom-Based Assessment for Language Teaching

Scholars have discussed the difference between assessment **of** learning and assessment **for** learning (Colby-Kelly & Turner, 2007; Leung, 2007). Substituting the preposition **for** distinguishes assessment embedded in the learning processes from assessment focused outside of it. A final exam that evaluates the knowledge students have absorbed by the end of the coursework would be an assessment **of** learning.

Contrast that with this example of assessment *for* learning: Students are assigned to write the questions for a chapter test that requires them to think about what they should have learned from the material. Then they critique the questions in class and develop an answer key. These activities would focus students on learning goals and their progress toward them. A third variant on this theme is assessment *as* learning (Earl & Katz, 2006), which sees assessment as encouraging students' metacognitive thinking about their learning. It creates an atmosphere where students are active in thinking about and evaluating their learning to make decisions based on their progress.

A similar pairing is between **dynamic** assessment and **static** assessment. Again, a final exam would exemplify **static** assessment; once teachers have developed or adopted a final exam, the test will remain the same when all students take it. A **dynamic** assessment could be a test, too, but one that includes interaction between the teacher and students. During such assessment, the teacher provides guidance to help a student proceed. It is not just about correct or incorrect answers but about how a student engages when support, or scaffolding, comes from the teacher, who models the process of learning.

From these two dichotomies come two approaches related to assessment in language learning classrooms: **Assessment for Learning** (AfL) and **Dynamic Assessment** (DA). Neither approach is unique to language teaching; each has a history in the field of education. Yet over the past decade, national ministries of education, school districts, curriculum developers, and researchers have increased the amount of attention both of these approaches have received.

ASSESSMENT FOR LEARNING (AfL)

The Assessment Reform Group (2002) in the U.K. emphasized the goal of AfL as "the process of seeking and interpreting evidence for use by learners and their teachers to decide where learners are in their learning, where they need to go, and how best to get there." AfL views assessment as a tool rather than simply a measure of learning. It focuses on the relationship across teaching, learning, and assessment—

in particular, how assessment can be used to inform both teaching and learning (Lee, 2007b).

Process writing is an example of AfL that has been widely accepted in second language teaching. In this approach, students work through multiple drafts of their writing projects and receive feedback at each stage. Peer review is often incorporated to provide additional feedback. While serving as reviewers of each other's writing, students develop critical judgment and awareness of the criteria for good writing. An example of AfL from a second grade ESL classroom is the use of signal cards to check comprehension: Each child has three cards colored red, green, or yellow on his or her desk. Periodically during a lesson, the teacher asks the learners about their understanding of the lesson, and they each hold up a card showing that they do not understand (red), need more time (yellow), or are ready to move on (green). This show-ing of cards gives the teacher a quick and visible way to check in with learners; in addition, it does not require a great deal of language, which can facilitate communication for young learners and those at a begin-ning level of English proficiency.

A number of research studies have looked at AfL in second lan-guage classrooms (Colby-Kelly & Turner, 2007; Ellery, 2008; Lee, 2007a, 2007b). Based on this research, formative assessment and AfL appear to be recognized and valued by teachers in general. Yet when practice and perceptions are thoroughly investigated, classroom assess-ment still seems teacher-directed and overshadowed by summative feedback. Thus, the potential learning opportunity that AfL affords is shortchanged.

Colby-Kelly and Turner (2007) studied a pre-university English for Academic Purposes (EAP) program that had integrated AfL into its cur-riculum. Specifically, the study focused on the assessment of speaking performances and feedback. Researchers collected data from nine teachers and 42 students and found substantial use of AfL by teachers, particularly in feedback to students. Although teachers reported strong support for this type of assessment, the researchers found that feedback was most often between the teacher and an individual student, rather than between the teacher and groups or among peers. Such feedback

was given privately and discreetly, as if it were something negative that might embarrass students. Teachers were also hesitant in their views regarding the usefulness of peer and self-assessment. Despite having a clear plan for AfL, the teachers' interpretation of feedback seemed at times to contrast with the spirit of AfL, which is to help learners see the benefits of others' input as a natural part of learning. The message from assessment needs to be on learning how to learn, not simply on being judged.

As previously mentioned, second language writing classrooms that have adopted a process approach to teaching writing are already using AfL (Ellery, 2008). In a research study, Lee (2007b) investigated the function of feedback in AfL. She collected feedback from 26 teachers on more than 174 writing assignments from secondary classrooms and then interviewed both students and teachers. She used six characteristics of AfL to analyze the teachers' feedback and found that their comments tended to focus more on summative assessment and error correction. Such negative feedback led to generalizations with little concrete meaning to students. In interviews with these teachers, they explained this tendency as a response to substantial pressure in a highly exam-focused educational system. This top-down pressure included emphasis on accuracy in written products, which was also conveyed by the teachers in their feedback. Student involvement in assessment was minimal; even more distressing, the teachers admitted that their feedback was probably not that useful to students. Lee (2007b) emphasizes the need to seek the overlap between exams and classroom assessment in AfL; in the contexts she studied, the two remained separate, and as a result, high-stakes exams seemed to drive the teaching and assessment in the classroom in an unproductive manner.

Assessment for Learning is a detailed plan to integrate formative assessment and involve learners in the learning process. It entails a commitment from teachers to guide students to become self-assessors, which requires practice, modeling, and time. Ideally, AfL is adopted from the beginning of a learning sequence or grade level, so that it is not just one class or one teacher trying to make this significant change

in how assessment is perceived and what roles teacher and student perform in this endeavor.

DYNAMIC ASSESSMENT

Dynamic Assessment (DA) came to the field of education from the concept in sociocultural theory. The Zone of Proximal Development (ZPD), an idea attributed to Lev Vygotsky, maintains that a novice learner will progress better with the guidance of an expert than he or she will alone. Lantolf and Poehner have brought the idea of DA to the field of language learning through a number of theoretical and practice-focused articles (Lantolf & Poehner, 2007; Poehner, 2007, 2008, 2009; Poehner & Lantolf, 2005).

Dynamic assessment allows us to gauge a learner's ability to respond to assistance by a teacher or someone guiding the learner. Assessment and mediation are fully integrated, which gives the learner support to complete a learning unit. The teacher's intervention is part of the assessment, and DA is used to evaluate the student's response. This information allows us to measure how learners will learn, as well as how they can perform the task after receiving some guidance. Another critical aspect of DA is how well learners are able to transfer what they gained from the intervention to other situations. DA, then, is not a single-shot assessment but is continuous and iterative throughout the term.

In DA, the teacher guides the students with leading questions to help structure their thinking about a problem or task. Articles by Poehner provide many examples of DA in foreign language classrooms when it is expanded to groups and the whole class (Poehner, 2009; Poehner & Compernolle, 2011). DA can be illustrated by a teacher-to-student interaction. This example is from a writing conference, Lia observed. Notice how mediation is used to guide the language learner.

S: [Reading essay aloud] The person I admire is no here. She is my mother.

T: Is *no here* important to know? I mean, what is most important in this opening sentence? The person or the location?

S: No here. She's not here. She is in away.

T: What is most important to know first, *who* you admire or *where* that person is?

S: Who. My mother.

T: So could you put that in the first sentence?

S: The person I admire is my mother. She is no here.

T: What about *no here*? Is *no* the right form?

S: No, *not*. . . . She is not here.

T: Great! Let's move forward.

Oskoz (2005) investigated the use of DA to evaluate synchronous computer-mediated communication (SCMC) in foreign language classrooms, such as in online chatting. She explains that SCMC is a process-oriented learning activity, and so assessment used with such interaction should be focused on process, rather than just the product, making DA a natural fit. In her study of five university-level Spanish classes, students were recorded interacting through online activities, such as information gaps and jigsaw puzzles. These interactions were analyzed using a five-level scale adapted from previous research to evaluate learners' attention to and correction of errors. Using this scale, Oskoz considered how the SCMC transcripts indicate a learner's level in terms of self-error correction as well as responses to a partner's guidance in recognizing and selecting errors. In her conclusion, Oskoz establishes that DA has the potential to provide rich information for evaluating students' interlanguage development through SCMC. This finding suggests that DA is well suited to certain kinds of tasks or classroom activities, particularly those that emphasize process.

Kozulin and Garb (2004) conducted a DA study following a testing-learning-testing format in a language program in Israel with 13 English-as-a-third-language learners from Ethiopia. These students started with a pre-test session using six items from a standardized placement test. A mediation session followed in which the pre-test material was discussed in relation to the students' pre-existing knowledge and what strategies they would need to answer the test items. A post-test had similar items with different content from the pre-test. Not

surprisingly, these learners did better on the post-test once the mediation had occurred but, more important, the researchers could observe which learners benefited most from the mediation. Analyzing the change in scores from pre- to post-test provided information on a student's learning potential and his or her individual learning needs. Although Kozulin and Garb (2004) admit that DA may be more time consuming to carry out than static testing, with mediation incorporated into classroom learning, this difference may disappear. They see DA as a critical process in understanding students' learning needs and highlight the need to prepare teachers as mediators. This aspect of DA is an important one to consider because the mediation depends on teachers providing useful input and feedback rather than teachers simply just correcting students or giving them answers.

Advocates of the DA approach believe that it allows for a more nuanced understanding of learners' development. It is less structured in terms of beginning and ending measures with more focus on the interactions that occur throughout the intervention. While we are including DA in a chapter with formative assessment, DA proponents claim that there is a distinction between these two approaches: DA should always be systematic, and although formative assessment is often planned, it can also be more spontaneous. DA is directed at long-term development, which highlights its similarities to AfL. DA and AfL embody different approaches to assessing students even though they both promote learning that shifts away from the traditional comprehensive final exam. They offer teachers insight into learners' progress and create new ways to evaluate and individualize instruction. AfL brings students into the assessment process to promote lifelong learning skills. DA brings assessment and teacher scaffolding into close proximity. In both approaches, assessment is ongoing and iterative throughout a term. Implementing DA or AfL should be done with curricular goals in mind and with deliberate consideration of alignment with summative assessment as well as other standardized testing used to evaluate learners.

What We Can Do . . .

Research shows that classroom-based assessment provides valuable information about learning and development to both teachers and learners, making it an insightful part of language teaching. How can we make it part of daily classroom activities?

1. Create an ongoing cycle of assessment, instruction, and course goals.

Planning should promote continuity between summative and formative assessment and also provide useful checkpoints on students' progress throughout the term. It is better to know by Week 5 that students are not able to write complex sentences or that they have not grasped the past perfect tense than to make such a discovery at the end of the term when there is no time to remedy it. Best practices in formative assessment require planning at the outset and continuous monitoring as the course progresses. Figure 2.2 illustrates this process. Notice the double arrows on the right side of the figure, moving back and forth between weekly course objectives, instructional planning, and formative assessment. These three activities support and inform each other. With summative assessment, the process tends to be more linear.

FIGURE 2.2: Assessment in Curriculum Planning
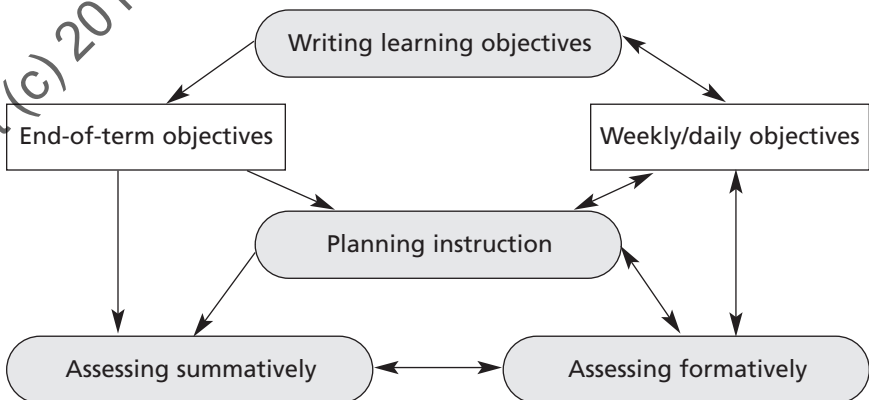
As part of developing their classes, teachers should incorporate an assessment plan. This plan should mesh with the daily lesson plans of a course, which are aligned with the overall curriculum of the course. An assessment plan could take many forms, but it must be a useful reference for the instructor. Aspects to include in the plan are: instructional objectives, assessments and activities, and procedures for scoring and providing feedback (see an example in Table 2.1).

The plan in Table 2.1 illustrates the alignment between assessment and instruction. In the example, the teacher has distinguished between activities, assessments, and feedback/scoring; marked those assessments that will be scored for grading; and included three uses for the results. The level of detail and categories depends on the teacher's needs and style. However, the plan should be followed closely through-

TABLE 2.1: An Assessment Plan for a Second Language Writing Class

Instructional Objective	Activities (A)	Assessments (Aa)	Scoring & Feedback	Decisions
Developing summarization skills	A1. Whole class model of summary writing A2. Pairs practice summary writing A3. Whole class critiques pairs' summaries	Aa1. Student highlights key ideas in reading Aa2. Student writes a summary in a template Aa3. Student writes an independent summary	Whole class builds rubric from critique of summary A3 Students use rubric to self-assess Aa2 & Aa3 Teacher uses rubric to score Aa3	A2 & A3 used to decide if students are ready for assessment Aa3 Results of self-assessment and Aa3 used to determine if more summary work is needed Score from Aa3 used for students' final grade

out the term, remembering that it is dynamic—changes can and should be made—but that such changes should be recorded in the plan.

In addition, such a plan can help teachers think through the instructional decisions that can be based on the assessments and eventually plan how to use the outcomes. Assessment can inform us of the progress of individual students and the class as a whole, as well as help us to decide how we teach a certain lesson in the future. Including these decisions in the assessment plan will help teachers remember all the information they have gleaned.

2. Emphasize the feedback aspect of assessment.

The research discussed shows the value of feedback in formative assessment, AfL, and DA. While a grade or a number communicates some information to learners and teachers, learning occurs when clear explanations and suggestions for improvement are provided. Giving useful feedback is a skill that teachers should cultivate and monitor. Finding ways to make comments comprehensible entails knowing what wording will be understood by language learners (especially second language learners). Having the class discuss grading criteria or rubrics is a way to increase the value of feedback later on. Giving examples of "good" performances can also guide learners. Activities that allow students practice using the assignment criteria can make expectations clearer. Some other approaches to emphasizing feedback include:

- Having one-on-one conferences with students to discuss their assessment performances. If students can lead the conversation in conferences, teachers might discover issues they hadn't considered, such as misunderstood expectations or confusion over language forms or functions.
- Requiring revisions based on feedback. Selective revisions require students to focus on one area for improvement, such as word variety when writing or fluency in speaking. Giving students an opportunity to

use the teacher's feedback for improvement comes
through second chances. It also allows teachers to see
how well students can use the feedback (mediation) to
learn.

- Checking with students to learn whether the feedback
being given is clear. Using jargon or abstract terminol-
ogy can make feedback confusing and potentially
useless. Examples of vague terms include: *awkward
wording, conciseness,* or *misplaced modifier*. Give students
an example and make sure the language used in feed-
back has been covered in class.
- Giving a new grade or revised grade based on improve-
ments from feedback. While a long-term goal is to take
emphasis off grades, students still care and see grades
as indicators of what a teacher values. Rewarding
attention to feedback with grading can increase, in
students' eyes, the value of the feedback and the learn-
ing that is gained from following it.

Since these feedback practices may be more difficult to employ when
class sizes are large, teachers must look for efficient strategies. Using a
rubric that students are familiar with is one approach (this will be dis-
cussed in detail in Myth 3). Another is to incorporate peer and self-
assessment after students have been trained to give helpful feedback.
Although a good deal of work is required to establish this approach at
the beginning of a course, it will pay off later on as students can
become sources of feedback.

3. Involve students in the assessment process.

Students need time to learn which criteria are used in assessment and
how to assess their own or others' work. Sharing models or examples is
an effective method to train students on peer and self-assessment. It is
also critical, especially initially, to help students see the value in this
practice because some students may believe that the only important

feedback comes from the teacher. One strategy from a writing classroom is to ask students to provide feedback to each other on early drafts. Then, as they rewrite, ask them to reflect on what input they used from peers in their own writing and how it changed their final paper.

In addition to peer and self-assessment, students can be asked to reflect on their learning throughout the course. While they may not be accustomed to thinking about it or articulating it, this can be cultivated through regular practice. Self-assessment can be as simple as the example of students holding up a red, yellow, or green card, indicating whether they are ready to move to a new topic. Another example could be asking students, as they prepare to hand in a major assignment, to write answers to three questions: What did you learn from this assignment? What would you like to practice more in the next assignment? What do you hope to learn as you complete the next assignment? This reflection can be a 15-minute quickwrite or even just some words or phrases to answer each question; whatever the method, it may provide useful information to the teacher.

Another way to involve students is to include them in the development of assessments and rubrics. Working in groups, students can create possible questions for a midterm and, later, the whole class can critique them. In designing and critiquing questions, students need to review important material from the course, consider how to elicit knowledge and abilities related to this content, and deliberate on what "correct" answers might be. This process provides many learning opportunities and results in a collaborative exam.

Another approach is to ask students to design a rubric for a performance assessment (see Myth 3 for more on performance assessment). Given examples of strong and weak performances, students in small groups discuss and place these performances in various categories. This technique could result in concrete criteria for evaluation. The teacher can monitor and mediate the process of collecting student-generated criteria. Issues can be highlighted, such as weighting criteria in a rubric (what is most important?) and discussing criteria that are difficult to measure. The activity culminates in the use of the rubric to evaluate students' assessment performances.

Assessment Myths
Applying Second Language Research to Classroom Teaching
Lia Plakans and Atta Gebril
http://www.press.umich.edu/5056216/assessment_myths
Michigan ELT, 2015
38 Assessment Myths

In all of these examples, it is important to realize that the first time students participate in assessment development, they may find it quite challenging and the results will not be perfect. Even for professional language testers, revision and reflection are critical parts of the assessment process. After a class-generated test is taken or a rubric used, students should review its success and learn from aspects that did not work as expected. Then the activities can be repeated, giving students an understanding of the evolution of their learning.