

# 1

## The Process of Developing Assessment

---

We have seen that assessment covers a range of activities from everyday observation of students' performance in class to large-scale standardized exams. Some teachers will be involved in a full range of assessment activities, while others will mainly be responsible for producing informal assessments for their own classes. However, at one time or another, almost all teachers are consumers of tests prepared by other people, so regardless of their personal involvement in actually developing assessment, teachers can benefit from understanding the processes involved. This chapter provides a guide to the assessment development process.

Assessment includes the phases of planning, development, administration, analysis, feedback, and reflection. Depending on teaching load and other professional responsibilities, a teacher can be working in several different phases at any one time. Let's look at how this applies in the case of Ms. Wright, an assessment leader in her high school.

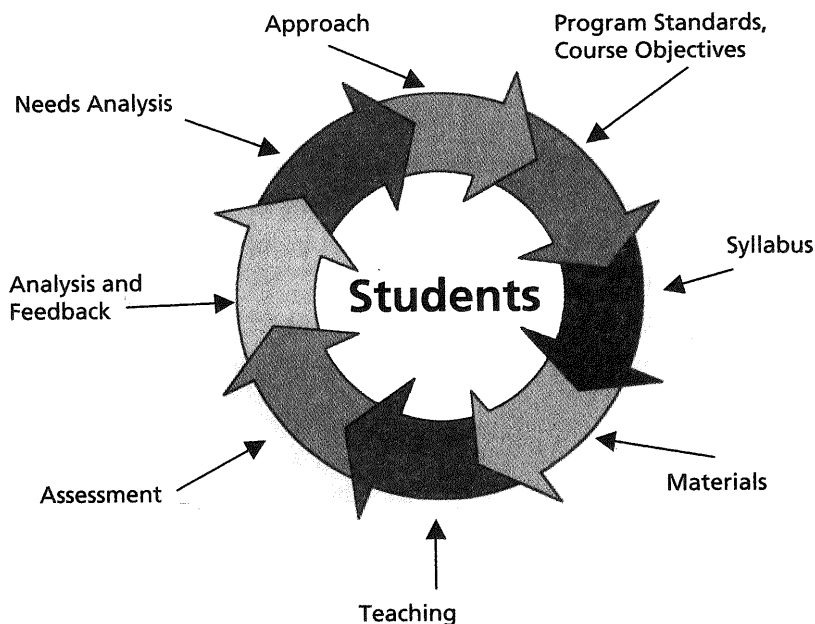
If we were to visit Ms. Wright in early November, halfway through the fall semester, we would learn that she had already taken these steps toward assessment of her students:

- started planning in August by doing an inventory of her Grade 12 course, ensuring that outcomes closely matched assessment specifications
- met with her colleagues to develop a schedule of different types of assessment spaced throughout the academic year
- ensured that all stakeholders (students, parents, colleagues, administration) had information about when assessments

occur, what they entail, and how much each assessment is worth

- administered and analyzed diagnostic exams to her classes in September and adjusted her instruction to the needs of her students
- revisited previous midterm and final exams to review results and select items for recycling based on item analysis conducted after the last test administration
- asked colleagues to prepare new test items well in advance of exams to allow time for editing
- organized workshops on speaking and writing to ensure inter-rater reliability
- blocked out time to conduct a preliminary analysis soon after the midterm exam
- scheduled a meeting with administrators to discuss midterm results

**Figure 1: Assessment in the Teaching/Learning Cycle**



Assessment is an integral part of the entire curriculum cycle, not something tacked on as an afterthought to teaching. Therefore, decisions about how to assess students must be considered from the very beginning of curriculum design or course planning. Once a needs analysis has established the goals and approach for an English program, *standards* are developed that define the overall aims for a particular level of instruction. These standards are then converted to more specific *course objectives* or *outcomes* that state what a student can be expected to achieve or accomplish in a particular course. It is important that the outcomes are worded in terms of actual student performance because they form the basis for the development of assessment *specifications*, which are the planning documents or "recipes" for particular assessments such as tests and projects.

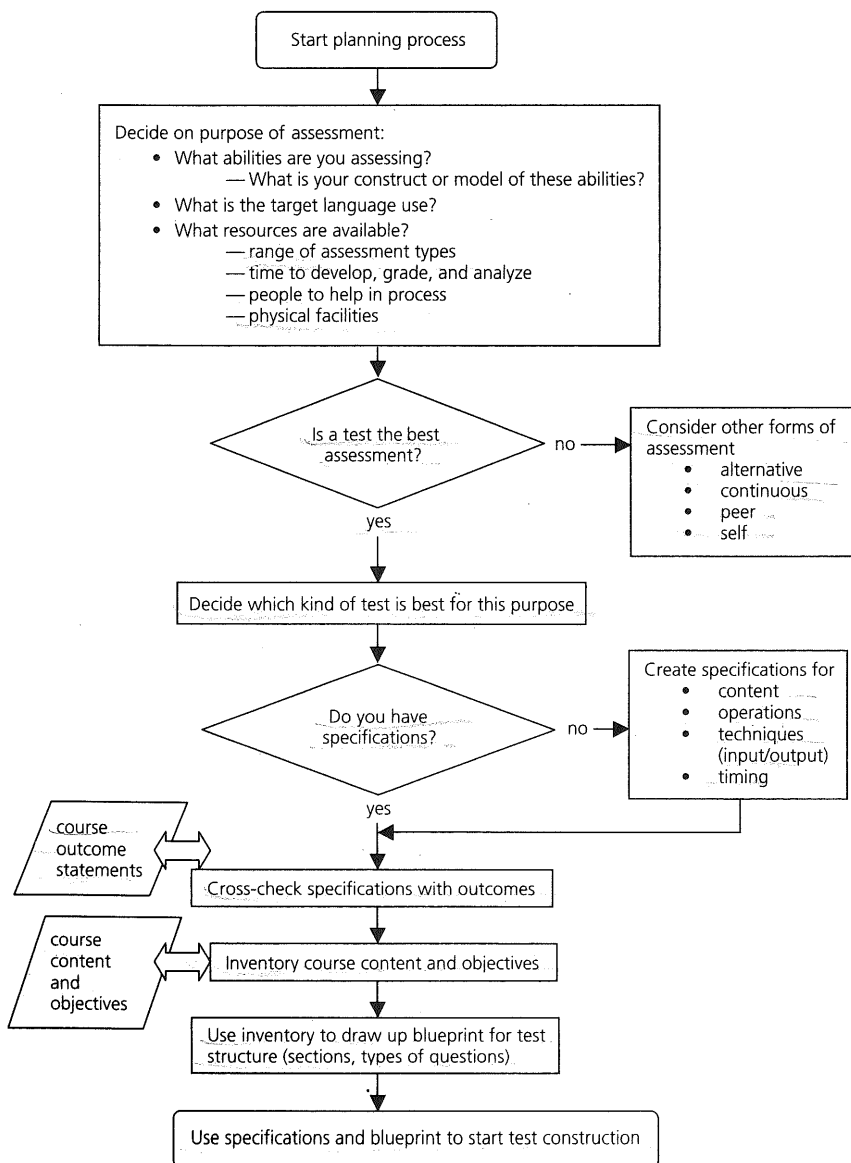
An outcome such as "Students will study present tenses" is too vague to be transformed into a test specification. If the outcomes are restated as "Students will use the simple present to describe facts, routines, and states of being" and "Students will use the present continuous (progressive) to describe an activity currently in progress," then it is much easier to create specifications that check that a student understands which tense to use in a particular circumstance. You can then choose whether to test these tenses separately or together, select formats that suit your purpose, and decide whether to have students produce answers or simply identify correct responses.

Looking again at how assessment fits in with the rest of the curriculum, we note the importance of analysis and feedback. Administrators are always eager to get results such as grades from assessments, but it is equally important to make time for analysis. Thorough analysis can identify constructive changes for other components of the program such as syllabus sequencing, textbook choice, or teaching strategies. Analysis is the basis for helpful feedback to students, teachers, and administrators. Assessment coupled with analysis can improve instruction; assessment alone cannot.

## The Assessment Process

The six major steps in the assessment process are: (1) planning, (2) development, (3) administration, (4) analysis, (5) feedback, and (6) reflection. In turn, each step consists of a number of component steps. This flow chart will help you follow the first stages of the process.

### Planning



# Planning

## *Choosing Assessment for Your Needs*

Several steps are important in planning for assessment. First, you must consider *why* you are assessing and choose a type of assessment that fits your needs. What is the purpose of this assessment, and what kind of information do you need to get from it? Is a test the best means of assessment at this point, or would some form of alternative assessment do the job better? What abilities do you want to measure, and what kind of mental model, or *construct*, do you have of these abilities? For example, do you consider listening to be predominantly a receptive skill, or is listening so closely paired with speaking in interactive situations that you must assess the two skills together? For your purposes at this time, is it important to assess a skill directly by having students produce writing or is it sufficient to indirectly test some aspects of their writing?

Bachman and Palmer (1996) emphasize the importance of "target language use (TLU) domain," which they define as "tasks that the test taker is likely to encounter outside of the test itself, and to which we want our inferences about language ability to generalize" (p. 44). They further distinguish "real-life domains" that resemble communication situations students will encounter in daily life from "language instruction domains" featured in teaching and learning situations. For a student planning to work in an office, learning how to take messages would be an example of the former, while note-taking during lectures exemplifies the latter. In both cases, teachers need to take the target language use into account in the initial stages of their assessment planning and choose assessment tasks that reflect TLU domains in realistic or authentic ways.

If you are assessing progress or achievement in a particular part of the syllabus, you need to "map" the content and main objectives of this section of the course. Remember that you cannot assess everything, so you have to make choices about *what* to assess. Some teachers find it helpful to visualize assessment as an album of student progress that contains photographs and mementos of a wide range of work. Just as a snapshot captures a single image, a test or quiz shows a student's performance at one point in time. The mementos are samples of other kinds of student performance such as journal entries, reports, or graphics used in a presentation. All of these together offer a broader picture of the student's linguistic ability. Thus, in deciding what to assess, you also have to decide the best means of assessment for those objectives.

As you map the material to be assessed, there are several other factors to be considered: What weighting do you assign to the objectives? Are they equally important, or are some more fundamental to the course as a whole? Is this

assessment focused on recent material, or does it comprehensively include material from earlier in the course? Which skills do you plan to assess, and will you test them separately or integrate them? Sometimes time and resources constrain the skills that you can practically assess, but it is important to avoid the trap of choosing items or tasks simply because they are easy to create or grade. As always, testing should reflect teaching and the amount of time spent on something in the classroom.

Mapping out the course content and objectives is not the only kind of inventory. At this stage of assessment planning, you must also take stock of other kinds of resources that may determine your choices. What realistic assessment options do you have in your teaching situation? If all your colleagues use tests and quizzes, can you opt for portfolios and interviews? How much time do you have to design, develop, administer, grade, and analyze assessment? Do you have the physical facilities to support your choice? For example, if you decide to have students videotape each other's presentations, is this feasible? How much lead time do you need to print and collate paper-and-pencil exams? Computer-based testing may sound great, but do you have the appropriate software, hardware, and technical support? These are a simple handful of important aspects to consider in determining what your assessment will look like.

Autonomy is another factor in planning assessment. Typically, assessment is coordinated with other colleagues in a department, with teachers using common tests for midterm and final examinations as well as agreeing on alternative assessment tasks for a course. This arrangement may mean that you have autonomy for some kinds of classroom assessment but are expected to contribute to the design and grading of assessments done on a larger scale. In other cases, notably at the college level, teachers have more autonomy in planning which kinds of assessment to use for their own classes. It can be a real advantage to work collaboratively as part of an assessment team because each person benefits from the input and constructive suggestions of other people. If you do work by yourself, find colleagues who teach similar courses and are willing to work with you and give feedback. In either a centralized or autonomous situation, it is useful to develop specifications to ensure continuity and reliability from one instructor or semester to another.

## Specifications

A *specification* is a detailed description of exactly *what* is being assessed and *how* it is being done. In large institutions and for standardized public examinations, specifications become official documents that clearly state all the components and criteria for assessment. However, for the average classroom teacher, much simpler specifications provide an opportunity to clarify your assessment decisions. When several colleagues contribute individual items or sections to a "home-grown" assessment, specifications provide a common set of criteria for development and evaluation. By agreeing to use a common "recipe" or "formula," all contributors share a clear idea of expectations. An assessment instrument built on specifications is coherent and cohesive. If a test has multiple versions, specifications provide a kind of "quality control" so that the versions are truly comparable and thus reliable. Moreover, the use of specifications contributes to transparency and accountability because the underlying rationale is made very explicit.

Specifications can be simple or complex, depending on the context for assessment. As a rule, the more formal and higher-stakes the assessment, the more detailed specifications need to be to ensure validity and reliability. There are several excellent language testing books that provide detailed discussions of specification development. For example, Alderson, Clapham, and Wall's (1995) chapter on test specifications concludes with a useful checklist of 21 components (p. 38), while Davidson and Lynch's (2002) entire book is devoted to writing and using language test specifications. Davidson and Lynch define the essential components of specifications. For classroom purposes, far simpler specifications might include:

- a general description of the assessment
- a list of skills to be tested and operations students should be able to do
- the techniques for assessing those skills
  - the formats and tasks to be used
  - the types of prompts given for each task
  - the expected type of response for each task
  - the timing for the task
- the expected level of performance and grading criteria

Examples of specifications are provided in each of the skills chapters (i.e., Chapters 3–6).

In discussing item types and tasks, H. D. Brown (2004) makes a useful distinction between elicitation and response modes. Elicitation modes refer

to ways in which responses are prompted, whereas response modes refer to ways a student can respond to a question. Students listen to an oral prompt or read a written prompt then respond through speaking or writing (p. 51). For example, students could listen to a dialogue as an oral prompt and then write short answers in response. Within each mode, there are many different options for formats. It is important to avoid skill contamination by requiring too much prompt reading for a listening task or giving a long listening prompt for a writing task because that tests memory and not listening skills. The chart that follows makes these combinations of prompts and responses clearer.

<b>Prompt</b>	Oral (listening)	Oral (listening)	Written (reading)	Written (reading)
<b>Response</b>	Speaking	Writing	Speaking	Writing

Some of the most common item formats and assessment tasks are detailed in Chapter 2. Sometimes the range of options seems daunting, especially to teachers without much experience in writing exams. Hughes (2003) makes the practical suggestion of using professionally designed exams as sources for inspiration (p. 59). Using published materials as models for writing your own versions is quite different from the practice of adapting or copying exams that were developed for other circumstances. Teachers who have to produce many assessments often keep a file of interesting formats or ideas that they modify to suit their own assessment situations. Make note of topics that appear in textbooks or on standardized exams and collect potential assessment material related to these topics.

A close inspection of the formats used in standardized examinations can be beneficial for both students and teachers. As a consequence of the No Child Left Behind policy, American students now take more high-stakes standardized exams than in the recent past. The results are used to judge teacher and school performance as well as that of students. An analysis of how the exams are organized and how the items are built often clarifies the intent of the test designers and their priorities. Professional testing organizations develop their assessments based on specifications. If you can deduce what these specifications are, you have a better understanding of how high-stakes exams are constructed, and you can also incorporate some of their features in your own assessments. This knowledge will benefit your students because they will be familiar with the operations and types of tasks that they will encounter later. In their guide to writing specifications, Davidson and Lynch



(2002) call this analysis of underlying specifications "reverse engineering" (pp. 41–44).

After you have your specifications well in hand, cross-check them with the course outcome statements to make sure the things you have decided to assess align with the major course objectives. Assessment design is an iterative or looping process in which you often return to your starting point, all in the interest of ensuring continuity between teaching and assessment.

Previous exams written to the same specifications and thoroughly analyzed after previous administrations are a tried-and-true source for exam items. If the exam was administered under secure conditions and kept secure, it is possible to recycle some items. The most logical candidates for recycling in a short period of time are discrete grammar or vocabulary items. Items that have fared well in item analysis can be slightly modified and used again. Exam sections that depend on long reading texts or listening passages are best kept secure for several years before recycling.

Although specifications usually refer to the form and content of tests or examinations (Davidson & Lynch, 2002), they are just as useful for other forms of assessment. In a multiple measures assessment plan, it is advisable to have specifications for any assessments that will be used by more than one teacher to ensure reliability between classes. For example, if 12 teachers have students working on projects, the expectations for what each project will include and how it will be graded should be clear to everyone involved.

## Constructing the Assessment

At this point, you have used your specifications for the overall design of the assessment and to write sections and individual items. If you worked as part of a team, your colleagues have carefully examined items you wrote as you have scrutinized theirs. Despite good intentions, all item writers produce some items that need to be edited or even rejected. A question that is very clear to the writer can be interpreted in a very different way by a fresh reader. For example, students sometimes produce unanticipated responses for short answers or gap-fill items or have an entirely different interpretation of the prompt or task. It is far better to catch ambiguities and misunderstandings at the test construction stage than later when the test is administered!

The next step is to prepare an answer key and scoring system for writing and speaking. Specific suggestions for grading will be given in Chapters 3, 4, 5,

and 6 that focus on skills. The answer key should give any alternative answers for open-ended questions and specify the degree of accuracy expected in spelling, for example. The length or duration of production should also be made clear (e.g., *write 250 words, speak for two minutes*, etc.). Decide on cut-off points or acceptable levels of mastery but be prepared to adjust them later. Design the answer key so that it is clear and ready to use.

Once the assessment is assembled, it is advisable to pilot it. Ideally, the test should be trialed with a group that is very similar to those who will use it, perhaps at another school or location. Don't tell students that they are taking the exam as a trial because that will affect their scores. If a trial with similar students is not possible, have colleagues take the test, adjusting the timing to allow for their level of competency.

Next, compare the answer key and scoring system with the results from the trial. Were there any unexpected answers that now must be considered? Are some items unclear or ambiguous? Are there any typographical errors or other physical/layout problems? Make any adjustments and finalize plans to reproduce the exam. Check that all necessary resources are available or reserved. Do a final proofread for any problems that may have crept in when you made changes. Double-check the numbering of items, sections, and pages. Electronically secure or anchor graphics so they don't "migrate" to unintended pages. No matter how good you believe your test is, always try it out on a human being before administering it to your actual target group. You may be surprised at certain results.

Be sure to back up the exam both electronically and in hard copies. Print the answer key or scoring sheet when you produce the exam. Keeping practicality in mind, produce the exam well in advance and store it securely. Nothing is more frustrating than a malfunctioning photocopy machine during exam week. Some textbook publishers now "bundle" computer-based testing (CBT) software such as ExamView® with their books. Such software is easy to use to create classroom or online tests. Tutorials typically accompany the software.

## Preparing Students

Students need accurate information about assessment, and they need to develop good test-taking skills. In our coverage of the assessment process, we will focus on providing information to students since test-taking skills themselves are addressed in Chapter 7.

Transparency means that students have accurate information about when

assessments will occur, what they will cover in terms of skills and material, how much the assessments are worth, and when students can reasonably get their results. If at all possible, provide an assessment schedule at the beginning of a school year or semester to hand out with the syllabus. Be clear but firm about due dates and policies regarding assessments not completed by those dates. For long-term assessments such as portfolios or projects, break down due dates into several interim dates instead of having the entire work due at once.

Many institutions now require students to have outcome statements from the beginning of a course of study so that they know what is expected of them. As you work through the syllabus, be sure to draw attention to these outcomes and clarify what they mean for students and what the consequences are of not meeting them. Point out that assessment tasks are designed to check that students have achieved the outcomes. Some teachers post outcomes in the classroom as a reminder of the goals to be achieved. Stress that the goals are achievable, and be positive about the progress students are making toward these goals.

## Assessment in the Larger Cycle

In this book, we devote separate chapters to developing student test-taking skills (Chapter 7), the administration of assessment (Chapter 8), and techniques for analyzing exams (Chapter 9). Remember, assessment is part of a larger cycle so schedule time for analysis, feedback, and reflection. Administrators typically understand why we need time to prepare and grade examinations, but they need to see that we conduct assessment for several reasons that matter to all involved.

First, we want information about how our students are progressing. We want to confirm that they are placed at an appropriate level and are moving forward with their language skills. We want to give them the results of formative assessment as quickly as possible and make helpful suggestions for further improvement.

Second, we want to know whether our assessment instruments themselves are working for us. Do exams assess the skills we purport to teach? For example, if business students need telephone skills in the workplace, do we use performance assessment to check their skills? Do our assessment techniques yield useful information, or do we need to make some changes in how and what we assess?

Third, we want feedback about all aspects of our programs from syllabi to

materials to instruction. Are the outcomes we have developed appropriate for our students? Are teachers able to cover material in the syllabus? Are the course materials performing as we had hoped or expected? If we notice that many students have problems in a particular area, what can we do to compensate? In general, have teaching techniques been effective in helping students to master material? If not, how could we make improvements? Good assessment can answer these questions, but only if we take the time to analyze results and communicate them to colleagues.

Last, assessment gives us a chance to take a big step and rethink our English program's current aims and whether they fit the needs of current students. Technological innovations push us to consider different goals and different ways of assessment. Twenty years ago, English language learners mostly dealt with printed texts, but today the Internet has changed all that. Now we have to consider whether our instruction provides students with the means to independently access Internet resources and evaluate whether or not they are germane. English viewed as an international language means that we no longer focus on native-speaker standards and imparting cultural information about places where English is the primary language. Instead, our focus is on clear communication in a global context. Clearly, assessment is a useful tool for evaluating program aims.

## ***Ten Things to Remember about Developing Assessment***

**1. Assessment is an integral part of the teaching/learning cycle.**

It involves planning, development, administration, analysis, feedback, and reflection.

**2. Teachers are often involved with several phases of assessment at the same time.**

It is not unusual to be analyzing a previous exam while you are writing new items.

**3. Transparency and accountability are the hallmarks of good assessment.**

Be open about your rationale, your procedures, and your results.

**4. Specifications are tools that link learning outcomes with assessment instruments.**

Use specifications to promote validity and reliability.

**5. Assessment development is a collegial activity.**

Benefit from the contributions and feedback from your colleagues. Work together to review new items and guidelines for alternative assessment. Strive for consensus on marking.

**6. Prepare grading criteria and answer keys along with tests.**

Thinking through grading issues clarifies ambiguous tasks and items at a stage when you can still remedy the situation. Don't wait until after giving an examination to develop answer keys.

**7. Professional development in assessment benefits everyone.**

Schedule sessions on item writing, grading, analysis, and feedback to enhance skills.

**8. The cycle does not stop with obtaining grades. Schedule time for thorough analysis.**

Get the most from your assessments by understanding how tests are performed.

**9. Provide feedback in a timely and accessible manner.**

Tailor your feedback to your audience. Make feedback to students and administrators something they can actually use.

**10. Plan time for reflection on assessment.**

Did your assessment serve its purpose well? Was it part of the students' learning experience? What improvements do you plan to make the next time? Write these things down so you don't forget them!