17 *Reading assessment*

An understanding of the principles and uses of assessment is essential for *all* teachers, and in particular for teachers of reading. (Snow, Griffin, & Burns, 2005: 179)

Reading assessment has great power to inform researchers, teachers, administrators, and policy makers. Assessment practices can significantly benefit the learning environment or they can inflict great harm. Reading assessment, therefore, needs to be treated with great care, attention, and respect. Teachers, especially, have a responsibility to understand the uses and the impacts of reading assessment and be mindful of the consequences of assessment.

Reading assessments are used for many purposes, but all appropriate uses begin from an understanding of the reading construct, an awareness of the development of reading abilities, and an effort to reflect the construct in assessment tasks. The first five chapters of this book, in effect, represent a reasonable interpretation of the construct of reading ability. Chapters 10 through 15 describe key aspects of the reading development process. The complexity of the construct of reading, as well as its development, also reveals the potential complexity of reading assessment.

Reading assessment can be intimidating and sometimes overwhelming for many teachers and administrators; thus, a first goal of this chapter is to present a straightforward framework that categorizes the many uses and purposes for assessment. A fairly simple, yet thorough framework should allow readers to sort through their own assessment experiences in a way that gives interpretive force to the framework. The chapter then outlines and describes a number of major options under each category in the assessment framework. These assessment options are equally applicable in both L1 and L2 contexts, though important L2 test and assessment practices are noted where relevant. No effort at comprehensiveness is intended for assessment practices and descriptions. A number of very good books provide detailed descriptions and discussions of the many options noted for assessment practices (e.g., Alderson, 2000; McKenna & Stahl, 2004). The third section considers a number of innovations and

challenges for reading assessment. The fourth section addresses a set of further issues for reading assessment.

Goals for reading assessment

Reading assessments are meant to provide feedback on the skills, processes, and knowledge resources that represent reading abilities (Chapters 1–5), though it is important to note that different assessment practices may assume different theories of reading and reading development. Assessment in general can be categorized in a number of ways, and all assessment frameworks serve important purposes. Commonly, assessment has been categorized in terms of (a) norm-reference and criterion-reference testing; (b) formative and summative assessment; (c) formal and informal (or alternative) assessment; and (d) proficiency, achievement, placement, and diagnostic assessment. For the purposes of this book, reading assessment is organized and described in terms of five basic assessment purposes listed in Table 17.1.

Table 17.1. *Five purposes for reading assessment*

1. Reading-proficiency assessment (standardized testing)
2. Assessment of classroom learning
3. Assessment *for* learning (supporting student learning is the purpose)
4. Assessment of curricular effectiveness
5. Assessment for research purposes

There is an inevitable overlap among specific test uses across these categories, but these categories, nonetheless, serve as a useful framework for organizing reading assessment.

Reading-proficiency assessment

Assessment of reading proficiency is important as a way to understand students' overall reading abilities (based on some assumed construct of reading) and to determine if students are appropriately prepared for further learning and educational advancement. Commonly, this type of assessment is referred to as *standardized testing*, although local groups and researchers also develop proficiency tests of different types. In most respects, proficiency assessment represents high-stakes testing because decisions are often made about students' future educational goals and opportunities. Alternatively, this type of assessment may lead to special education or reading-disability designations – labels that, once applied,

are hard to remove from a student's record. Reading-proficiency assessment is also sometimes used for student placement, for policy decisions, for curriculum changes, or for program, teacher, or institutional evaluations.

Assessment of classroom learning

Assessment of reading improvement in classroom settings involves the measurement of skills and knowledge gained over a period of time and is commonly referred to as *summative* or *achievement testing*. Sometimes, proficiency assessments are used to measure student progress from year to year (as in a final exam), but this type of reading assessment does not capture ongoing student gains made in reading skills in the classroom. Year-end testing actually measures growth in proficiency from year to year rather than measuring gains in reading abilities *based on what was taught in class*. Much more commonly, assessment of classroom learning uses tasks that reflect the material taught in class and the skills practiced. Typically, the teacher, teacher groups, or curriculum groups (or textbook-materials writers) develop these tests, and they are responsible for deciding what represents a measure of success, as well as what steps to take as a result of assessment outcomes.

Teachers have multiple opportunities to assess student learning at several points in any semester using common techniques (e.g., end-of-unit tests, quizzes of various types, postreading comprehension questions, etc.), but some classroom assessment alternatives are less obvious. *Informal and alternative assessment* options are central for the effective assessment of learning (e.g., student observations, self-reporting measures, progress charts, engagement and group work, group outcomes assessment, interviews), and they usually provide converging evidence over time for the appropriate summative assessment at the end of the school year. Assessment of learning can be either *normative* (how students compare to each other) or *criterion-based* (how well students perform on curriculum standards and established learning goals). These two testing purposes should lead to somewhat different tests and different scoring. To give the simplest example, normative testing would discourage every student from receiving an "A," but criterion-based tests may include all students receiving an "A."

Assessment for learning

Assessment for learning involves a type of reading assessment that is not commonly discussed and is somewhat innovative in discussions of L2 assessment. This assessment purpose is intended to support and promote student learning, in this case, the improvement of reading abilities.

Performance evaluation or a record of outcomes is not the goal; instead, the goal is to provide immediate feedback on tasks and to teach students to engage in more effective learning. In many respects, this approach appears to overlap with the assessment of classroom learning, but this is true only with respect to many of the reading tasks performed, not to the follow-up feedback and interaction between the teacher and the students. Assessment for learning engages students in their own learning and responds to indicators of nonunderstanding or weak performances with ongoing remediation and fine-tuning of instruction. There are two general types of "assessment for learning" practices: One involves the use of recognizable classroom assessment activities to provide helpful feedback for learning; the second involves specific assessment for learning practices to support students directly in their day-to-day learning. These unique assessment-for-learning practices are described in the next major section of this chapter.

Assessment of curricular effectiveness

Assessment of curricular effectiveness and program evaluation is not specific to reading but is relevant for the development and / or review of reading curricula. Assessment outcomes that apply to curricular effectiveness include standardized testing, cumulative records over years that indicate gains or losses in student outcomes, interviews with teachers, students, and school administrators on summative test performance, feedback from institutions that receive graduates from the program or school, and innovative assessments that highlight specific school or program goals (e.g., project work, motivation, extensive reading, writing skills, or collaboration and group work). Evaluations of curricular success and teacher effectiveness represent different types of evaluation goals and extend beyond the immediate goals of student assessment, but they are important considerations for any large-scale assessment of reading curricula as well as programmatic needs analysis. This topic will not be developed further in this chapter, but important teacher and program evaluation ideas and resources are described in Brown (1995), Lynch (1996), Rea-Dickins & Germaine (1998), Richards (2001), and Thornton, Burch, and El-Araby (2003).

Assessment for research purposes

Assessment for research purposes is a topic that is not generally addressed in assessment chapters, but it is one that is very important for reading-research results as well as for their implications for reading instruction. Research studies sometimes use standardized assessment instruments to measure student levels or student instructional outcomes. In other

studies, however, researchers develop their own reading-assessment measures for a variety of reasons. Regardless of reason, research-developed measures need to conform to expected requirements for any appropriate and fair assessment practice. The measures, first and foremost, need to be valid; that is, reliable, construct-relevant, useful, fair, and responsible (with respect to consequences). Reading research can have a powerful impact on teaching and on students' learning experiences. Assessment measures are a part of all of these research conclusions and they need to be trustworthy. Given that students are likely to perform somewhat differently even across different standardized measures (Cutting & Scarborough, 2006), it is important to ensure that tests are developed and used appropriately. The value of multiple measures in any research context must also be stressed.

As the above framework indicates, reading-assessment practices (much like all educational assessment) can cover a wide range of purposes and uses, and each purpose or use includes a number of specific tasks and measurement options. In the section to follow, a subset of these tasks and practices is described. Moreover, the discussion will be limited primarily to L2 reading-assessment contexts. Most, if not all, of the activities outlined are equally applicable to L1 contexts, though not necessarily with the same assessment tools and resources. However, it is not possible to do justice to the field of reading assessment as a whole (nor even for L2 reading) in a single chapter (see Alderson, 2000; Urquhart & Weir, 1998; Weir, 2000).

L2 reading-assessment practices and resources

> Any single technique for assessment will necessarily be limited in the picture it can provide. . . . We should always be aware that the techniques we use will be imperfect, and therefore we should always seek to use multiple methods and techniques, and we should be modest in the claims we make. (Alderson, 2000: 270)

In this description of L2 assessment practices, for ease of explanation, we link reading-proficiency assessment with standardized testing and assessment of learning primarily with classroom-based measures. There is certainly an amount of overlap between standardized and classroom settings, and some of these instances are noted, but this simplification eliminates the need to detail all crossover points.

Standardized L2 reading assessment

What all standardized reading tests have in common is an effort to reflect the construct of reading-comprehension abilities in one form or another.

Standardized assessment makes a serious effort to capture crucial aspects of the component abilities of comprehension. Drawing on these assumptions for standardized test construction, and considering the component abilities outlined in this book, standardized reading assessment should seek to translate (aspects of) the reading construct listed in Table 17.2 into an effective reading test.

Table 17.2. *Major component abilities for reading comprehension*

1. Fluency and reading speed
2. Automaticity and rapid word recognition
3. Search processes
4. Vocabulary knowledge
5. Morphological knowledge
6. Syntactic knowledge
7. Text-structure awareness and discourse organization
8. Main-ideas comprehension
9. Recall of relevant details
10. Inferences about text information
11. Strategic-processing abilities
12. Summarization abilities
13. Synthesis skills
14. Evaluation and critical reading

Among the challenges to consider for reading assessment is how such an array of component abilities can best be captured within the operational constraints of standardized testing, what new assessment tasks might be developed, and what component abilities (e.g., grammar) might best be assessed indirectly. While it is possible to outline the many component abilities of reading comprehension, it is less straightforward to capture all of these abilities in reading-assessment tasks. Standardized assessment practices are far more constrained by concerns of validity, reliability, time, cost, usability, and consequence than classroom assessment practices. These concerns limit the types of reading-assessment tasks that can be used. Also, the context for standardized assessment precludes any strong assumption of a match to authentic reading in the "real world." When students read a text as part of standardized assessment, they know that they are reading for an assessment purpose. Nonetheless, more realistic texts, tasks, and contexts are helpful as long as they do not pretend to be authentic "real-world" reading tasks.

A further complication for standardized reading assessment is that different tasks and task types are appropriate at different proficiency levels. How reading-assessment tasks and task types should change with growing L2 proficiency is an area that has not been intensively investigated (although it is logically assumed, and appropriately so, in many

standardized multilevel assessment batteries). Commonly, assessment practices tend to focus on specific target populations within a more restricted proficiency range, and tasks may not be valid for populations at much lower or higher proficiencies. Having a theory of reading development as one of the validity supports for assessment practices might provide a more complete understanding of L2 reading abilities, their expected patterns of growth, and rationales for using different types of assessment tasks at differing proficiency levels.

Until fairly recently, standardized L2 reading assessment had not been overly concerned with the development of reading assessment in terms of an evidence-based construct of reading abilities tied to the group of students being assessed (see Mislevy, Almond, & Lukas, 2003). However, there are now a number of examples of major standardized assessments being developed from an initial set of claims about the nature of L2 reading ability and a set of tasks that would measure the relevant component skills.

The development of the IELTS (International English Language Testing System) represents one example of a standardized test built from construct assumptions and the gathering of appropriate evidence (Clapham, 1996). Similarly, efforts to redesign the TOEFL (Test of English as a Foreign Language) as TOEFL®iBT (internet-based testing) required the development of an appropriate L2 reading construct (among other language skills) as well as evidence to support assessment tasks that would measure this construct (see Chapelle, Burns, & Grabe, 1997; Chapelle, Enright, & Jamieson, 2008). Additional approaches to L2 standardized assessment that are built from claims about reading abilities include the suite of Cambridge English proficiency exams (Weir & Milanovic, 2003) and the Advanced English Reading Test in China (Weir, 2000). These approaches to L2 reading assessment strongly document arguments for an L2 reading construct, the importance of specific components of reading ability, the types of tasks that can assess these component abilities, and the creation of overall tests that generate evidence for the claims made (thus building a validity argument for the appropriateness of the test).

It is important to look at the types of tasks developed for standardized reading tests, consider how these major tests incorporate and reflect the reading construct, and how they engage L2 learners in fair and appropriate assessment tasks. The primary purpose of assessment tasks is to collect information to make inferences about students' reading abilities. Different reading tasks should help provide information about many component reading abilities as well as reading comprehension more generally. Reading assessment tasks come in many recognizable forms in standardized assessment (as well as a few uncommon options). The list in Table 17.3 summarizes most major task options used in

standardized reading assessments. The items in Table 17.3 are reviewed in Alderson (2000), Hughes (2003), and Weir and Milanovic (2003), and multiple useful examples are given.

Table 17.3. *Standardized reading assessment task formats*

1. Cloze
2. Gap-filling formats (rational cloze formats)
3. C-tests (retain initial letters of words removed)
4. Cloze elide (remove extra word)
5. Text segment ordering
6. Text gap
7. Choosing from a "heading bank" for identified paragraphs
8. Multiple-choice
9. Sentence completion
10. Matching (and multiple matching) techniques
11. Classification into groups
12. Dichotomous items (T / F / not stated, Y / N)
13. Editing
14. Short answer
15. Free recall
16. Summary (1 sentence, 2 sentences, 5–6 sentences)
17. Information transfer (graphs, tables, flow charts, outlines, maps)
18. Project performance
19. Skimming
20. Scanning

Many of these task formats are well-known and widely used in standardized tests. I will comment selectively on a number of them, and identify formats that are less common but that do appear in standardized assessments. Cloze assessments with random n-th word deletions (every sixth word, or every seventh word) are not automatically valid assessments of reading abilities, particularly when students are expected to write in the missing words. Such tests become production measures and are not appropriate for L2 reading assessment. Much more useful options are gap-filling measures (rational cloze formats) that target specific words purposefully (e.g., prepositions, verbs) rather than delete every seventh word (for example). However, even with gap-filling formats, a reading measure should not ask students to fill in words (as a production task) that they do not know or have not already seen from reading a text beforehand (unlike short-answer formats in which students have read a nonmutilated text beforehand).

C-tests are variants of cloze formats, but rather than deleting whole words, the initial letter or syllable of a targeted word remains, and students use this clue, along with the sentence context, to determine the missing word. This option is less of a production task. Cloze elide, in which "extra" words are meant to be struck out, have the advantage of

not requiring any production. In both cases, however, it is not easy to make persuasive construct-validity arguments for these formats.

Text-segment ordering and text-gap formats involve the moving around of whole sentences or paragraphs, or the selection of the right space in the text to supply a sentence or paragraph. Text-gap formats can be tricky when multiple gaps are created and a list of sentences or paragraphs is provided to insert in the correct spaces. These formats amount to a type of multiple matching task. Choosing from a heading bank to label identified paragraphs is a similar type of task. The strength of these types of tasks is that they call on knowledge of discourse signals and discourse structuring to be answered successfully. They require several comprehension skills for appropriate task completion.

A number of task formats in Table 17.3 are relatively uncommon. Free-recall formats simply ask test takers to make a list of ideas they remember from a text they have just read. These responses are matched up against a list established by the test maker. Summary formats can be straightforward though difficult to score. Alternative summary formats can include, for example, choosing the best from among three summary options and identifying the weaknesses of unacceptable options. Information-transfer formats, especially visual representations of text information, have powerful construct-validity support. However, they can be difficult to score and can have very high item interdependence. Project-performance evaluation is a newer task format that evaluates test takers as they read texts and then perform in groups to carry out a larger project. It is an interesting option, but is problematic on several validity grounds (giving individual scores based on group interactions and a holistic task). Skimming and scanning tasks are well-known to teachers but are not common as standardized reading assessment tasks.

In closing this discussion of standardized assessment, it is important to emphasize that all formats need to go through a careful validation process that includes a theoretical justification, a feasibility study (Does the task work the way that it is supposed to?), a piloting study for reliability and item performance, and a fairness review. While individual teachers are seldom called upon to create standardized tests, programs and schools are sometimes expected to generate a standardized assessment. It is important that medium- and high-stakes tests be constructed and used carefully and in the fairest possible way.

Classroom-based assessment practices

> A first concern for classroom teachers is collecting and using reading assessment information . . . to shape instruction and learning. . . . A robust classroom assessment program continually provides detailed information about students' current competencies and next steps. (Afflerbach, 2007: 268)

> Inevitably in contexts where informal teacher- or classroom-based techniques are used or advocated, little reference is made to their validity, accuracy, or reliability, and much more is made of their "usefulness" and "completeness." (Alderson, 2000: 268–9)

Classroom assessment allows for a much wider range of tasks and student observations. Testing formats in classroom contexts can include all of the assessment formats identified for standardized assessment (Table 17.3). Classroom reading assessments can also make use of informal reading inventories or miscue analysis (reading aloud one-on-one with an evaluator who notes errors on a record sheet and then determines what progress a student has made or what instructional support is needed by the student).

A number of other informal, or alternative, assessment types are also well recognized and commonly used in classroom contexts. Almost any language task that is a useful teaching task can be used as an assessment task. What might be lost in the way of relatively weak validity or consistency for any given reading task or measurement in the classroom setting is, in principle, countered by the continual nature of assessment practices of all types in this context. Several types of informal assessment options are available to teachers, including the following:

1. Observations
2. Self-reporting measures
3. Progress charts
4. Performance inventories
5. Participation and engagement records
6. Portfolios

Each general type of informal assessment category can be carried out by means of several specific tasks and in different formats (see Table 17.4). Regardless of informal assessment formats used, teachers and administrators have a responsibility to focus on appropriate tasks and interpretations of task outcomes so that students are not evaluated unfairly. Table 17.4 identifies a large number of informal assessment options that can be used in ongoing classroom reading-skills assessment. The items in this table are grouped according to the six major informal assessment types noted above.

Many of the assessment options noted below are fairly simple and straightforward for teachers to use as means to gather important information about students' reading abilities. A key issue for informal reading assessment includes the need for multiple assessment formats for any decisions about student abilities or student progress. In addition, informal formats should be operationalized in ways that provide more

Table 17.4. *Informal assessment formats*

1. Have students read aloud in class and evaluate their reading.
2. Keep a record of student responses to questions in class after a reading.
3. Keep notes on student participation in class discussions on a reading.
4. Observe what reading material is read during free reading or SSR.
5. Observe how much time students spend on tasks during free reading or SSR.
6. Have students do paired readings and observe.
7. Observe students reading with an audiotape or listen to an audiotaped reading.

— — —

8. Have students list strategies they have used while reading.
9. Have students list words they want to know after reading and why.
10. Have students keep diaries or reading journals.
11. Have students write simple book reports.
12. Have students recommend books.
13. Ask students about their reasons for choosing certain answers in reading tasks and activities.
14. Ask students about their reading progress.
15. Ask students about their goals for reading with various texts and tasks.

— — —

16. Keep charts of student readings.
17. Keep charts of student reading-rate growth.
18. Record how far a student reads on an extended reading task.

— — —

19. Have a student read and then have a discussion on the text (one-on-one).
20. Have a student read aloud for the teacher / tester and make notes, or use a checklist or note miscues on the text (one-on-one).
21. Have students do think-alouds while reading (one-on-one).

— — —

22. Have students enact a scene / episode / event from a text.
23. Note the uses of texts in a multistep project and discuss.
24. Have students fill out simple questionnaires of interests and engagement levels in various tasks.

— — —

25. Create student portfolios of reading activities or progress indicators.

objective assessment than after-the-fact judgment or overall subjective assessment. Informal assessment can be made more objective when a teacher knows that certain assessment information will be used for grading, ranking, placement, or advancement. Observations can be carried out on a regular schedule with certain tasks and activities. Notes should be recorded consistently after specific tasks, and comments on student performances can be recorded at breaks in classes or at the end of the day. Records for assessment purposes should be kept consistently and carefully. Portfolios and projects can be constructed with clear goals in mind and assessed according to a preestablished set of grading criteria; the grading criteria need to be applied consistently to all students. The

goal is not to formalize informal assessment, but to remember to evaluate a wider range of student performances, thereby making informal assessment more effective and fairer to students.

When using various (and multiple) student performances to inform assessment, decide on a grading or commenting plan before evaluating students. Assessments involving students reading in class, participating in discussions, contributing to a project, reporting on a book they read, or performing in a scripted project (a theater reading, a news report, etc.) should include multiple assessment points (e.g., observe each student every 15 minutes, note performance of multiple steps in a performance). Reviews of student reading records, reading and fluency charts, and reading portfolios should include notes and comments on students' activities at multiple points in the school semester. Otherwise, it is often the case that the most recent one or two recollections dominate the assessment.

Self-assessment by students is an important component of informal assessment and there are multiple options that teachers can explore. Self-assessments can require students to (a) chart their progress on various skills; (b) note what they are reading and why; (c) explain their goals for reading and their reading choices; (d) list reading strategies that they use or want to use; and (e) assess their own evolving reading portfolios according to a few simple criteria. Through self-assessment, students develop greater self-awareness that feeds into assessment *for* learning when their self-assessments are reviewed, discussed, and reflected on. In fact, informal assessment practices, when objectivized in some way so that students participate in the assessment, can promote assessment for learning on a continual basis.

Assessment for learning practices

> It is generally acknowledged that increased use of formative assessment (or assessment for learning) leads to higher quality learning. (Wiliam et al., 2004: 49)

Assessment for learning is an alternative and somewhat innovative way to think about the goals of assessment and operationalize the notion that assessment should be used to promote learning (rather than to check learning). As such, assessment for learning involves more of a philosophy toward teaching and student learning than a separate set of assessment practices; it is not specifically addressed to reading, although it easily can be so. In effect, the crucial goals in assessment for learning involve teaching procedures that use assessment information as major opportunities for learning and the development of more effective skills over time. The key, in this case, is not to provide answers, but to enhance learning, work through misunderstandings that are apparent from

student performance, develop effective learning strategies, and encourage student self-awareness and motivation to improve.

This goal is often easier to state than to carry out consistently in the classroom. However, Black and Wiliam (1998, 2005; Wiliam, 2007/2008) note that five reviews of research on this topic have synthesized more than 4,000 research studies (across all areas of student learning) over the past 40 years. The conclusions of this research demonstrate that assessment for learning practices can double the rate of student learning (Wiliam, 2007/2008).

Assessment for learning is an approach that is well-suited to reading instruction. In its simplest form, the teacher gathers feedback on student performance (e.g., on reading-related activities) on a continual basis and engages students in improving their learning based on teacher responses. Teachers learn to respond to student signals of noncomprehension through teacher observations, outcomes of students' weak performance, or specific feedback mechanisms that students can use. In many cases, the techniques used can be associated with effective comprehension strategies instruction, though these strategy uses are in response to formative assessment activities. (See Black & Wiliam, 1998, 2005; James et al., 2006; Wiliam 2007/2008; Wiliam & Leahy, 2007; Wiliam & Thompson, 2007, for additional perspectives on assessment for learning.) The 15 ideas and techniques for assessment for learning outlined in Table 17.5 apply to any learning and assessment context, but it should be evident that these techniques are ideally suited to reading tasks and reading-comprehension development.

These 15 examples of engaging in assessment for learning represent only a subset of what teachers and students can do together to enhance learning through ongoing assessment feedback (see Black & Wiliam, 1998; Black et al., 2004; UK Qualifications and Curriculum Authority, 2008; Wiliam et al., 2004). In addition, many, if not most, of the informal assessment practices noted in the previous section (as well as standard assessment formats) create outstanding opportunities for teachers to engage students and groups of students in feedback. When implementing assessment for learning, the first step is to agree upon feedback mechanisms from students to teachers that allow students to signal difficulties that they are having. Responses from teachers should (a) address skills needed to improve learning; (b) encourage greater student awareness of what successful outcomes would look like; and (c) provide opportunities to help students become more successful. Because carrying out assessment for learning on a consistent basis can be difficult for teachers at the outset, Black et al. (2004) and Wiliam (2007/2008) offer a number of guidelines for implementing these practices.

Whichever assessment for learning concepts are chosen, they should be used on a regular basis so that students know what to expect, how to respond, and how to use feedback from the teacher. Students also need

Table 17.5. *Assessment for learning techniques*

1. Teachers wait for 3–5 seconds after asking a question, without answering the question when students are silent, or without switching to a strong student for the answer.
2. Teachers ask students to generate good questions about reading texts that could be included on a test of some sort. Students learn how to formulate good questions about the texts they read, and their questions then appear on quizzes and comprehension checks.
3. Teachers move from more traditional question-and-answer sequences about reading passages to questions that begin a discussion with students about their understanding of the text.
4. Teachers withhold grades on student work until students respond to comments on their assignments (e.g., multiple text syntheses, filled-in graphic organizers). In many cases, teachers do not provide grades at all, only comments.
5. Teachers deal with wrong answers or performance difficulties in ways that engage students in finding good answers and achieving task success. Teachers provide feedback to encourage student learning.
6. Teachers engage in "why" questions and "why" follow-up questions to generate discussions about an answer.
7. Students engage in self-assessment, monitoring their comprehension and their levels of success in carrying out during-reading and postreading activities.
8. Students work in groups to answer questions and are given time to work out answers together.
9. Students set learning goals at the outset of instruction.
10. Students assess the work of other students, usually through comments rather than specific grades.
11. Students select a peer who will answer a question, for example, by drawing a student's name from a hat.
12. Students learn to monitor their noncomprehension of a text or task and signal for assistance when they have difficulty.
13. Students review their own learning effectiveness.
14. Students redo their answers on a reading task or a quiz, often with peer support.
15. Students use "traffic-light" discs (green on one side, red on the other) to silently signal to the teacher that they are doing well or having trouble when working on their own or in groups.

to be encouraged to see assessment for learning practices in a positive light. This need for a positive learning cycle around brief, consistent, and informative assessments means that teachers must not use the assessment procedures primarily to sort, evaluate, or grade students. Both teacher and students need to be sold on the positive power of assessment practices in the classroom (Black & Wiliam, 1998; Black, et al., 2004). To carry out assessment for learning effectively, teachers need a significant amount of training, and students need consistent support and reinforcement. Assessment for learning works when it is used consistently and positively over an extended period of time. In L2 assessment and learning

contexts, assessment for learning has been receiving increasing attention in the UK, in Australia, and in Hong Kong. In some cases, these applications involve many similar ideas and approaches to those outlined above. In other cases, the concepts of assessment for learning, and the larger notion of formative assessment, have been adapted into standardized assessment schemes for overall student performance. Examples of these efforts, and related discussion, are provided by Colby-Kelly and Turner (2007), Davison (2007), Lee (2007), Leung (2004), and Rea-Dickens (2004). As of yet, there have been no efforts to implement assessment for learning practices in L2 reading contexts, either in classroom settings or in more standardized assessment versions.

Placement and diagnostic assessment practices

Most of the assessment formats identified in this chapter provide the foundation for placement and diagnostic assessment. The key point about assessments for placement and diagnostic purposes is that they are usually more locally driven: The class, program, and institute have specific needs for placement and for the diagnosis of students' strengths and weaknesses. It is true that standardized tests are commonly used for placement purposes (e.g., GORT [Gray Oral Reading Test] or TOWRE [Test of Word Reading Efficiency] for young L1 students; TOEFL or IELTS for L2 students), but programs and institutions often require additional assessment information and give follow-up assessments of crucial language skills, including reading skills. Locally based placement tests often also focus on specific skills that a program feels are important and need to be assessed. For reading, these skills might include comprehension of academic material, discourse awareness for reading difficult texts, and writing from text resources. At more basic levels of L2 reading ability, students can be assessed and placed according to vocabulary knowledge, oral passage reading, grammar knowledge, and basic comprehension.

Diagnostic assessment can be used for placement purposes, although the goals for placement and diagnosis should be distinct. Diagnostic assessment should present a battery of skills to students on a given ability level that may cause difficulties or (alternatively) should already be well-learned by students. Results should indicate the need for specific teaching practices and possible tutorial work that is designed specifically to address the weaknesses of each student. In many cases, diagnostic assessments lead to a plan for individualized work in addition to more generalized coursework for the whole class. Diagnostic assessments can focus on pseudo-word reading, vocabulary knowledge, sight-word reading for common words, analysis of complex syntactic structures, the ability to make reasonable inferences, and the ability to summarize main ideas, for example.

Research-based assessment practices

A final assessment practice topic addresses the role of assessment in research studies. In many cases, research studies make use of already existing standardized tests or create adapted versions from these tests. In other cases, researchers create their own tests using testing formats already identified in this chapter. In yet other cases, research uses measures that were developed for the specific research question and that are not common assessment formats. For example, reading sentences with alternating CaPiTaL and small letter cases is not a typical reading-assessment technique, but it does measure the impact of visual form on sentence processing, and indicates different levels of visual processing interference for L2 students from different L1s. Similarly, oral passage-reading tasks are not generally used for L2 assessment purposes but are useful to examine reading fluency in the right contexts. Regardless of assessment instruments and formats used for research, they all need to be validated under the various criteria for a validity argument (construct validity, reliability, usability, fairness, consequences).

Most common assessment measures that have been developed by researchers can be categorized under one or more of the six general options listed in Table 17.6.

Table 17.6. *Types of reading assessments for research purposes*

1. Time measures and response-time measures
2. Priming measures (speed and preference decisions based on co-occurring associations)
3. Frequency of occurrence measures (including category-level coding with qualitative data)
4. Specific skills and knowledge measures (including recall measures of various types)
5. Self-reporting measures (checklists, diaries, interviews, questionnaires)
6. Performance measures with information from texts

Reading research uses all of these common types of measures as well as a few that are unique and innovative. Regardless of measures used, they all involve assessment practices and need to conform to expectations applied to reading assessment more generally.

Issues and innovations in L2 reading assessment

In this section, I identify some possible themes for L2 reading assessment that might better capture the skills and abilities used in reading for various purposes. Within the area of reading for academic purposes, there might be a need for different types of reading measures that do

not currently fit easily within common assessment instruments. While almost everyone would agree that test takers need to be given a variety of types of tasks reflecting different skills, the issue is what might count as tasks that capture new but important aspects of reading comprehension. Adding different types of assessment tasks to the standard repertoire also complicates validity concerns for any reading test. These are interesting challenges that such assessment tasks pose, and any new task that is given credibility as a valid assessment task must add further knowledge about a test taker's reading abilities. Below are six options for new reading assessment tasks, but they are only possibilities that are still in need of validation.

First, an option that has not been explored seriously in L2 reading assessment is to consider innovative ways to assess vocabulary knowledge. At present, vocabulary assessment and instruction have been rather static. But recent conceptualizations of vocabulary instruction in English L1 contexts raise new possibilities for vocabulary assessment as well (Baumann & Kame'enui, 2004; Hiebert & Kamil, 2005; Stahl & Nagy, 2006; Wagner, Muse, & Tannenbaum, 2007b). A comparable surge of research activity on vocabulary instruction in L2 settings would be helpful in suggesting options and opportunities for vocabulary tasks as part of reading assessment. Just to suggest two possibilities, test takers could be asked to carry out sorting or classification tasks with large groups of words according to specified category labels, or they could be asked to take words from a list to fill in a relevant diagram. This latter task might be very useful with lower-proficiency students.

Second, another option for reading assessment is to develop tasks that encourage students to read longer texts (for advanced assessment, 700–1,200 words, assuming a rate of 120–150 wpm). There are currently some tests that require test takers to read over, or skim, longer texts in brief time periods and answer general questions. In contrast, this option would ask test takers actually to read for main ideas, key details, connections among sets of information, and integration of information. Such items would need to presume a minimum reading rate and level of vocabulary knowledge to allow such a task with a longer text. A test option might involve, with computers, reading a passage of 750 words (read at 150 wpm) that then disappears after 5 minutes. Test takers might then answer questions, fill in outlines, list major ideas, or click-and-drag 10 statements from a list of 20 statements about the text that are true. Alternatively, tests might have a passage of 750 words, and when students have finished reading it, they press a button; the text disappears and the questions appear (providing both rate and comprehension). These types of tasks also impose expectations about reading fluency as part of the subconstruct being measured.

Third, adding a time limit to the above tasks raises the more general topic of assessing reading fluency. Should fluency be assessed as part of

a reading comprehension measure? Are there aspects of fluency that are supportable as part of the construct of reading and that are feasible as assessment tasks? Obvious candidates would include exploring word-identification fluency skills and simple reading-rate skills (word lists, oral reading for 1 minute, silent reading on a computer, timed reading, assessment of rereading). These formats might be more useful at lower L2 proficiency levels. At higher levels, oral passage-reading measures might be relevant and fairly simple as a task type under the right conditions and scoring rubric (see Fuchs et al., 2001; Jenkins et al., 2003a). Such tests would ultimately place less emphasis on task authenticity and direct measurement of reading comprehension and more emphasis on prior extensive reading practice. Perhaps a task that directly supports extensive reading practice would be an entire test working with one or two long texts, with a variety of assessment tasks being produced around the text(s). The limitations of such a test format would also need to be addressed. Such a performance test would almost be like a professional or clinical performance assessment and would require a considerable amount of extensive reading as general preparation.

Fourth, assessment of discourse structure is another option that can be explored further. With respect to text factors, we need to know how reading-assessment difficulty can vary by text choice (genre, length, complexity, vocabulary, organization). Will literary texts vs. general narrative texts vs. informational texts vs. persuasive texts generate very different assessment outcomes? Discourse-structure knowledge might also be assessed effectively through various types of discourse-awareness tasks (see Chapter 12), and especially through the use of extended graphic organizers (Jiang, 2007). This use of graphic organizers is currently employed in some tests to a limited extent (as part of "information transfer"), but it should be explored further.

Fifth, strategies and metacognitive awareness have been interesting topics for the past decade as possibilities for reading assessment. However, this area has yet to become a major option in assessment practice (aside from summary practice and multiple-choice inference-based questions). At issue is how to assess these notions in useful ways. Some commonly assessed concepts, such as inferencing and monitoring, may be more basic cognitive processes than conscious reading strategies (see Chapters 3 and 10). Some promising strategies are amenable to individualized assessment (predicting, stating main ideas, summarizing, using context, forming questions) through student think-alouds in classroom contexts. However, in contexts other than one-on-one interviews, it is difficult to assess most strategy uses in reading. One assessment approach might be to have students self-report on strategy uses, but one of the limiting factors for this type of assessment is that questionnaires and self-reporting surveys only indicate strategic awareness and not actual strategy use while reading. An option for developing strategy use while

reading could involve a multi-task decision-making process. In some cases, the decisions made can reflect strategic actions on the part of the student, particularly if students report the reasons for their decisions as they work through multiple tasks and decision points. One example of a test that would allow such a format is the *iskills* test (Information and Communication Technology [ICT] literacy test from Educational Testing Service; www.ets.org/ictliteracy).

A further variation on assessing reading-strategy abilities could involve assessment of specific strategies that generate a product as an outcome of using the strategy. All of the following strategies, commonly identified as important for comprehension skills, could be developed in formats that generate a product as a way to assess strategic response to texts (see Table 17.7).

Table 17.7. *Strategy tasks that generate assessment outcomes*

1. Categorize statements as explicitly mentioned, inferrable, or not invoked in any way in the text.
2. Choose the most relevant background knowledge.
3. Develop a synthesis statement.
4. Decide on the best summary statement (monitoring).
5. Evaluate alternatives from multiple brief texts.
6. Explain the purpose of the text.
7. Form a summary.
8. Form questions about the text.
9. Identify the most relevant text-structuring signals in response to a prompt.
10. Make an inference and explain why the inference is appropriate.
11. Make appropriate associations to key phrases.
12. Map a concept described by a text (making a simple visual graphic).
13. Paraphrase a small text segment.
14. Predict upcoming text continuations.
15. Sort main-idea statements from supporting information (and both from incorrect statements).

The challenge for assessment specialists is how to develop an assessment format around a text or series of short texts that can be used for a multistep set of tasks to tap reading-strategy uses (see also Alderson, 2000).

Sixth, a final option for assessment innovations is to consider ways in which to test reading to integrate information and evaluate information. Tests might include tasks that assess synthesis skills, evaluation skills, or content monitoring while reading. Computer-based tasks have been developed that allow for the rearrangement of information, using a click-and-drag interface. More elaborate versions of integration and evaluation could be developed in which students read competing sets of

information and need to choose the better information for task completion. Versions of such tasks already exist as part of the ICT literacy instrument, *iskills,* developed by ETS for postsecondary technological literacy (www.ets.org/ictliteracy). These tasks deserve consideration for a range of more advanced L2 reading-assessment tasks.

It can be a fascinating exercise to consider possible "what if" assessment tasks related to reading abilities. I would just like to close by noting that the six exploratory ideas above are not meant to be seen as arbitrary options because of some loose connection to reading. They are all defensible to one extent or another as part of the reading construct described in this book. Ultimately, any ideas for new reading-assessment tasks must connect back to a coherent and plausible construct statement, and they need to provide additional information that better assesses the reading abilities relevant to a given set of students, in a specific situation, and for subsets of possible proficiency levels. But with the appropriate theoretical constraints set up at the outset, the potential development of new reading-assessment options can be interesting, challenging, and even, for an assessment specialist, fun.

Further issues for reading assessment

Several additional issues could be discussed, but I focus on two that are particularly important for the appropriate uses of assessments in reading contexts. The first is the recognition – on the parts of assessors, teachers, and administrators – of the consequences of assessment. The second is the importance of teacher training for effective and appropriate reading assessment.

Consequences of assessment

All assessments have consequences. In a classroom setting, students may feel that they did not perform well. Teachers may be disappointed in a specific performance by a student and let it show even if no grade is involved. Informal assessments may inadvertently be made based on one or two salient experiences rather than a fair accumulation of evidence over time and across tasks. Teachers and administrators may use grades, tests, and evaluation reports to make decisions about students that should require additional input. Summative achievement tests certainly have consequences, as do standardized tests. Even in an assessment for learning contexts, a steady diet of difficulties and poor performances can take a toll on both students and teachers. The point is that assessment, no matter the context, is serious business and must be handled with care, purpose, and expertise.

Most important, teachers cannot opt out of the consequences of assessment. Even when teachers state that they do not believe in standardized assessment or grading, they are generating consequences of assessment. They are leaving assessment to others who may be less informed about the students, their progress, and their needs. When teachers do not take assessment seriously as their responsibility, they give up their ability to advocate for students in assessment contexts or support fair assessment practices. Teachers need to develop expertise in assessment practices of all types if they are to ensure fair uses of assessment and appropriate consequences of assessments. Only when many teachers have true assessment expertise will their views be heard by administrators and policy consultants.

Assessment is sometimes treated as a bad word. But all of us engage in assessment all the time, both in our daily lives and in the classroom. We assess the value of most daily activities we engage in and most purchases we make. We assess our friends, our work, our leisure activities, and what we eat. We assess students' homework, their group work, their responses to pre-, during-, and postreading activities, and their level of engagement in SSR in class. Assessments also provide information for continuous learning. It should not be surprising, then, that assessments are essential to learning institutions, teachers, and students. Honest, fair, and appropriate assessments are needed by students if they are to learn most effectively. Taking this view of assessment also highlights the consequences of assessment. It is everyone's responsibility to ensure that these consequences are not harmful or unfair.

Teacher training for reading assessment

> There is little research that describes how teachers develop as assessment experts or that demonstrates what types of classroom assessment training most benefits teachers and their students. (Afflerbach, 2007: 278)

The above discussion of assessment consequences inevitably places a responsibility on teachers to ensure that assessments of all types are fair and appropriate. This responsibility, in turn, means that teachers must develop expertise in assessment practices and uses of the resulting outcomes. One obvious way to promote teacher expertise in assessment is to change teacher-development programs so that a greater emphasis is placed on assessment skills. However, assessment training itself needs to change to engage teachers much in the way that teachers have a responsibility to engage students in their own learning. As Afflerbach (2007) notes in the quote above, there is surprisingly little research on how teachers develop assessment expertise, or how they *can* develop this expertise. Teachers will not develop expertise in assessment in spite of

educational training institutions. Teacher-training programs also have a responsibility to ensure that teachers understand assessment practices and purposes, carry out assessments fairly and appropriately, and use assessments for effective learning purposes (see also Snow, Griffin, & Burns, 2005).

At the same time, there are a number of ways in which teachers can help themselves. When assessments are recognized as essential aspects of learning, teachers can explore specific aspects of assessment practices that appeal to them. It is not difficult for teachers to develop expertise in informal assessment practices and learn to use them appropriately. Teachers may also want to explore assessment for learning practices as a process of continual supportive feedback for student learning. Teachers may want to learn about specific standardized assessments, perhaps in study groups, to understand technical specifications, construct-validity arguments, and consequences of test uses. Well-informed teachers who can challenge or complement expert consultants will go a long way to the implementation of fair and appropriate assessment practices. Teacher engagement in assessment issues also suggests that researchers and teacher trainers need to know more about how teachers can develop assessment expertise. Assessment is far too important in the lives of students to leave matters of fairness and consequences to others. Assessment also has the potential to promote effective learning outcomes, and who could be against that?

Implications for instruction

The above section on responsibility for assessment and teacher development for assessment expertise has both a policy dimension and an instructional dimension. When teachers recognize the potential of assessment in student learning and the need to take responsibility for appropriate assessment practices, many implications for instruction emerge. The most central of these implications is that appropriate and effective assessment practices will lead to better student learning. The most obvious example of this is a consistent and ongoing effort to provide assessment for learning.

Many teachers might imagine assessment for learning as a process of weekly quizzes intended to measure learning progress. However, assessment for learning is not focused mainly on quizzes, graded homework, and tests. Rather, it is intended to provide immediate information to teachers who then adjust instruction accordingly to meet students' learning needs. In this way, assessment is simply a major component of instruction itself. Teachers can receive feedback from students in multiple ways, many of which involve informal types of assessment. The key,

for learning purposes, is the continual process of collecting feedback from students, and then acting quickly on that information to enhance learning. Doing this well, and remembering to do this continually, is not an easy process for teachers to incorporate in their teaching routines if they are not socialized into this practice. So a major implication for instruction and student learning involves teachers becoming adept and comfortable in collecting continuous feedback on student learning and then acting on this information to enhance learning.

Wiliam (2007/2008) offers a number of important suggestions for incorporating assessment for learning into ongoing teacher practices. These recommendations include (a) limiting new ideas to two or three classroom innovations at most at any time; (b) allowing sufficient time to implement and get comfortable with teaching innovations; (c) developing detailed action plans for implementing innovations; (d) working together with other teachers as local groups to implement and evaluate changes; (e) developing clear structural formats for teacher group meetings; and (f) meeting for two years (once per month) to work through changes. These suggestions would seem to be good advice for any instructional changes that are intended to be long term.

On a small scale, a teacher or group of teachers can engage in action research projects to explore assessment for learning options and their impacts on student learning, perhaps even as preliminary steps before making any long-term commitments (see Grabe & Stoller, 2002). Projects might include ways to teach students self-assessment of reading comprehension or task performance. They might involve ways for students to indicate, through some signaling mechanism, that they do not understand or are having difficulties. They may involve student and teacher discussions to identify sources of comprehension difficulty and strategies for addressing these difficulties. Projects may also involve actual assessments that can be a resource for further instruction with students. The point of assessment for learning is not that common assessment formats are inappropriate (e.g,. quizzes, comprehension questions, unit tests), but that assessment formats should be means for learning about difficulties and then working to address these difficulties. The key for teacher exploration is how to turn this information from assessment formats into effective learning opportunities.

Informal assessments often provide important opportunities for student learning. Usually there is a product or a performance that students and teachers can work on together, pointing to specific skills and strategies that can improve learning. Students can explore how to improve their performance for future cycles in concrete ways. Again, action research projects specifically focused on how to use informal assessments in support of learning would be a way to build teacher expertise and student learning.

Two final issues that deserve mention involve (a) the relationship between learning activities and assessment activities in the classroom, and (b) the relationship between informal assessment and formal assessment in the classroom. In the case of the former, can all learning activities in a class also become assessment opportunities? If so, how; if not, why not? Conversely, can all assessment activities also become learning activities? In the case of the link between informal and formal assessment, when might this linkage be emphasized and how might the two be linked up in ways that support effective feedback to students? These are both issues that would make useful and productive action research projects for a group of teachers to explore. Neither has been discussed extensively in either assessment or teacher training literature, but they would be very useful topics to incorporate into future teacher-development practices.