CHAPTER 3

Assessing Comprehension
What, How, and for What Purpose

Overview

Schools are expected to teach children to read, listen, and watch a variety of visual presentations, and students are expected to do these in order to learn new things. At the heart of reading, listening, and viewing (indeed their very soul) is that incredibly complex process called comprehension. Reading, listening, and viewing basically involve two skills: decoding written, spoken, or visual stimuli, and comprehension. Decoding can be assessed in relatively uncomplicated ways, but “there is no uniform comprehension process to be measured” (Kintsch & Kintsch, 2005, p. 86). As indicated in Chapter 1, comprehension varies, depending on the student, the text, and the context of the situation.

Comprehension is something that can be examined only indirectly. We cannot actually see what is occurring in a student’s head as he or she comprehends. We can only observe “indirect symptoms and artifacts of its occurrence” (Pearson & Hamm, 2005, p. 19). If a student correctly answers a question, we infer that...
comprehension has occurred. Similarly, if a student composes a coherent summary, accurately fills in a diagram, or correctly completes a problem, we make the assumption that the student has comprehended. Because we can only observe the products of comprehension, “any attempt to assess reading comprehension is inherently imperfect” (Francis, Fletcher, Catts, & Tomblin, 2005, p. 376).

As described in Chapter 1, comprehension is an extremely complex entity. “There are many different processes entailed in the broad thing called ‘comprehension,’ and ‘comprehension’ proceeds very differently for different kinds of text, different topics and different reading purposes” (Duke, 2005, p. 93). However, we often talk of comprehension as if it were relatively straightforward, simple, and unitary. Educators discuss students’ comprehension scores on standardized tests. Some districts list comprehension as an item on the report card. Students are referred to as good or poor comprehenders. Many formal measures, as well as classroom activities, use general titles such as “reading,” “language arts,” “social studies,” and “math” to describe their intended purpose. This masks the specific purpose of the activity and often leads to confusion in interpretation of results with regard to comprehension. Referring to comprehension in these ways suggests that it is a single entity that can be assessed as such. Nothing could be farther from the truth.

Assessing What?

What should educators assess in regard to comprehension? This is intimately connected with how we have defined comprehension. Educators need to know exactly what they assess as comprehension; however, this has always been a complex and inexact science.

In regard to schools and classrooms, comprehension is generally defined as a series of subskills. Pearson and Hamm (2005) describe the genesis of this definition as emanating from the work of Davis (1944), who conceptualized comprehension as involving two major factors: word knowledge and reasoning. Davis sorted these factors into the following general categories: remembering word meaning; recognizing word meaning in context; understanding explicitly stated content; connecting ideas; drawing inferences; following passage organization; recognizing literary devices; and understanding the author’s purpose, mood, and tone. Although Davis focused on reading comprehension, the same skills can also be applied to listening and, with some modifications, to viewing.

Over time, the lists of comprehension skills grew longer and longer and evolved into what was referred to as scope and sequence charts. Such charts were very lengthy and could include as many as 30 different comprehension skills (Pearson & Hamm, 2005). However, these skills were seldom discrete entities.
For example, finding the main idea is very similar to determining importance, although both might be on the same chart. Making predictions and drawing inferences were often listed as different skills, but a prediction is a form of inference. And are drawing conclusions and making judgments really different, or are they both facets of the same component, inferential comprehension? The scope and sequence charts became the basis for comprehension testing. Each skill was repeatedly assessed and reassessed at different levels until a student attained some semblance of mastery. Test items required a single “right” answer, even after educators acknowledged that experience and prior knowledge could influence a student’s response. Such testing required complex management systems to facilitate classroom and district implementation.

Lengthy scope and sequence charts have not gone away. Many districts have simply recycled them, using different terminology. Skills are now called performance expectations, benchmarks, learning objectives, or grade-level targets. However, they are still confusing and impossible to keep track of in regard to individual students. What do I mean by this? Consider a classroom of 30 students. Suppose the teacher is expected to address and assess 30 grade-level targets carried over from the old scope and sequence chart. This means that the teacher has to keep track of 900 different points of assessment, an impossible task. For this reason and despite district efforts to tie these performance expectations to state standards, I suspect that many have gathered dust on classroom shelves. Comprehension of subject matter text has been similarly confusing. Some state standards for comprehension of science and social studies, for example, are extremely broad; others are quite narrow and delineate a large number of specific facts that should be learned at a particular grade level. Textbooks compound the problem. The number of facts included within a single chapter is overwhelming. Are all of these equally important? I suspect not. And although facts are important, there is much more to comprehension than understanding and remembering facts. If educators are to assess comprehension in a more valid and meaningful way, then perhaps we need a different system.

Components of Comprehension

As mentioned repeatedly, comprehension is extremely complex and we have sought ways to simplify it in order to make both instruction and assessment more manageable. Wiggins and McTighe (1998) describe six facets of understanding or comprehension. The first is explanation or understanding of why and how. For example, a student can explain why the ancient Egyptians mummified their dead or how a cactus stores water in order to survive. The second facet is interpretation. A student recognizes the importance or significance of an event or idea. For example, the student is aware of the far-reaching effects of the
Homestead Act on settlement in the Great Plains, or distinguishes family and cultural influences that led Patricia McKissack to become a writer. The third facet is application or the ability to use knowledge in new situations and contexts. A student applies his or her interpretation of the influences on Patricia McKissack’s life to a different author, or applies understanding of the Homestead Act to the effects of other government initiatives. The fourth facet is perspective. A student understands different perspectives on an issue and evaluates the validity of each. For example, the student examines the Treaty of Versailles from both the German and the Allies’ point of view, understands the position of both sides, and evaluates the justice of the treaty in this regard. The fifth facet is empathy for the feelings and views of others. Basically, a student is able to walk in another’s shoes and imagine the thoughts and feelings of others. For example, the student understands slavery and discrimination from the point of view of Biddy Mason (a former slave) or Malcolm X. The final facet is self-knowledge, knowledge about what we think and why we think it. A student recognizes his or her ideas or biases about a specific culture and is able to identify the source and evaluate the validity of these perceptions.

Bloom’s taxonomy of educational objectives (Bloom & Krathwohl, 1956) is a familiar device for breaking comprehension into manageable parts: knowledge, comprehension, application, analysis, synthesis, and evaluation. Knowledge is defined as recall of information, usually factual or literal in nature. Comprehension is the ability to interpret information in one’s own words, and application entails applying learned content to new situations. Analysis is the identification of patterns and relationships. Synthesis involves organizing knowledge from different areas, and evaluation includes making judgments and assessing the value of ideas and theories. This taxonomy has been used for many years to describe different facets of the comprehension process.

Both Wiggens and McTighe (1998) and Bloom and Krathwohl (1956) focus on six separate components of comprehension, and there is some similarity between their categories. For example, both address application and move well beyond the factual or literal level. To put it another way, both systems recognize the importance of doing something with the facts that were understood. However, the systems are more different than alike. Yet because comprehension is such a complex entity, both probably tap into the comprehension process in some way.

Can educators use these two systems effectively to describe the comprehension of their students? Of course. However, the issue is the ease with which this might occur. Six components may be too many to handle efficiently in the world of the busy classroom. In addition, distinguishing between the separate components is not always easy. The line between interpretation and perspective, or application and evaluation, for example, often tends to blur, and many
activities actually represent more than one component. No matter how well intentioned and dedicated an educator is, describing a specific comprehension assessment in terms of six separate elements may be too unwieldy for effective long-term use.

The RAND Reading Study Group (2002) offers a more efficient taxonomy and differentiates between three components of comprehension: knowledge, application, and engagement. What does the student know? Can the student use this knowledge to do something new? Is the student engaged or actively involved? It is interesting that Wiggins and McTighe (1998), Bloom and Krathwohl (1956), and the RAND Group (2002) all identify application as an important component of comprehension, which seems to support Kintsch’s (2005) definition of learning as application.

An examination of question taxonomies (explained in Chapter 4) suggests that the categories of literal, inferential, and application may represent a viable system for describing comprehension in the classroom. Dividing comprehension into literal, inferential, and application presents more specificity than simply calling an assessment activity “comprehension.” Three components are more manageable than the six proposed by Wiggins and McTighe (1998) and Bloom and Krathwohl (1956) and offer educators something that is more controllable than recycled scope and sequence charts. Elements of state standards can easily be tied to these three components. If a teacher or school uniformly describes student performance in terms of a manageable number of chosen components, comprehension assessment becomes more meaningful and more manageable.

Verbs as Indicators of Comprehension Components

How does a teacher know if a comprehension activity is literal or inferential in nature, or if it involves application? It is relatively easy to identify literal comprehension; the student answers a question, offers a comment, or identifies a concept that has been explicitly stated in the text. Unfortunately, differentiating inferential and application components is not as clearcut, and teachers do not have time to engage in lengthy examinations of questions or activities to determine which component of comprehension is best represented. However, we do have a guideline that, although not perfect, makes a relatively workable distinction between inferential and application comprehension, at least as far as we are able to take it at this point in time. We have verbs that describe what students are supposed to do, words like describe, analyze, evaluate, compare, justify, and so forth. A teacher can use these performance or question words to determine whether a question, activity, or student comment represents inferential or application comprehension. There may still be times when a teacher is unsure, but
applying the guideline allows one to be right more often than wrong in the long run. Remember, comprehension assessment is still an inexact science, but differentiating it according to three manageable components is better than some past and present practices.

**Text and Topic**

In addition to choosing a workable model of comprehension components, we also have to consider the underlying text and the topic of that text. Is the student reading or listening to a narrative? Is the student grappling with expository material? Is the student viewing a video of familiar or unfamiliar content? Many students comprehend well in reading narratives but falter when dealing with unfamiliar expository material. Caldwell and Leslie (2004/2005) described proficient middle school readers who performed very differently with narrative versus expository text. They were able to effectively retell and answer questions about a narrative describing a family’s capture during the Vietnam war, but failed miserably to comprehend an expository selection on viral replication. The difference between comprehending a family’s struggle in a war and comprehending viral replication points out that, in addition to the text genre, topic is another issue in assessing comprehension.

Comprehension can also depend on familiarity and knowledge. It is often assumed that there is a single grade level for a written text and that fifth graders, for example, should be able to read, listen, and/or understand all text written at that level. This is an erroneous belief. I have asked many students to read passages on familiar and unfamiliar topics at the same grade level. Students who comprehended a narrative selection on the familiar topic of Abraham Lincoln evinced serious difficulties in comprehending an expository passage on temperature and humidity. Similarly, fourth graders who read and comprehended a passage on Johnny Appleseed struggled with a fourth-grade selection on plant structure. Is it fair to describe a student as having difficulties in comprehension on the basis of performance with a text structure and/or topic that is relatively unfamiliar? In reality, we can expect that students may comprehend very differently in literature, social studies, math, and science.

We do not presently have, nor can we ever construct, a single assessment that adequately measures something as complex and multifaceted as comprehension. There is no such thing as a single measure of comprehension that we can administer, score, record in a grade book, and relate to parents. Our assessment of comprehension must necessarily involve multiple samples of different kinds of behavior. When we choose or construct a test or assessment activity, we must determine what components of comprehension we are assessing and qualify the word *comprehension* with appropriate modifiers such as *literal*, *inferential*, and
application. This applies to published instruments as well as assessments constructed by the teacher. Knowing exactly what an assessment instrument measures is the first step toward establishing validity. Perhaps the most difficult task facing educators regarding comprehension assessment is to stop thinking and talking about it as a single entity and instead attach appropriate modifiers to the word comprehension.

Assessing How?

We all know how fast some plants grow. The seed germinates and the first leaves poke through the soil. The plant grows taller and bushier, blossoms appear, and finally there is a tomato or squash ready to be picked and enjoyed. As we weed and water our garden, we note changes in the plant but we do not actually see the process unfold. However, time-lapse photography can record the entire progression and reduce the time so that over a matter of minutes, we can watch a growth process that spans several months. Unfortunately, we cannot do the same with comprehension; we cannot see it as it happens. There is no way we can get into a person’s head and watch what occurs when comprehension “clicks.” We have to be content with “residual traces of the comprehension process—indirect indexes of its occurrence” (Pearson & Hamm, 2005, p. 62).

In Chapter 1, I discussed how comprehension varies, depending on the knowledge and skill of the comprehender and the structure and difficulty level of the text. Comprehension also varies according to how we measure it, according to those indirect indices of its occurrence that we choose to assess its presence or absence. Do we ask students to retell or answer questions? Do we expect students to answer questions from memory, or do we allow them to locate the answers in the text? What form do our questions take? Do we ask for factual or literal recall, or do we ask higher-level inferential questions? Do we use formats such as multiple choice, fill in the blanks, and true/false, or do we assign essay questions? Do we assess comprehension by evaluating research papers, group projects, lab experiments, or oral presentations?

How we assess comprehension also relates to our methods of scoring the assessment instruments that we choose. If we choose questions as our index of comprehension, do we weight all questions the same? Or do we assign more weight to questions that tap higher-level comprehension skills such as inferring and applying? If we assign essay or open-ended questions, do we construct a rubric to guide our scoring, thus establishing a measure of reliability for the process and ensuring that all students are evaluated in a relatively consistent way?

How we assess and score comprehension can have a powerful effect on the process itself and may underestimate or overestimate a student’s comprehension. I used to regularly employ true/false items in a graduate class until I overheard
one student say, “If she had asked me to write the definition, I could have done
this, but I got all mixed up with true and false. I hate true and false questions!” I
took the student aside and asked her to define the concept in question and she
did so quite adequately. That was some years ago, and it taught me a valuable les-
son. Not all comprehension assessments are equal.

Questions

There are many ways to assess comprehension. Perhaps the most common test-
ing format is asking students to answer questions. However, questions take many
forms and not all are similarly sensitive to the existence of comprehension. A stu-
dent may correctly answer a multiple-choice question by clever guesswork with-
out ever really understanding what was read. With true/false questions, students
have a 50% chance of being correct. There are questions that students can
answer just from their own knowledge without ever reading or understanding
the text. Educators will always ask questions to evaluate comprehension. The
trick is to know what a student’s answer tells us about his or her comprehension.
Has the student comprehended or learned? Does the question tap literal under-
standing of facts or understanding at a higher inferential level? Chapter 4
includes guidelines for using questioning as an effective measure of comprehen-
sion assessment.

Open-Ended Assessment

Another index of comprehension is gained by evaluating students through open-
ended tasks such as answering essay questions; doing science experiments, pro-
jects, and presentations, and developing portfolios. By open-ended I mean that a
student must construct something, as opposed to choosing a correct answer from
two or more choices. We can learn a lot more from open-ended assessments than
from multiple-choice and true/false questions, but only if we know exactly what
we are looking for. An open-ended assessment probably encompasses all three
components of comprehension—literal, inferential, and application—and it
may be impossible to separate these during the assessment process. However, an
educator can decide which of the three is most clearly represented by the assign-
ment and use this for a more fine-tuned description of a student’s comprehen-
sion. Scoring open-ended assessments presents challenges. A true/false question
is either right or wrong, but 25 students turning in the same essay question can
represent 25 very different interpretations of the assignment. Moreover, student
performance variables can obscure comprehension. A student may understand
the material quite well, but difficulties with writing, organization, or oral presen-
tation can suggest the opposite. Chapter 5 addresses guidelines for describing
and assessing comprehension through open-ended activities.
Student Dialogue

Teachers listen to students every day and, by so doing, can learn quite a bit about their comprehension or lack of it. Unfortunately, much dialogue in the classroom is teacher centered and offers students few opportunities for discussing what they have read, heard, or viewed. Fostering student-centered dialogue through practices such as thoughtful talk and thinking aloud provides another window on the comprehension process, and students’ dialogue can be evaluated according to the nature of their comments: literal, inferential, and application. For many teachers, this represents a dramatic shift in how they structure classroom dialogue. Chapter 6 addresses guidelines for fostering student-centered dialogue and using it to assess comprehension.

Word Comprehension

There is a strong relationship between word knowledge and comprehension. To put it simply, people who know more word meanings tend to comprehend more effectively. Although there is almost uniform agreement that word knowledge should be taught in all classrooms, little attention has been paid to the assessment of vocabulary knowledge. Good assessment is always tied to instruction, and the same activities recommended for teaching vocabulary can be used to assess it. As demonstrated in Chapter 7, these same activities can be designated as literal, inferential, or application in nature, thus tying word knowledge to the comprehension process.

Comprehension Proxies

There are also a variety of assessments that we can label comprehension proxies. A proxy is an individual appointed to act for another; basically, a proxy is a substitute. Popular comprehension proxies are measures of word recognition and oral reading fluency. These are used as proxies because they are strongly correlated with comprehension. Correlation is a statistic that shows a relationship between two sets of scores. Two correlated variables share something in common. Obviously, the ability to accurately and quickly recognize words is necessary for comprehension but, as indicated in Chapter 1, it is not the whole of comprehension. Because assessments of word recognition and fluency are relatively easy to administer and score, it is tempting to use these as comprehension proxies. There are other measures, such as the cloze technique and the sentence verification technique, that can also be called proxies. In using the cloze technique, students fill in missing words in a piece of text. With the sentence verification technique, they indicate whether or not a specific sentence was present in the text they read or listened to.
There are problems with the use of comprehension proxies no matter how attractive the process may seem. First, they are, at best, only an indirect measure of the comprehension process. The ability to fill in a missing word sheds little light on a student’s ability to understand facts, draw inferences, and/or apply knowledge. Second, educators often interpret correlation as establishing a causal relationship, and they regard word recognition and fluency as causes of comprehension. Correlation establishes the existence of a relationship but not a causal one. There are students who read accurately and fluently and still demonstrate comprehension problems. Comprehension proxies may have a place in the assessment cycle, but they are not substitutes for more direct assessments of comprehension. Comprehension proxies are addressed in more detail in Chapter 8.

Assessing for What Purpose?

The basic purpose of any assessment is to gather information in order to make a decision of some sort. In regard to comprehension assessment, the primary purpose is to gather information in order to decide if students are comprehending and learning. Farr (1992) states that “the bottom line in selecting and using any assessment should be whether it helps the students” (p. 46). Comprehension assessment should assist educators in evaluating the effectiveness of schools, in fostering student growth, and in designing and planning instruction.

School Effectiveness

One purpose of comprehension assessment (and the one receiving most emphasis at present) is to judge the effectiveness of schools. The No Child Left Behind (NCLB) Act, signed into law in 2002, requires that schools document the adequate yearly progress of their students. The instruments chosen to do this are formal standardized tests. These are the measures most often reported in the media and responded to by the general public. These are the measures that possess face validity in that they are generally (and perhaps unfortunately) accepted as valid measures of student comprehension and learning. But are they? To what extent do standardized measures actually assess comprehension? Chapter 10 focuses on what standardized measures tell us and what they don’t.

Assessments using formal standardized measures are becoming more and more high-stakes evaluations in that state and federal approval and funding depend on acceptable levels of performance by students. Unfortunately, at the present time few schools or districts have any alternatives to the standardized tests. That is, if student performance is suspect on the standardized measure, there are no schoolwide alternatives based on state standards, district goals, grade-level curricula, or classroom expectations. Even if schools or districts have
implemented schoolwide assessment programs, rarely are the data summarized and disseminated to the public in such a way as to counter the standardized results. Or, to put it another way, rarely are the data used to build a case for the face validity of less formal assessment instruments.

**Student Learning and Progress**

Another purpose of comprehension assessment is to measure the learning of students and to monitor their progress across a school year. Such assessment is at present very idiosyncratic across teachers and schools. Although schools and districts may construct a uniform plan of what to assess based on state standards, how this is translated into classroom practice generally varies from teacher to teacher and from grade to grade. Measuring learning and monitoring progress are tied to grading practices, and there is no agreed-upon system for designing assessments or assigning a specific grade (Marzano, 2000).

Some schools and districts mandate certain assessments at specific grade levels, such as the use of running records at the primary level. Content teachers often use identical unit tests to assess student learning. However, for the most part, teachers in different classrooms design different assessments to measure comprehension. This lack of uniformity is both a strength and a weakness. It is a strength in that it dignifies the knowledge, professionalism, and creativity of individual educators. It is a weakness in that the reliability of assessments across a school or grade level is suspect. Chapter 9 focuses on classroom assessment and grading practices.

**Designing, Planning, and Improving Instruction**

Using assessment of student performance to plan instruction is formative in nature and measures teacher effectiveness more than student learning. The teacher assigns a short activity to ascertain whether certain material needs to be retaught or presented in a different way. Allied to this practice is the provision of feedback to students, which is closely allied to the promotion of student self-reflection on performance. “What did I know?” “What am I still confused about?” “How can I improve?” These are questions that are seldom tied to the evaluation process in our schools. Typical classroom practice provides feedback in the form of a single grade. Students know the grade they received but not often why they received it. Assessment should be a collaborative activity involving both teacher and students. Unless the assessment actively includes the student, the following scenario is all too common: A teacher carefully evaluates a paper or project and provides extensive comments to help each student to understand his or her strengths and weaknesses. When such papers are returned to the students, many totally ignore the comments, look at the grade and toss the paper into the trash.
Assessing Comprehension

Pulling It All Together: Sampling and Categorizing

How can we pull together what we know about comprehension assessment in a workable fashion? That is, how can we address the what, how, and purpose of comprehension assessment in such a way as to offer specific guidelines to students, parents, and the general public, as well as to the educator? Whatever method we choose will be imperfect and inexact, but it can only improve upon the present methods of assessing and scoring that elusive and complex entity called comprehension. We can design and implement a workable system for assessing comprehension by using two components that are integral parts of our everyday life: taking samples and categorizing objects, events, and experiences.

Using Samples

Comprehension assessment relies on the use of samples of performance, but samples that are clearly labeled as to what is being assessed. Perhaps an analogy will help. Suppose you visited a new restaurant and did not enjoy the taste or presentation of the food. If a friend asks you about your experience, you probably reply, “The food wasn’t too good.” As a result, your friend may decide to avoid that particular establishment. Now, was your assessment a complete or fair evaluation of the restaurant’s food? After all, you tasted only your selected salad, a single entrée, and the dessert you chose. Could it be that other salads, entrées, and desserts might have been very different? In order to evaluate the total menu, you would have to taste all the foods on the menu. This would be a massive task, and even then it may not represent a fair appraisal. Different cooks cook on different nights, and ingredients can vary in their quality and freshness. A dish ordered on Saturday during a peak time may be very different from the same dish ordered on a relatively calm night. It is probably not realistic or possible to completely assess all the dishes offered by the restaurant. Instead, when we evaluate restaurants, we use samples to determine if the restaurant meets our expectations. We also naturally group our samples into general categories such as appetizers, salads, entrées, and so on. We might report on a dining experience by praising an appetizer, criticizing an entrée, and going into raptures over a dessert. And because we inherently understand this idea of sample usage, even if we have a bad experience on one night or in one category, we generally return to give the cook another try.

Our lives are basically made up of samples and we inherently understand this. A sample is a small part of a whole; it may or may not reflect the quality of the whole. We use samples to make decisions about the wholes they represent. In other words, samples of the world around us provide data, information, or evidence for the decisions we make. The traffic on the freeway at 6:30 in the morning, the attitude of a salesperson late in the day, the reaction of a coworker to a
national issue, and the music we hear on a specific radio station are all samples, and we react to them in a variety of ways. Usually we evaluate several samples before making a judgment and taking action. For example, we may have a first impression of a new coworker, but we generally interact with that individual over several instances before deciding that the person is someone we might like as a friend. We may take the same route to school several times before deciding that the traffic is too heavy and choosing an alternate path. We probably allow for one or two instances of rude salespeople before we decide to avoid a store. Generally, we take samples, look for patterns, and make decisions based on these patterns, not on single instances. Of course, sometimes a single sample is so pleasant or unpleasant that our decision is formed immediately. I well remember my first taste of anchovies and I have never attempted them since.

We need to carry our everyday practice of sample analysis into comprehension assessment. Instead of looking at a test or activity as representing a student's comprehension, we should regard it as only one sample of comprehension. It may be an accurate sample, but it may not be. For example, think about high-stakes testing and the pressure put on students by both parents and teachers. This can lead to a high level of anxiety, which may have a very negative effect on comprehension. Consider a group project in which negative group dynamics hinders comprehension. I remember when mandated certification tests in my state moved to a required computer format. At that point in time, many students found this awkward and unfamiliar and their scores indicated their unease. Because so much can go wrong with a single assessment, collecting multiple samples makes good sense.

Our role is to look for patterns in multiple samples of comprehension behaviors. But don’t educators already do this? Not if they label an activity or assessment “comprehension” without categorizing it in some more specific way. Not if they describe it in a grade book as “Chapter 4 Test.” Not if they use one form of assessment such as mandated standardized assessments as the primary basis for deciding about a student’s ability. An educator who takes the grades of all tests, projects, worksheets, and homework assignments and averages them for the report card is basically saying that all are of equal importance and all reflect the comprehension process in a similar way. Given the contents of Chapter 1 and the first part of this chapter, we should actively question the validity of that assumption.

Categorization

Using samples for decision making works only if it is combined with categorization. Categorizing objects, events, and emotions is a natural cognitive process, so natural that we are not even aware of doing it. We categorize samples as good or bad, needed or unneeded, acceptable or unacceptable, tasty or nasty, convenient or inconvenient, and so forth. This natural categorization process is an integral
part of our decision-making procedure. And so, if we are to make a decision about the quality of a student’s comprehension, we need a system for categorizing comprehension samples. We need a simple coding system that can allow educators to detect patterns of comprehension behavior. When an assessment sample is entered into a grade book, it should be accompanied by a code that specifically categorizes it. Many grade books use one- or two-word descriptions of an assessment activity, such as page number, unit/chapter title, or a brief description. Although these suggest the subject matter topic, they are inadequate to categorize that complex entity called comprehension.

What are the basic elements of comprehension categorization? Consider some possible elements and whether they are workable for school and classroom assessment. The first is mode of input. Was comprehension assessed during or after reading, listening, or viewing? In most cases, it is impossible to separate these components. During an instructional unit or classroom activity, a teacher may ask students to read segments of the textbook and the print on a media presentation, view illustrations and computer simulations, and listen to the teacher’s explanations and the comments of their peers. How can this complexity be easily categorized as representing primarily reading, listening, or viewing? It can’t. In such cases, categorizing the mode of input is clearly unworkable.

However, there are instances when categorization of input may be important. Differentiating the mode of input is important when assessing students who are struggling academically. This usually occurs when a student is not reading at the level of his or her peers. Because listening comprehension is strongly related to reading comprehension (Daneman, 1991; Sticht & James, 1984), it makes sense to determine if the listening comprehension process is similarly impaired. Many students who are learning English as a second language can understand through listening much more than they can through reading, and it makes good sense to separate the two. In fact, students who are English language learners are often given the option of taking tests in an auditory as opposed to visual format. In summary, whereas an assessment specialist will probably differentiate mode of input, a classroom teacher will generally not do so.

Another comprehension category is type of text, narrative versus expository. For content teachers who teach social studies, science, math, or other specific subjects, this is not an issue. Almost all text used in a content classroom is expository in nature. However, English teachers and reading teachers should pay attention to possible comprehension differences in narrative versus expository material. A variety of genres are presented in English and reading classes, and it is important to differentiate narrative from expository. Many students demonstrate much higher comprehension after reading a narrative selection than after reading expository text. Assessment should inform instructional planning, and this can suggest a need for greater attention to comprehension of expository text structure.

Another category is familiarity or unfamiliarity of the topic. Again, for
math, social studies, science, and other subject matter disciplines, this is probably not an issue. Most, if not all, text is unfamiliar in nature. However, in the reading or language arts classroom, familiarity can be an issue. I recall watching a teacher read a story to a second-grade class of urban children and assess their understanding through discussion. Unfortunately, the story was about a farm, and none of the children had any idea of the nature and purpose of a farm. Although some had gone to the zoo and watched monkeys and lions, none had seen a pig, chicken, or cow. None understood what such animals eat or how they must be cared for. Their comprehension reflected this lack of familiarity. Another time, in the same classroom, a totally different picture emerged as the children eagerly discussed their reactions to a story about washing the family car. If the teacher intended to keep some record of both experiences, one should certainly have been categorized as unfamiliar and the other as familiar.

Yet another category is the format of the assessment activity. Although there are many ways to categorize assessment formats, I suggest that, for the purpose of describing comprehension, teachers focus on the following two kinds: selected response assessments and constructed response assessments. Selected response items are often referred to as objective assessments. The student is presented with several possible choices for an answer and chooses one. Most standardized instruments use the selected response format. Constructed response assessment requires that the student construct something. It can be a written answer, an oral summary, or a project of some sort. We often call such assessments open-ended or subjective assessments. All assessments can fall into these two broad categories, which allows for a simple form of categorization. Did the students select an answer or construct the answer? Very often, teachers find very real comprehension differences across the two kinds of assessment formats.

But what about an assessment that utilizes both formats, for example, a unit test that employs a series of selected response questions with several constructed response questions? It is important to separate the two when analyzing student performance. They represent different forms of comprehension, and we cannot assume that a student will do well in both. This subject is discussed in more detail in later chapters.

In addition to the aforementioned elements of comprehension categorization (mode of input, type of text, familiarity, and format), educators should choose a limited number of components to describe the nature of the comprehension process itself. Perhaps the simplest way is to categorize comprehension as literal, inferential, and application.

The following table suggests categories of comprehension with simple coding symbols that can be easily entered into a grade book. Note that the categories Assessment Format and Comprehension Component are listed first and are italicized. This is to stress that these are the basic and required codings for all assessments. Mode of Input, Type of Text, and Familiarity may have relevance for
some teachers in some specific situations, but differentiating comprehension in terms of assessment format and comprehension component is basic to detecting patterns of comprehension in all students. This allows an educator to move away from thinking of comprehension as a unitary entity and provides modifiers for educators to use in describing comprehension performance.

Categories of Comprehension

<table>
<thead>
<tr>
<th>Assessment format</th>
<th>Comprehension component</th>
<th>Mode of input</th>
<th>Type of text</th>
<th>Familiarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRF—Selected response format</td>
<td>L—Literal</td>
<td>R—Reading</td>
<td>N—Narrative</td>
<td>F—Familiar</td>
</tr>
<tr>
<td>CRF—Constructed response format</td>
<td>I—Inferential</td>
<td>Ls—Listening</td>
<td>E—Expository</td>
<td>U—Unfamiliar</td>
</tr>
<tr>
<td></td>
<td>A—Application</td>
<td>V—Viewing</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>C—Combination</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The following table summarizes categories that may be applicable for different educators.

Assessment of Comprehension Components

<table>
<thead>
<tr>
<th>Assessment specialist</th>
<th>Mode of input</th>
<th>Type of text</th>
<th>Familiarity</th>
<th>Assessment format: SRF, CRF</th>
<th>Comprehension component: Literal Inferential Application</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading teacher</td>
<td>Type of text</td>
<td>Familiarity</td>
<td>Assessment format: SRF, CRF</td>
<td>Comprehension component: Literal Inferential Application</td>
<td></td>
</tr>
<tr>
<td>English teacher</td>
<td>Type of text</td>
<td>Familiarity</td>
<td>Assessment format: SRF, CRF</td>
<td>Comprehension component: Literal Inferential Application</td>
<td></td>
</tr>
<tr>
<td>Content specialist</td>
<td></td>
<td></td>
<td>Assessment format: SRF, CRF</td>
<td>Comprehension component: Literal Inferential Application</td>
<td></td>
</tr>
</tbody>
</table>
Suppose an educator assessed comprehension following a science unit using a traditional question-based unit test. Using this coding system, the educator might describe the assessment as usual in the grade book (e.g., Test Chapter 6 or Test Structure of Matter) but append the following: SRF (the assessment employed selected response items such as multiple-choice and true/false) and L (the assessment primarily involved literal or factual comprehension).

Consider another example. The English or reading teacher assesses students’ comprehension following discussion of a short story and records this by coding N (it was a narrative), U (it was on an unfamiliar topic or theme), CRF (the assessment employed an open-ended constructed response activity), and L/I (the assessment involved both literal and inferential comprehension). Of course, in the grade book, there would be two columns: one for a literal score and one for an inferential score.

When it is time to enter report card grades, talk with students, or confer with parents, the educator identifies any salient patterns. An assessment specialist would probably note any clear differences in reading and listening comprehension. Students who are English language learners often demonstrate better listening and viewing comprehension than reading comprehension. Young children who are just beginning to make sense of letters and sounds also show better listening comprehension. In fact, it is important to note this, because poor listening comprehension can be an indication of problems with language development in general.

The assessment specialist, the reading teacher, and the English teacher would note any comprehension differences between narrative and expository text. A student who displays acceptable comprehension with narrative text but falters with expository material is not a poor comprehender. This student is a poor comprehender of a certain kind of text, but comprehension in general seems intact. I would certainly look like a poor comprehender if you assessed my ability to interpret income tax directions. For content teachers of math, social studies, and science, almost all of the written or oral text used in the classroom is expository in nature, so this level of analysis can be skipped.

The assessment specialist and the reading teacher would also note any differences between familiar and unfamiliar material. Inability to comprehend familiar text is more serious than inability to comprehend unfamiliar topics. For many content areas, all subject matter is basically unfamiliar, so this level of analysis can be skipped.

All educators should determine whether a student experiences difficulty with an assessment format. Some students do well with selected response questions but experience difficulty with constructed response assessments’ essay questions. Some students can answer multiple-choice questions but are not able to conduct a science experiment or coordinate an open-ended project. At this level of analy-
sis, you are identifying information that not only describes your students but also suggests areas for instructional emphasis.

Finally, which comprehension components showed a student’s strengths? Which ones need more instruction? Again, you are not only determining a student’s strengths and weaknesses, you are also discovering how to focus subsequent instruction.

At first glance this process seems unwieldy and, like any new endeavor, it takes a bit of getting used to. In the beginning it may seem quite time-consuming, but as you become familiar with the process, it takes much less time. However, this process offers a workable system for categorizing multiple samples of a student’s comprehension behavior. It allows the educator to use general but meaningful descriptors to specify comprehension. I return to the notion of sample categorization again and again in succeeding chapters.

Comprehension Assessment and Response to Intervention

Response to intervention (RTI, Fuchs & Fuchs, 2006) will no doubt have a profound influence on an assessment of student comprehension, learning, and progress. The passage of the Individuals with Disabilities Education Act (IDEA, 2004) allowed schools and districts to classify students as learning disabled on the basis of their response to instruction. In the past, classification of learning disability rested on an identified discrepancy between level of intelligence, as measured by a standardized IQ test, and level of functioning in listening, thinking, speaking, reading, writing, spelling, and/or mathematical calculation. That is, a child with normal intelligence who was functioning below the level of his or her peers in reading, writing, or math, for example, could be considered learning disabled, and the school or district would receive federal funding to provide instructional intervention. Today, the identification of learning disability can rest on lack of student response to instruction. “Response to Intervention is, simply put, a process of implementing high-quality, scientifically validated instructional practices based on learner needs, monitoring student progress, and adjusting instruction based on the student’s response. When a student’s response is dramatically inferior to that of his peers, the student may be determined to have a learning disability” (Bender & Shores, 2007, p. 7).

RTI is often conceptualized as embodying three tiers. The first is classroom instruction. Assuming the presence of effective classroom instruction and valid and sensitive classroom assessment, lack of student progress can signal a need to move to the next tier. Tier 2 provides focused intervention in addition to regular classroom instruction. It can be delivered by the classroom teacher or by person-
nel external to the classroom such as reading specialists or subject matter experts. Such instruction usually involves small-group instruction offered two to three times a week (McCook, 2006). Lack of progress in this tier leads to Tier 3, individualized intervention and the probable identification of a learning disability.

RTI emphasizes “the role of data for decision-making” (Brown-Chidsey & Steege, 2005, p. 18). Student progress is monitored using “short, quick and easy-to-administer probes” (McCook, 2006, p. 14). In Tier 1, universal classroom screening occurs three times a year and identifies those students who are not progressing as well as their peers. In Tiers 2 and 3, the same measures are used more frequently to identify progress as a result of intervention. Classroom screening establishes a baseline for scores, and the performances of individual students are compared to that baseline to determine progress or lack of it.

What forms do the short and quick assessments take? They primarily focus on what I describe in Chapter 8 as proxies, that is, measures that predict comprehension but do not directly assess it. The most popular RTI assessment is oral reading fluency, a measure of the number of words read correctly in 1 or 2 minutes. Other RTI probes assess number of sounds and letters recognized and number of words correctly spelled, all within a 1- to 3-minute time frame (Wright, 2007). If assessed, comprehension generally takes two forms, neither of which parallel what readers, listeners, and viewers do when they comprehend. The student orally reads a passage for 1 minute and then retells it. Retelling is scored as the number of words in the retelling offered within a 1-minute time frame; the accuracy of the retelling is not judged. A second form of assessment is a 3-minute maze task in which the student reads a passage with words deleted and selects a missing word from three or four choices.

At the present time, the RTI assessment and intervention model primarily centers on reading performance (Bender & Shores, 2007; Wright, 2007; McCook, 2006; Brown-Chidsey & Steege, 2005) involving sound-based or phonological measures. Admittedly, these are highly predictive of future reading comprehension performance. However, Gersten and Dimino (2006) state that although such assessment is not inaccurate, “it is incomplete; it is only one part of the picture” (p. 104). There are other nonphonological measures that predict reading comprehension, such as oral language proficiency, expressive vocabulary, and passage recall. However, these do not necessarily lend themselves to short, quick, and easy-to-administer RTI probes.

RTI, as presently conceptualized, seems focused in the elementary school, but there is certainly a chance that the initiative will extend into the middle and secondary levels. When and if this occurs, assessment of comprehension will have to move beyond comprehension proxies and measures of letter and sound identification. Because federal funding is closely tied to RTI, the validity and reliability of comprehension instruction and assessment will be of increased
importance in the next years and subject to a closer form of scrutiny than in the past.

RTI is basically quite new, and it remains to be seen how various schools and districts will interpret and implement the fairly general federal guidelines. As always, the devil is in the details. It is not a purpose of this book to suggest specific guidelines for RTI implementation. However, it is important to examine RTI in relation to classroom comprehension assessment. All teachers basically represent Tier 1, and their instruction and assessment should be sensitive enough to identify children in need of additional and more focused instruction whether or not a classification of learning disability ever applies. In succeeding chapters, I briefly examine the RTI paradigm in regard to comprehension assessment.

**Summary**

The following table provides a summary of the contents of this chapter.

<table>
<thead>
<tr>
<th>What am I assessing?</th>
<th>How am I assessing?</th>
<th>Why am I assessing?</th>
<th>Pulling it together</th>
</tr>
</thead>
<tbody>
<tr>
<td>Components of comprehension: literal, inferential, application</td>
<td>Formal standardized assessments</td>
<td>To evaluate school effectiveness.</td>
<td>Taking multiple samples</td>
</tr>
<tr>
<td>Mode: reading, listening, viewing</td>
<td>Informal classroom assessments</td>
<td>To evaluate student earning. RTI.</td>
<td>Categorizing comprehension</td>
</tr>
<tr>
<td>Text: narrative, expository, familiar, unfamiliar</td>
<td>Selected response assessments</td>
<td>To determine why students struggle academically. RTI.</td>
<td>Coding comprehension categories</td>
</tr>
<tr>
<td>Topic: literature, mathematics, social studies, science</td>
<td>Constructed response assessments</td>
<td>To evaluate instructional effectiveness.</td>
<td></td>
</tr>
<tr>
<td>Questions</td>
<td>To provide student feedback.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student dialogue</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comprehension proxies, RTI</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Professional Development Activities for Increasing Understanding of Comprehension Assessment

- Examine a published test that you have used in the past or are planning to use. Identify sections that assess literal comprehension. Which ones assess inferential comprehension? Are there any that assess application or learning?
- Examine how you code entries in your grade book. Does your system allow you to note patterns across assessment formats and the three components of comprehension?
- Choose the most recent entries in your grade book. How would you code them with regard to assessment format and comprehension component?