Assessing Students and Texts

It’s not as simple as testing. It’s like thinking, if we weigh the cow, the cow’s going to get fatter.
—STATE LEGISLATOR

How effectively are students learning to use reading, writing, talking, and viewing as tools to comprehend and respond to material in content areas? Assessing students and texts to provide this kind of information means that there is a direct connection between teaching and learning, between instruction and the improvement of practice. Assessment in content area classrooms means that students and teachers are actively engaged in a process of evaluation and self-evaluation. Instead of measuring learning exclusively by a score on a standardized test or proficiency exam, the learning process includes assessment of authentic tasks.

Teachers and students want useful assessment; that is, they want to make sense of how and what is taught and learned at any given time. Teachers want to make instructional decisions based on their students’ content literacy skills, concepts, and performance. They must also deal with the very real pressure of state and federal mandates for standards-based curriculum and testing. Yet as the state legislator realized, testing alone cannot yield improvements in student learning.

To understand assessment, you need to differentiate between two major contrasting approaches: a formal, high-stakes approach and an informal, authentic approach. When standards were initially developed by professional
organizations and state governments, testing was thought to be necessary in order to ensure that schools would meet high standards of achievement. Soon, students’ performances on state-mandated tests became the focus of debate among educators, policy makers, and constituency groups. The public’s attention today is often on this formal, high-stakes approach to assessment. Many teachers have become adept at alternative, authentic assessment practices to help them make decisions about instruction appropriate for each student. As depicted in the graphic organizer, portfolios, observations, anecdotal records, checklists, interviews, inventories, and conferences with students are some of the methods and techniques that make authentic assessment possible.

Assessing for instruction should, first and foremost, provide the opportunity to gather and interpret useful information about students as they learn, including their prior knowledge; their attitudes toward reading, writing, and subject matter; and their ability to use content literacy to learn with texts. Through the portfolio assessment process—collecting authentic evidence of student work over time—teachers and students gather useful information about an individual’s comprehension and response to content area material. The organizing principle of this chapter maintains that assessment should be useful, authentic, and responsive to teacher decision making: **Instructional assessment is a process of gathering and using multiple sources of relevant information about students for instructional purposes.**
Teachers sometimes know intuitively that what they do in class is working. More often, however, information for making decisions is best obtained through careful observation of students. Their strengths and weaknesses as they interact with one another and with texts can be assessed as they participate in small groups, contribute to class discussions, respond to questions, and complete written assignments. This approach to assessment is informal and authentic; it is student centered and classroom based. This approach, however, isn’t the only one operating in schools today. If teachers are in a school district guided by standards-based curricula, they need to understand the differences between high-stakes and authentic approaches to assessment.

High-Stakes Testing and Authentic Approaches to Assessment

The two major views of assessment, high-stakes and authentic, are like different sides of the same coin. They represent the almost opposite perspectives of policy makers on one side, and teachers on the other. The policy makers are responding to the public and its demands for assurances that students will leave school well prepared to enter either the workforce or college. Teachers and other educators are calling for better, more authentic assessment practices that will improve instruction and result in learning. As Tierney (1998) put it, one focuses on “something you do to students,” and the other focuses on “something you do with them or help them do for themselves” (p. 378).

Authentic methods often include some combination of observations, interviews, anecdotal records, and student-selected performances and products. The information gained from an authentic assessment can be organized into a rich description or portrait of your content area classroom or into student...
Concerns that emerge, whether about individual students or about the delivery of instructional strategies, are likely to make sense because they come directly from the classroom context and often result from teacher–student or student–student interaction.

Consider how an authentic approach differs from a more formal, high-stakes one. In Table 2.1, the two approaches are compared in several categories. Certainly, there are many gray areas in assessment, where the formal and informal overlap. In this table, however, differences between the two approaches are emphasized. Traditional, formal assessments are product oriented. They are more tangible and can be obtained at predetermined points in time. Authentic assessments are informal and process oriented. The process is ongoing, providing as much information about the student as learner as about the product. Together, they permit a more balanced approach through a combination of traditional, formal and authentic, informal practices. The end result is an understanding of why particular results are obtained in formal assessment, which informs the how of the teacher decision-making process.

**High-Stakes Testing: Some Issues and Concerns**

Never have the stakes been higher. With virtually every state adopting content standards in multiple content areas, such as English, mathematics, social studies, and science, mandated, standardized testing systems have been developed and put in place throughout the United States. Thus, although standardized testing has been used to evaluate student achievement since Thorndike developed the first standardized tests in the early part of the twentieth century, the amount of mandatory testing has increased. And the stakes, significant rewards and penalties, have risen. For example, school districts must follow state regulations, which are written to comply with federal education law. When Ohio legislators recently passed its education bill, $400 million in U.S. Department of Education aid to Ohio schools was preserved!

There are several issues and concerns about the role of high-stakes testing that are being discussed widely. Proponents of high-stakes testing contend that such testing is a sound strategy to use to ensure that standards are met and students are achieving at an appropriate level of proficiency. They seek to effectively end the practice of social promotion, that is, promoting students from one grade level to the next regardless of whether the students have demonstrated on tests the potential to work successfully at the next grade level. In recent years, mandatory tests have been administered to students at younger and younger ages and with greater frequency than ever before (Hoffman et al. 1999).

As the use of mandated, high-stakes testing grows, questions have been posed about the validity of certain assessment tools. In some states, studies have called into question the reasons behind students gains on state mandated assessment, suggesting that students’ improved performances on standardized tests could be attributed not only to achievement but also to factors such as the increased class
time spent on test preparation, students’ growing familiarity with test questions and procedures, and a large proportion of low-scoring students being exempted from taking tests in order to avoid the sanctions connected with low test scores (Koretz & Barron 1998; Klein, Hamilton, McCaffrey, & Stecher 2000). “Teachers are falling into line and teaching to the test not because they agree with instruction that is driven by standardized testing, but because the consequences of low test scores
School leaders have acknowledged the challenge of blending students’ needs with needed scores and have expressed additional concerns about mandated high-stakes tests. Some assert that the tests are not grounded in child development theory, pointing out that students of the same chronological age should not be expected to be at the same point in terms of cognitive development and academic achievement. Others express concern that the high-stakes nature of the tests will result in classroom instruction that focuses more on the drilling of skills and less on the application of knowledge (Mraz 2000).

In response to concerns about the use of high-stakes testing, the International Reading Association issued a position statement that advised against attaching rewards and penalties to single test scores and encouraged the use of multiple measures to inform important decisions as well as the use of assessment tools that honor the complexity of reading (Ransom et al. 1999).

Additionally, no single test can meet the needs of all groups who require information about school and student performance: Different constituencies need different types of information, presented in different forms, and made available at different times. Legislators and the general public may benefit from information provided by norm-referenced tests that are administered on an annual basis. Parents, teachers, and students need information specific to individual students on a more consistent basis. Observations, portfolios, samples of student work over time, and conferences about student progress are more effective than standardized tests in providing that type of information. The purpose of the assessment selected, and the goals for its use, should be carefully considered so that the assessment tool selected will ultimately serve to provide information that can be used to improve learning opportunities for all students (Farr 1992).

A former education advisor at the state level explained, “I think that assessment is something that shouldn’t be a surprise, nor should it be put up as a barrier or hurdle. It’s just part of the process. We need good assessments at all times for all students, and teachers need to be trained in assessment” (Mraz 2002, p. 79).

In some states, initial attempts at establishing assessment programs raised concerns and resulted in programmatic and legislative adjustments. A former education advisor at the state level explained, “Initially, the thinking was that establishing high standards for all [students], would help to improve instruction. In fact, I think the proficiency test was really designed as an early warning system to decide which students needed additional assistance and intervention as they moved through the grades. So, the test was established for one purpose, and then was used for another purpose. It was designed as a warning system, but then it became a pass–fail problem, which became a political problem in that 40,000 students were not successful” (Mraz 2002, p. 81).
State Standards and Accountability

According to the Education Commission of the States (ECS), state policy makers have been actively engaged in setting standards, assessing student reading performance, and imposing consequences for students who do not meet reading standards. In some states, comprehensive plans designed to link standards with assessment have been developed. Here are snapshots of how two states took action.

North Carolina’s ABCs Accountability Model is organized around the goals of ensuring strong accountability, emphasizing the basics and high educational standards, and providing schools and districts with some degree of local control. Schools are rewarded for improvements in student achievement as well as for overall percentages of students performing at or above grade level.

Under the accountability model, public school students are required to meet statewide standards, also referred to as gateways, for promotion from grades 3, 5, and 8, and for high school graduation. Students who do not meet the standards are given opportunities for retesting as well as interventions, such as additional instructional opportunities, extra support in smaller classes, personalized education plans, or increased monitoring and evaluations. The responsibility for designing strategies to reach the state standards and ensuring that constituent groups, such as educators, parents, students, and community members, understand and participate in implementing the standards is delegated to superintendents and local boards of education (North Carolina Public Schools 2003).

In Ohio, recent education reforms, based on the recommendations of the Governor’s Commission on Student Success, a group of educators and community members appointed by the governor, has sought to establish an aligned system of standards, assessments, and accountability (ODE 2003). Changes were made to previous education legislation. Under the 1997 Senate Bill 55, the Fourth Grade Reading Guarantee prohibited school districts from promoting to the fifth grade any student who had not passed the reading portion of the fourth-grade off-year proficiency test (OPT), unless the student was excused from taking the test because of a documented disability or because the student’s principal and reading teacher agreed that the student was academically prepared, as defined in the district’s promotion policy for fifth grade (ODE 1999).

In 2001, Senate Bill 1 resulted in several revisions to the original bill, including a redefinition of the proficiency levels: “Basic” and “below basic” levels of performance were added to the original “proficient” and “advanced” levels of performance on the fourth-grade state reading test, resulting in the lowering of the required pass score from 217 to 198.

Proficiency tests given in kindergarten through eighth grade are scheduled to be replaced by achievement tests aligned to academic content standards and diagnostic tests designed to improve student comprehension of content area standards (ODE 2002). Intervention services will be provided as needed, based on a student’s test scores and classroom performance. New Ohio Graduation Tests, which measure the level of reading, writing, mathematics, science, and social studies skills expected of students at the end of the tenth grade, are in the process of being implemented.
Federal Legislation

In addition to decisions made at the state level, the federal government has played a role in the standards and assessment movement. The Improving America’s Schools Act and Goals 2000: Educate America Act (1994) were based on the concept of standards-based reform, that is, using federal resources to assist states in developing and implementing challenging standards for all students. Under the plan, states and school districts were granted the flexibility to implement programs in ways that met the needs of their students (Congressional Digest 1999). By 1998, new legislation in the form of the Reading Excellence Act included an unprecedented, and, some argued, restrictive, definition of reading and acceptable reading research (Tierney 2002).

In 2002, the No Child Left Behind Act (NCLB) instituted new accountability requirements. Schools that fail to show annual improvement on mandatory assessments risk losing part of their federal funding. Schools that fail to raise test scores over several years could risk being restaffed (Toppo 2001). Critics of the legislation fear that, although high standards of educational achievement are desirable, NCLB could unfairly penalize schools while actually lowering standards as states adjust their proficiency requirements downward in order to preserve federal funding. Additional concerns that NCLB contains a disproportionate number of mandates in relation to the funding offered to schools to fulfill those mandates have also been raised (Maguire 2001).

Standardized Testing: What Teachers Need to Know

Standardized reading tests are formal, usually machine-scorable instruments in which scores for the tested group are compared with standards established by an original normative population. The purpose of a standardized reading test is to show where students rank in relation to other students based on a single performance.

To make sense of test information and to determine how relevant or useful it may be, you need to be thoroughly familiar with the language, purposes, and legitimate uses of standardized tests. For example, as a test user, it is your responsibility to know about the norming and standardization of the reading test used by your school district. Consult a test manual for an explanation of what the test is about, the rationale behind its development, and a clear description of what the test purports to measure. Not only should test instructions for administering and scoring be clearly spelled out but also information related to norms, reliability, and validity should be easily defined and made available.

Norms represent average scores of a sampling of students selected for testing according to factors such as age, sex, race, grade, or socioeconomic status. Once a test maker determines norm scores, those scores become the basis for comparing the test performance of individuals or groups to the performance of those who were included in the norming sample. Representativeness, therefore, is a key concept in understanding student scores. It’s crucial to make sure that the norming
sample used in devising the reading test resembles the characteristics of the students you teach.

Norms are extrapolated from raw scores. A raw score is the number of items a student answers correctly on a test. Raw scores are converted to other kinds of scores so that comparisons can be made among individuals or groups of students. Three such conversions—percentile scores, stanine scores, and grade-equivalent scores—are often represented by test makers as they report scores.

Percentile scores describe the relative standing of a student at a particular grade level. For example, the percentile score of 85 of a student in the fifth grade means that his or her score is equal to or higher than the scores of 85 percent of comparable fifth graders.

Stanine scores are raw scores that have been transformed to a common standard to permit comparison. In this respect, stanines represent one of several types of standard scores. Because standard scores have the same mean and standard deviation, they permit the direct comparison of student performance across tests and subtests. The term stanine refers to a standard nine-point scale, in which the distribution of scores on a test is divided into nine parts. Each stanine indicates a single digit ranging from 1 to 9 in numerical value. Thus a stanine of 5 is at the midpoint of the scale and represents average performance. Stanines 6, 7, 8, and 9 indicate increasingly better performance; stanines 4, 3, 2, and 1 represent decreasing performance. As teachers, we can use stanines effectively to view a student’s approximate place above or below the average in the norming group.

Grade-equivalent scores provide information about reading-test performance as it relates to students at various grade levels. A grade-equivalent score is a questionable abstraction. It suggests that growth in reading progresses throughout a school year at a constant rate; for example, a student with a grade-equivalent score of 7.4 is supposedly performing at a level that is average for students who have completed four months of the seventh grade. At best, this is a silly and spurious interpretation: “Based on what is known about human development generally and language growth specifically, such an assumption [underlying grade-equivalent scores] makes little sense when applied to a human process as complex as learning to read” (Vacca, Vacca, & Gove 2000, p. 530).

Reliability refers to the consistency or stability of a student’s test scores. A teacher must raise the question, “Can similar test results be achieved under different conditions?” Suppose your students were to take a reading test on Monday, their first day back from vacation, and then take an equivalent form of the same test on Thursday. Would their scores be about the same? If so, the test may indeed be reliable.

Validity, by contrast, tells the teacher whether the test is measuring what it purports to measure. Validity, without question, is one of the most important characteristics of a test. If the test purports to measure reading comprehension, what is the test maker’s concept of reading comprehension? Answers to this question provide insight into the construct validity of a test. Other aspects of validity include content validity (Does the test reflect the domain or content area being examined?) and predictive validity (Does the test predict future performance?).
Standardized test results are probably more useful at the building or district, not the classroom, level. A school, for example, may wish to compare itself in reading performance to a state or national norm. Or local districtwide norms may be compared with national norms, a process that is sometimes necessary when a district is applying for federal or state funds. In general, information from standardized tests may help screen for students who have major difficulties in reading, compare general reading-achievement levels or different classes or grades of students, assess group reading achievement, and assess the reading growth of groups of students (Allington & Strange 1980).

However, you need useful information about students’ text-related behavior and background knowledge. You would be guilty of misusing standardized test results if you were to extrapolate about a student’s background knowledge or ability to comprehend course materials on the basis of standardized reading-test performance. Alternatives to high-stakes, formal assessments are found in an informal, authentic approach to assessment. One of the most useful tools for inquiry into the classroom is observation.

**Authentic Assessment: The Teacher’s Role**

In a high-stakes approach to assessment, the *test* is the major tool; in an authentic approach, the *teacher* is the major tool. Who is better equipped to observe students, to provide feedback, and to serve as a key informant about the meaning of classroom events? You epitomize the process of assessing students in an ongoing, natural way because you are in a position to observe and collect information continuously (Valencia 1990). Consequently, you become an observer of the relevant interactive and independent behavior of students as they learn in the content area classroom.

Observation is one unobtrusive measure that ranges from the occasional noticing of unusual student behavior to frequent anecdotal jottings to regular and detailed written field notes. Besides the obvious opportunity to observe students’ oral and silent reading, there are other advantages to observation. Observing students’ appearance, posture, mannerisms, enthusiasm, or apathy may reveal information about self-image. However, unless you make a systematic effort to tune in to student performance, you may lose valuable insights. You have to be a good listener to and watcher of students. Observation should be a natural outgrowth of teaching; it increases teaching efficiency and effectiveness. Instructional decisions based on accurate observations help you zero in on what and how to teach in relation to communication tasks.

Today’s teachers are expected to meet the special needs of all students. Consequently, the challenges of teaching diverse learners in the classroom may cause nonspecialist teachers to feel frustrated and unprepared. Understanding and accepting differences in students can, however, lead to effective instructional adaptations. Here’s how Kim Browne, a seventh-grade teacher of language arts, used observational assessment to help deal with her “inclusion section”:

One of the most frequent questions I’m asked at parent meetings and IEP [individual educational plan] meetings is, “How does my child interact with his or her peers?” I
planned to collect data on each student by using a simple observation checklist when the students are participating in their literary circles after reading *Take me out to the ball game*. I keep an index card file on each student by the class period; my focus is on peer relationships, noting any overt behavior that may be indicative of boredom or confusion, as well as cooperative interactions. Additional observations can be added to a large label stuck to the back of the card.

In addition to the basic format for time sample or interval data, Kim included two other sections: *other information*, where she noted any support the student may be receiving in or out of school or if the student is on an IEP plan and asked a specific question about this student, and *tentative conclusions*, where she made comments about what she just observed and what to focus on in the next observation. Figure 2.1 illustrates Kim’s recent observation on Neil, a student with special needs in her late-morning section.

To record systematic observations, to note significant teaching–learning events, or simply to make note of classroom happenings, you need to keep a notebook or index cards on hand. Information collected purposefully constitutes *field notes*. They aid in classifying information, inferring patterns of behavior, and making predictions about the effectiveness of innovative instructional procedures. As they accumulate, field notes may serve as anecdotal records that provide documentary evidence of students’ interactions over periods of time.

Teachers and others who use informal, authentic tools to collect information almost always use more than one means of collecting data, a practice known as *triangulation*. This helps ensure that the information is valid and that what is learned from one source is corroborated by what is learned from another source. A fifth-grade science teacher recounted how he combined the taking of field notes with active listening and discussion to assess his students’ current achievement and future needs in the subject:

I briefly document on individual cards how students behave during experiments conducted individually, within a group, during reading assignments, during phases of a project, and during formal assessments. Knowing which students or what size group tends to enhance or distract a student’s ability to stay on task helps me organize a more effective instructional environment. When students meet to discuss their projects and the steps they followed, I listen carefully for strategies they used or neglected. I sometimes get insights into what a particular student offered this group; I get ideas for topics for future science lessons and projects or mini-lessons on time management, breaking up a topic into “chunks,” and so on.

In addition to providing valid information, informal assessment strategies are useful to teachers during parent–teacher conferences for discussing a student’s strengths and weaknesses. They also help build an ongoing record of progress that may be motivating for students to reflect on and useful for their other teachers in planning lessons in different subjects. And finally, the assessments themselves may provide...
meaningful portfolio entries from both a teacher’s and a student’s perspective, serving “as the essential link among curriculum, teaching, and learning” (Wilcox 1997, p. 223).

Many students want to establish a personal rapport with their teachers. They may talk of myriad subjects, seemingly unrelated to the unit. It is often during this informal chatter, however, that you find out about the students’ backgrounds, problems, and interests. This type of conversation, in which you assume the role of active listener, can provide suggestions about topics for future lessons and materials and help the student’s voice emerge.

Discussion, both casual and directed, is also an integral part of assessment. You need to make yourself available, both before and after class, for discussions about general topics, lessons, and assignments. For an assessment of reading comprehension, nothing replaces one-on-one discussion of the material, whether before, during, or after the actual reading. Finally, encourage students to verbalize
their positive and negative feelings about the class itself as well as about topics, readings, and content area activities.

A note of caution: It’s important to realize that “no matter how careful we are, we will be biased in many of our judgments” (MacGinitie 1993, p. 559). Yet teachers who observe with any sort of regularity soon discover that they are able to acquire enough information to process “in a meaningful and useful manner” (Fetterman 1989, p. 88). They can then make reliable decisions about instruction with observation and other techniques in portfolio assessment.

**Portfolio Assessment**

One of the most exciting and energizing developments in assessment is the emergence of portfolios. As a global, alternative, balanced practice in gathering information about students, *portfolio assessment* is a powerful concept that has immediate appeal and potential for accomplishing the following purposes:

- Providing and organizing information about the nature of students’ work and achievements
- Involving students themselves in reflecting on their capabilities and making decisions about their work
- Using the holistic nature of instruction as a base from which to consider attitudes, strategies, and responses
- Assisting in the planning of appropriate instruction to follow
- Showcasing work mutually selected by students and teacher
- Revealing diverse and special needs of students as well as talents
- Displaying multiple student-produced artifacts collected over time
- Integrating assessment into the daily instruction as a natural, vital part of teaching and learning
- Expanding both the quantity and the quality of evidence by means of a variety of indicators

Portfolios are vehicles for ongoing assessment. They are composed of purposeful collections that examine achievement, effort, improvement, and, most important, processes (selecting, comparing, sharing, self-evaluation, and goal setting) according to Tierney, Carter, and Desai (1991). As such, they lend themselves beautifully to instruction in content areas ranging from math and science to English, history, and health education.

Significant pieces that go into student portfolios are collaboratively chosen by teachers and students. Selections represent processes and activities more than products. A distinct value underlying the use of portfolios is a commitment to students’ evaluation of their own understanding and personal development.
Contrasting portfolios with traditional assessment procedures, Walker (1991) submits that instead of a contrived task representing knowledge of a subject, portfolios are an “authentic” assessment that measures the process of the construction of meaning. The students make choices about what to include; these choices in turn encourage self-reflection on their own development, their own evaluation of their learning, and personal goal setting. Advantages of portfolios are more easily visualized when compared with traditional assessment practices as displayed in Table 2.2, adapted from Tierney, Carter, and Desai (1991, p. 44).

### Adapting Portfolios to Content Area Classes

You can, by making some individual adjustments, adapt portfolios to meet your needs. Techniques such as interviewing, observing, and using checklists and inventories provide good sources of information about students in the classroom. The use of portfolios is in many ways a more practical method of organizing this type of information. Linek (1991) suggests that many kinds of data be collected

<table>
<thead>
<tr>
<th><strong>Portfolio</strong></th>
<th><strong>Testing</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Represents the range of learning activities in which students are engaged</td>
<td>Assesses students across a limited range of assignments that may not match what students do</td>
</tr>
<tr>
<td>Engages students in assessing their progress or accomplishments and establishing ongoing learning goals</td>
<td>Mechanically scored or scored by teachers who have little input</td>
</tr>
<tr>
<td>Measures each student’s achievement while allowing for individual differences between students</td>
<td>Assesses all students on the same dimensions</td>
</tr>
<tr>
<td>Represents a collaborative approach to assessment</td>
<td>Assessment process is not collaborative</td>
</tr>
<tr>
<td>Has a goal of student self-assessment</td>
<td>Student assessment is not a goal</td>
</tr>
<tr>
<td>Addresses improvement, effort, and achievement</td>
<td>Addresses achievement only</td>
</tr>
<tr>
<td>Links assessment and teaching to learning</td>
<td>Separates learning, testing, and teaching</td>
</tr>
</tbody>
</table>


---

*Content Area Reading: Literacy and Learning Across the Curriculum*, Eighth Edition, by Richard T. Vacca and JoAnne L. Vacca. Published by Allyn and Bacon. Copyright © 2005 by Pearson Education, Inc.
for a thorough documentation of attitudes, behaviors, achievements, improvement, thinking, and reflective self-evaluation. For example, students may begin a math course with poor attitudes and may constantly challenge the validity of the content by saying things such as, “What are we learning this for anyway? It’s got nothing to do with me and my life.” If you provide opportunities for functional application in realistic situations, comments may change over time to “Boy, I never realized how important this was going to be for getting a job in real life!”

Much more than a folder for housing daily work, a record file, or a grab bag, a portfolio is a comprehensive profile of each student’s progress and growth. Most professional associations have endorsed the use of portfolios. For example, if you are preparing to teach a math class, whether it’s arithmetic, algebra, or trig, consult the National Council for Teachers of Mathematics (NCTM) assessment guidelines. Then, decide with students what types of samples of student-produced work should be included. Best Practice Box 2.1 outlines a procedure for implementing portfolios.

Cherrie Jackman, a fifth-grade teacher, wanted to experiment with portfolios as an assessment tool for writing and science. Here’s how she described the process of implementation that she followed:

To begin implementing portfolios in my class, I followed certain steps:

1. I explained the concept of portfolios and discussed why they are important. We thought of how local businesses use portfolios, and how certain types of professions (architecture, art, journalism) depend on them.

2. I explained the purposes of our portfolio: to describe a portion of students’ work over the quarter, showing how it has improved; to reflect on and evaluate their own work in writing and science; and to compile a body of work that can travel with them from year to year.

3. We discussed the requirements for our portfolio: to select one or two pieces of work from science and writing that each student feels is representative of the best that he or she has done for the quarter; to add one piece for each subject area each quarter; and at the end of the school year, to evaluate students’ overall progress.

4. I gave examples of the kinds of contributions that would be appropriate: writing samples, self-evaluations (reflections) on a particular project, semantic maps, group projects, peer evaluations—all are acceptable pieces of work to place into the portfolio.

5. Finally, we discussed the ongoing process of conferencing that will occur in our classroom. I will meet with students on an individual basis to discuss work in progress and assist in deciding which pieces might be placed in the portfolio. Time in class will be scheduled during the week for students to write reflections, ask for peer evaluations, or hold discussions with teachers about the portfolios. Although the actual work may be done at another time (writing, science), the assessment of the work could be done during this regularly scheduled time.

Portfolios are a process! I really want students to understand that their portfolios are a work in progress. I want them to feel comfortable selecting a piece, critiquing others’ work, and asking questions. I want them to feel ownership for their own work!
An example of a portfolio contribution made by one of Cherrie’s students is a personal reflection on an experiment done in science (see Figure 2.2).

**Checklists and Interviews**

Informal assessment techniques, such as checklists, interviews, and content area reading inventories (discussed later in this chapter), are different from natural, open-ended observation. They often consist of categories or questions that have already been determined; they impose an a priori classification scheme on the
observation process. A checklist is designed to reveal categories of information the teacher has preselected. When constructing a checklist, you should know beforehand which reading and study tasks or attitudes you plan to observe. Individual items on the checklist then serve to guide your observations selectively.

The selectivity that a checklist offers is both its strength and its weakness as an observational tool. Checklists are obviously efficient because they guide your observations and allow you to zero in on certain kinds of behavior. But a checklist can also restrict observation by limiting the breadth of information recorded, excluding potentially valuable raw data. Figure 2.3 presents sample checklist items that may be adapted to specific instructional objectives in various content areas.

In addition to checklists, observations, and inventories, interviews should be considered part of the portfolio assessment repertoire. There are several advantages of using interviews, “be they formal, with a preplanned set of questions, or informal, such as a conversation about a book” (Valencia, McGinley, & Pearson 1990, p. 14). First, students and teachers interact in collaborative settings. Second, an open-ended question format is conducive to the sharing of students’ own

---

**FIGURE 2.2 A Personal Reflection for Science**

**Experiment:**

They’re All Wet.—Determine what effect soaking seeds has on the time it takes them to sprout. In a group of four, develop an experiment using the scientific procedure. Evaluate your group from a scientific and cooperative point of view.

**Reflection:**

I selected the experiment “They’re All Wet” as my best work in science for a number of reasons.

1. My group worked very well together. Everyone was assigned a job (reader, recorder, speaker, organizer), and everyone got to talk.
2. We wrote a sound hypothesis and design for our experiment because we took our time and we thought about the process.
3. We kept very good records of our observations, and then everyone participated in telling about them.
4. Even though our experiment did not prove our hypothesis, I learned many things from this experiment (see above).

Next time maybe my results will support my hypothesis, but I did learn the proper way to conduct an experiment.
## Sample Checklist Items for Observing Reading and Study Behavior

<table>
<thead>
<tr>
<th>Reading and Study Behavior</th>
<th>Fred</th>
<th>Pat</th>
<th>Frank</th>
<th>JoAnne</th>
<th>Jerry</th>
<th>Courtney</th>
<th>Mike</th>
<th>Mary</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Comprehension</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Follows the author’s message</td>
<td>A</td>
<td>B</td>
<td>B</td>
<td>A</td>
<td>D</td>
<td>C</td>
<td>F</td>
<td>C</td>
</tr>
<tr>
<td>2. Evaluates the relevancy of facts</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Questions the accuracy of statements</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Critical of an author’s bias</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Comprehends what the author means</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Follows text organization</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Can solve problems through reading</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Develops purposes for reading</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Makes predictions and takes risks</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Applies information to come up with new ideas</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Vocabulary</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Has a good grasp of technical terms in the subject under study</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Works out the meaning of an unknown word through context or structural analysis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Knows how to use a dictionary effectively</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Sees relationships among key terms</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Becomes interested in the derivation of technical terms</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Study Habits</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Concentrates while reading</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Understands better by reading orally than silently</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Has a well-defined purpose in mind when studying</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Knows how to take notes during lecture and discussion</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Can organize material through outlining</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Skims to find the answer to a specific question</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Reads everything slowly and carefully</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Makes use of book parts</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Understands charts, maps, tables in the text</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Summarizes information</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Grading Key:**  
- A = always (excellent)  
- B = usually (good)  
- C = sometimes (average)  
- D = seldom (poor)  
- E = never (unacceptable)
views. Third, it reveals to what extent students are in touch with their internal disposition toward reading subject matter material.

In general, there are several types of interviews: structured, semistructured, informal, and retrospective. As described by Fetterman (1989, pp. 48–50), these types blend and overlap in actual practice.

1. **Formally structured and semistructured.** Verbal approximations of a questionnaire; allow for comparison of responses put in the context of common group characteristics; useful in securing baseline data about students’ background experiences.

2. **Informal.** More like conversations; useful in discovering what students think and how one student’s perceptions compare with another’s; help identify shared values; useful in establishing and maintaining a healthy rapport.

3. **Retrospective.** Can be structured, semistructured, or informal; used to reconstruct the past, asking students to recall personal historical information; may highlight their values and reveal information about their worldviews.

One technique developed to interview students about the comprehension process is the Reading Comprehension Interview (RCI) (Wixson, Boskey, Yochum, & Alvermann 1984). Designed for grades 3 through 8, it takes about thirty minutes per student to administer in its entirety. The RCI explores students’ perceptions of (1) the purpose of reading in different instructional contexts and content areas, (2) reading task requirements, and (3) strategies the student uses in different contexts.

The RCI’s main uses are to help identify patterns of responses (in the whole group and individuals) that then serve as guides to instruction and to analyzing an individual’s flexibility in different reading activities.

Several questions on the RCI are particularly appropriate for content area reading. Although the RCI was developed for grades 3 through 8, high school teachers can make good diagnostic use of some of the questions.

Rather than interviewing each student individually, we suggest the following adaptation: Have each student keep a learning log. In these logs, students write to themselves about what they are learning. For example, they can choose to focus on problems they are having with a particular reading assignment or activity. A variation on this general purpose would be to ask students to respond to some of the more pertinent questions on the RCI—perhaps one or two at any one time over several weeks.

In relation to a particular content area textbook, examine the kinds of questions students can write about from the RCI:*

---

1. What is the most important reason for reading this kind of material? Why does your teacher want you to read this book?

2. Who’s the best reader you know in (content area)? What does he/she do that makes him/her such a good reader?

3. How good are you at reading this kind of material? How do you know?

4. What do you have to do to get a good grade in (content area) in your class?

5. If the teacher told you to remember the information in this story/chapter, what would be the best way to do this? Have you ever tried (name a strategy, e.g., outlining)?

6. If your teacher told you to find the answers to the questions in this book, what would be the best way to do this? Why? Have you ever tried (name a strategy, e.g., previewing)?

7. What is the hardest part about answering questions like the ones in this book? Does that make you do anything differently?

Having students respond to these questions in writing does not deny the importance of interviewing individuals. However, it does save an enormous amount of time while providing a teacher with a record of students’ perceptions of important reading tasks related to comprehension.

**Rubrics and Self-Assessments**

Students need to play a role in the assessment of their own literacy products and processes. Teachers who want to help students get more involved in assessment invite them to participate in setting goals and to share how they think and feel. What are the students’ perceptions of their achievements? McCullen (1998) described how she begins this process with middle-grade students:

I usually start by envisioning the possible outcomes of each assignment. Then the students and I develop a standard of excellence for each facet of the process and convert the outcomes into a rubric. Thus, before the students begin their research, they know the goals of the assignment and the scope of the evaluation. (p. 7)

Rubrics are categories that range from very simple and direct to comprehensive and detailed. Some are designed to help individual students self-assess; often they are designed to be used by small groups or by an individual student and teacher. In Figure 2.4, a basic rubric serves the dual purpose of involving each student in evaluating the group’s work on an inquiry project involving the Internet and in self-evaluating.

A more detailed rubric, shown in Figure 2.5, was developed in a seventh-grade life science class by the teacher and her students for a unit exploring the five senses. The teacher gave the students copies of the rubric in advance so they could monitor themselves. Using a scale of 0 to 3, students were graded individually and as part of a group by their teacher and by themselves. This rubric may be time consuming to develop.
Rubrics containing less detail and those developed in partnership with students may take less time to construct. They surely help involve students in assessing their own learning in an authentic, meaningful way that keeps the focus on why and how we do what we do.

Assessing Text Difficulty

Evaluating texts and assessing students’ interactions with texts are crucial tasks for content area teachers and students—and they call for sound judgment and decision making. One of the best reasons we know for making decisions about the quality of texts is that the assessment process puts you and students in touch with their textbooks. To judge well, you must approach text assessment in much the same manner as you make decisions about other aspects of content area instruction. Any assessment suffers to the extent that it relies on a single source or perspective on information rather than on multiple sources or perspectives.
## FIGURE 2.5  Detailed Rubric for an Inquiry Project on the Five Senses

<table>
<thead>
<tr>
<th>Group Evaluation</th>
<th>Individual Evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>3</strong></td>
<td><strong>3</strong></td>
</tr>
<tr>
<td>Worked well together every day</td>
<td>Used at least four sources, including one Website and one traditional source; correctly listed the sources</td>
</tr>
<tr>
<td>Thoroughly completed the lab activity</td>
<td>Thoroughly answered the assigned question</td>
</tr>
<tr>
<td>Developed a well-organized, very neatly presented handout that combined all group members’ work, including at least one visual aid</td>
<td>Came up with and answered thoroughly two related questions</td>
</tr>
<tr>
<td>Worked independently on most days</td>
<td>Participated in an experiment and engaged in a thoughtful reflection around that experiment</td>
</tr>
</tbody>
</table>

| **2**            | **2**                 |
| Worked well together most days | Used at least three sources, including one Website and one traditional source; listed the sources |
| Completed the lab activity with some effort | Thoroughly answered the assigned question |
| Developed a well-organized, fairly neatly presented handout that combined all group members’ work; may or may not have included a visual aid | Came up with and tried to answer two related questions |
| Worked independently on some days | Participated in an experiment and engaged in a thoughtful reflection around that experiment |
| | Cooperated with and helped group members most days |

| **1**            | **1**                 |
| May or may not have worked well together | Used at least two sources; listed the sources |
| Completed the lab activity | Answered the assigned question |
| Developed a handout that combined all group members’ work; did not include a visual aid | Came up with and tried to answer one related question |
| Did not work independently | Participated in an experiment and engaged in a reflection around that experiment |
| | Cooperated with and helped group members some days |

| **0**            | **0**                 |
| Did not work well together | Used fewer than two sources |
| Did not complete the lab activity | Did not answer the assigned question |
| Did not develop handout that combined all group members’ work | Did not come up with any related questions |
| Did not work independently | May have participated in an experiment but did not reflect on that experiment |
| | May or may not have cooperated |

### Grading Scale
- 70% of your grade is based on your individual score
- 30% of your grade is based on the group score

<table>
<thead>
<tr>
<th>Final Score</th>
<th>Letter Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.5–3.0</td>
<td>A</td>
</tr>
<tr>
<td>2.0–2.4</td>
<td>B</td>
</tr>
<tr>
<td>1.4–1.9</td>
<td>C</td>
</tr>
<tr>
<td>0.6–1.3</td>
<td>D</td>
</tr>
<tr>
<td>Below 0.6</td>
<td>F</td>
</tr>
</tbody>
</table>
Therefore, it makes sense to consider evidence in the student’s portfolio along with several other perspectives.

One source of information to consider is publisher-provided descriptions of the design, format, and organizational structure of the textbook along with grade-level readability designations. Another information source is your acquired knowledge of and interactions with the students in the class. A third is your own sense of what makes the textbook a useful tool. A fourth source is student perspective, so that instructional decisions are not made from an isolated teacher’s perception of the students’ perspectives. To complement professional judgment, several procedures can provide you with useful information: readability formulas, the Fry graph, the cloze procedure, readability checklists, and a content area framework for student analysis of reading assignments. The first order of business, then, if content area reading strategies are to involve students in taking control of their own learning, is to find out how students are interacting with the text.

**Content Area Reading Inventories**

Teacher-made tests provide another important indicator of how students interact with text materials in content areas. A teacher-made *content area reading inventory* (CARI) is an alternative to the standardized reading test. The CARI is informal. As opposed to the standard of success on a norm-referenced test, which is a comparison of the performance of the tested group with that of the original normative population, success on the CARI test is measured by performance on the task itself. The CARI measures performance on reading materials actually used in a course. The results of the CARI can give a teacher some good insights into how students read course material.

Administering a CARI involves several general steps. First, explain to your students the purpose of the test. Mention that it will be used for evaluation only, to help you plan instruction, and that grades will not be assigned. Second, briefly introduce the selected portion of the text to be read and give students an idea direction to guide silent reading. Third, if you want to find out how the class uses the textbook, consider an open-book evaluation, but if you want to determine students’ abilities to retain information, have them answer test questions without referring to the selection. Finally, discuss the results of the evaluation individually in conferences or collectively with the entire class.

A CARI can be administered piecemeal over several class sessions so that large chunks of instructional time will not be sacrificed. The bane of many content area instructors is spending an inordinate amount of time away from actual teaching.

A CARI elicits the information you need to adjust instruction and meet student needs. It should focus on students’ abilities to comprehend text and to read at appropriate rates of comprehension. Some authorities suggest that teachers also evaluate additional competency areas, such as study skills—skimming, scanning, outlining, taking notes, and so forth. We believe, however, that the best use of reading inventories in content areas is on a much smaller scale. A CARI should seek information related to basic reading tasks. For this reason, we recommend
that outlining, note taking, and other useful study techniques be assessed through observation and analysis of student work samples.

**Levels of Comprehension**

Teachers estimate their students’ abilities to comprehend text material at different levels of comprehension by using inventories similar to the one shown in Figure 2.6 for American history. The teacher wanted to assess how students responded at literal (getting the facts), inferential (making some interpretations), and applied (going beyond the material) levels of comprehension. At this time you can also determine a measure of reading rate in relation to comprehension.

You can construct a comprehension inventory using these steps:

1. **Select an appropriate reading selection from the second fifty pages of the book.** The selection need not include the entire unit or story but should be complete within itself in overall content. In most cases, two or three pages will provide a sufficient sample.

2. **Count the total number of words in the excerpt.**

3. **Read the excerpt, and formulate ten to twelve comprehension questions.** The first part of the test should ask an open-ended question such as, “What was the passage about?” Then develop three or more questions at each level of comprehension.

4. **Prepare a student response sheet.**

5. **Answer the questions.** Include specific page references for discussion purposes after the testing is completed.

While students read the material and take the test, the teacher observes, noting work habits and student behavior, especially of students who appear frustrated by the test. The American history teacher whose inventory is illustrated in Figure 2.6 allowed students to check their own work as the class discussed each question. Other teachers prefer to evaluate individual students’ responses to questions first and then to discuss them with students either individually or during the next class session.

**Rates of Comprehension**

To get an estimate of students’ rates of comprehension, follow these steps:

1. **Have students note the time it takes to read the selection.** This can be accomplished efficiently by recording the time in five-second intervals by using a “stopwatch” that is drawn on the board.

2. **As students complete the reading, they look up at the board to check the stopwatch.** The number within the circle represents the minutes that have elapsed. The numbers along the perimeter of the circle represent the number of seconds.
**General directions:** Read pages 595–600 in your textbook. Then look up at the board and note the time it took you to complete the selection. Record this time in the space provided on the response sheet. Close your book and answer the first question. You may then open your textbook to answer the remaining questions.

**STUDENT RESPONSE FORM**

*Reading time: ________ min. ________ sec.*

I. **Directions:** Close your book and answer the following question: In your own words, what was this section about? Use as much space as you need on the back of this page to complete your answer.

II. **Directions:** Open your book and answer the following questions.

1. To prevent the closing of banks throughout the country, President Roosevelt declared a national “bank holiday.”
   a. True
   b. False

2. The purpose of the Social Security Act was to abolish federal unemployment payments.
   a. True
   b. False

3. The National Recovery Administration employed men between the ages of 18 and 25 to build bridges, dig reservoirs, and develop parks.
   a. True
   b. False

4. President Roosevelt established the Federal Deposit Insurance Corporation to insure savings accounts against bank failures.
   a. True
   b. False

III. **Directions:** Answers to these questions are not directly stated by the author. You must “read between the lines” to answer them.

1. Give an example of how FDR’s first 100 days provided relief, reform, and recovery for the nation.

2. How is the Tennessee Valley Authority an example of President Roosevelt’s attempt to help the poorest segment of American society?

3. How did the purpose of the Civil Works Administration differ from the purpose of the Federal Emergency Relief Act?
3. Later, students or the teacher can figure out the students’ rates of reading in words per minute.

Example:
Words in selection: 1,500
Reading time: 4 minutes 30 seconds
Convert seconds to a decimal fraction. Then divide time into words.

\[
\frac{1,500}{4.5} = 333 \text{ words per minute}
\]

4. Determine the percentage of correct or reasonable answers on the comprehension test. Always evaluate and discuss students’ rates of reading in terms of their comprehension performance.

In summary, information you glean from a CARI will help you organize specific lessons and activities. You can decide on the background preparation needed, the length of reading assignments, and the reading activities when you apply your best judgment to the information you have learned from the assessment.

**Readability**

There are many readability formulas that classroom teachers can use to estimate textbook difficulty. Most popular formulas today are quick and easy to calculate. They typically involve a measure of sentence length and word difficulty to determine a grade-level score for text materials. This score supposedly indicates the reading achievement level that students need to comprehend the material. Because of their ease, readability formulas are used to make judgments about materials. These judgments are global and are not intended to be precise indicators of text difficulty.
A readability formula can best be described as a “rubber ruler” because the scores that it yields are estimates of text difficulty, not absolute levels. These estimates are often determined along a single dimension of an author’s writing style: sentence complexity (as measured by length) and vocabulary difficulty (also measured by length). These two variables are used to predict text difficulty. But even though they have been shown to be persistent correlates of readability, they only indirectly assess sentence complexity and vocabulary difficulty. Are long sentences always more difficult to comprehend than short ones? Are long words necessarily harder to understand than short ones? When a readability formula is used to rewrite materials by breaking long sentences into short ones, the inferential burden of the reader actually increases (Pearson 1974–1975).

Keep in mind that a readability formula doesn’t account for the experience and knowledge that readers bring to content material. Formulas are not designed to tap the variables operating in the reader. Our purpose, interest, motivation, and emotional state as well as the environment that we’re in during reading contribute to our ability to comprehend text.

The danger, according to Nelson (1978), is not in the use of readability formulas: “The danger is in promoting the faulty assumptions that matching the readability score of materials to the reading achievement scores of students will automatically yield comprehension” (p. 622). She makes these suggestions to content area teachers:

1. Learn to use a simple readability formula as an aid in evaluating text.
2. Whenever possible, provide materials containing the essential facts, concepts, and values of the subject at varying levels of readability within the reading range of your students.
3. Don’t assume that matching readability level of material to reading achievement level of students results in automatic comprehension. Remember there are many factors that affect reading difficulty besides those measured by readability formulas.
4. Don’t assume that rewriting text materials according to readability criteria results in automatic reading ease. Leave rewriting of text material to the linguists, researchers, and editors who have time to analyze and validate their manipulations.
5. Recognize that using a readability formula is no substitute for instruction. Assigning is not teaching. Subject area textbooks are not designed for independent reading. To enhance reading comprehension in your subject area, provide instruction which prepares students for the assignment, guides them in their reading, and reinforces new ideas through rereading and discussion. (pp. 624–625)

Within the spirit of these suggestions, let’s examine a popular readability formula and an alternative, the cloze procedure.

**The Fry Graph**

The readability graph developed by Edward Fry (1977) is a quick and simple readability formula. The graph was designed to identify the grade-level score for
materials from grade 1 through college. Two variables are used to predict the difficulty of the reading material: sentence length and word length. Sentence length is determined by the total number of sentences in a sample passage. Word length is determined by the total number of syllables in the passage. Fry recommended that three 100-word samples from the reading be used to calculate readability. The grade-level scores for each of the passages can then be averaged to determine overall readability. According to Fry, the readability graph predicts the difficulty of the material within one grade level. See Figure 2.7 for the graph and expanded directions for the Fry formula.

**Cloze Procedure**

The cloze procedure does not use a formula to estimate the difficulty of reading material. Originated by Wilson Taylor in 1953, a cloze test determines how well students can read a particular text or reading selection as a result of their interaction with the material. Simply defined, then, the cloze procedure is a method by which you systematically delete words from a text passage and then evaluate students’ abilities to accurately supply the words that were deleted. An encounter with a cloze passage should reveal the interplay between the prior knowledge that students bring to the reading task and their language competence. Knowing the extent of this interplay will be helpful in selecting materials and planning instructional procedures. Figure 2.8 presents part of a cloze test passage developed for an art history class. Here is how to construct, administer, score, and interpret a cloze test.

1. **Construction**
   a. Select a reading passage of approximately 275 words from material that students have not yet read but that you plan to assign.
   b. Leave the first sentence intact. Starting with the second sentence, select at random one of the first five words. Delete every fifth word thereafter, until you have a total of fifty words for deletion. Retain the remaining sentence of the last deleted word. Type one more sentence intact. For children below grade 4, deletion of every tenth word is often more appropriate.
   c. Leave an underlined blank of fifteen spaces for each deleted word as you type the passage.

2. **Administration**
   a. Inform students that they are not to use their textbooks or to work together in completing the cloze passage.
   b. Explain the task that students are to perform. Show how the cloze procedure works by providing several examples on the board.
   c. Allow students the time they need to complete the cloze passage.

3. **Scoring**
   a. Count as correct every exact word students apply. Do not count synonyms even though they may appear to be satisfactory. Counting

**Response Journal**

How did you do on the sample cloze passage? Do you think it is a useful assessment tool?
EXPANDED DIRECTIONS FOR WORKING READABILITY GRAPH

1. Randomly select three (3) sample passages and count out exactly 100 words each, beginning with the beginning of a sentence. Do count proper nouns, initializations, and numerals.

2. Count the number of sentences in the 100 words, estimating length of the fraction of the last sentence to the nearest one-tenth.

3. Count the total number of syllables in the 100-word passage. If you don’t have a hand counter available, an easy way is simply to put a mark above every syllable over one in each word; then, when you get to the end of the passage, count the number of marks and add 100. Small calculators can also be used as counters by pushing numeral 1, then pushing the + sign for each word or syllable.

4. Enter graph with average sentence length and average number of syllables; plot dot where the two lines intersect. Area where dot is plotted will give you the approximate grade level.

5. If a great deal of variability is found in syllable count or sentence count, putting more samples into the average is desirable.

6. A word is defined as a group of symbols with a space on either side; thus 1945 is one word.

7. A syllable is defined as a phonetic syllable. Generally, there are as many syllables as vowel sounds. For example, stopped is one syllable and wanted is two syllables. When counting syllables for numerals and initializations, count one syllable for each symbol. For example, 1945 is four syllables.

synonyms will not change the scores appreciably, but it will cause unnecessary hassles and haggling with students. Accepting synonyms also affects the reliability of the performance criteria, because they were established on exact word replacements.

b. Multiply the total number of exact word replacements by two to determine the student’s cloze percentage score.

c. Record the cloze scores on a sheet of paper for each class. For each class, you now have one to three instructional groups that can form the basis for differentiated assignments (see Figure 2.9).

4. Interpretation

   a. A score of 60 percent or higher indicates that the passage can be read competently by students. They may be able to read the material on their own without guidance.

   b. A score of 40 to 60 percent indicates that the passage can be read with some competency by students. The material will challenge students if they are given some form of reading guidance.

---

**FIGURE 2.8 Sample Portion of a Cloze Test**

If the symbol of Rome is the Colosseum, then Paris’s symbol is without doubt the Eiffel Tower. Both are monuments unique (1) planning and construction, both (2) admiration by their extraordinary (3), and bear witness (4) man’s inborn will to (5) something capable of demonstrating (6) measure of his genius. (7) tower was erected on (8) occasion of the World is (9) in 1889. These were the (10) of the Industrial Revolution, (11) progress and of scientific (12). The attempt was made (13) adapt every art to (14) new direction which life (15) taken and to make (16) human activity correspond to (17) new sensibility created by (18) changing times.

Answers to the Cloze Test sample may be found on page 67.

---

**FIGURE 2.9 Headings for a Cloze Performance Chart**

Subject ____________________________________________
Period ______________________________________________
Teacher _____________________________________________

<table>
<thead>
<tr>
<th>Below 40 percent</th>
<th>Between 40 and 60 percent</th>
<th>Above 60 percent</th>
</tr>
</thead>
</table>
c. A score below 40 percent indicates that the passage will probably be too difficult for students. They will need either a great deal of reading guidance to benefit from the material or more suitable material.

The cloze procedure is an alternative to a readability formula because it gives an indication of how students will actually perform with course materials. Unfortunately, the nature of the test itself will probably be foreign to students. They will be staring at a sea of blank spaces in running text, and having to provide words may seem a formidable task. Don’t expect a valid score the first time you administer the test. It’s important to discuss the purpose of the cloze test and to give students ample practice with and exposure to it.

**Readability Checklist**

Despite the many factors to be considered in text evaluation, teachers ultimately want texts that students will understand, be able to use, and want to use. To help guide your assessment and keep it manageable, a checklist that focuses on understandability, usability, and interestability is useful. One such checklist is shown in Figure 2.10; it is an adaptation of the Irwin and Davis (1980) Readability Checklist.

The domain of understandability provides information about how likely a given group of students is to comprehend the text adequately. It helps the teacher assess relationships between the students’ own schemata and conceptual knowledge and the text information. When teachers judge textbooks for possible difficulties, it is imperative to decide whether the author has considered the knowledge students bring to the text. The match between what the reader knows and the text will have a strong influence on the understandability of the material.

Armbruster and Anderson (1981) indicate that one way to judge the author’s assumptions about students’ background knowledge and experiences is to decide if enough relevant ideas are presented in a text to satisfy the author’s purpose. Often, authors use headings to suggest their purposes for text passages. Convert the headings to questions. If the passage content answers the questions, the authors have achieved their purposes and the passage is considerate. If an author hasn’t provided enough information to make a passage meaningful, the passage is inconsiderate.

The second major domain is usability. Is the text coherent, unified, and structured enough to be usable? Divided into two subsections on the Readability Checklist, this section provides information about the presentation and organization of content. It will help the teacher assess factors contributing to the day-to-day use of the text in teaching and the students’ use in learning. These items help pinpoint for a teacher exactly what needs supplementing or what may take additional preparation time or class time.

Essentially, a teacher’s response to these items is another way of deciding if a text is considerate or inconsiderate. A considerate text not only fits the reader’s prior knowledge but also helps “the reader to gather appropriate information with minimal cognitive effort”; an inconsiderate text “requires the reader to put forth extra effort” to compensate for poorly organized material (Armbruster & Anderson 1981, p. 3).
**FIGURE 2.10 General Textbook Readability Checklist**

In the blank before each item, indicate ☑ for “yes,” + for “to some extent,” or ❌ for “no” or “does not apply.”

**Understandability**

1. Are the assumptions about students’ vocabulary knowledge appropriate?
2. Are the assumptions about students’ prior knowledge of this content area appropriate?
3. Are the assumptions about students’ general experiential background appropriate?
4. Does the teacher’s manual provide the teacher with ways to develop and review the students’ conceptual and experiential background?
5. Are new concepts explicitly linked to the students’ prior knowledge or to their experiential background?
6. Does the text introduce abstract concepts by accompanying them with many concrete examples?
7. Does the text introduce new concepts one at a time, with a sufficient number of examples for each one?
8. Are definitions understandable and at a lower level of abstraction than the concept being defined?
9. Does the text avoid irrelevant details?
10. Does the text explicitly state important complex relationships (e.g., causality and conditionality) rather than always expecting the reader to infer them from the context?
11. Does the teacher’s manual provide lists of accessible resources containing alternative readings for very poor or very advanced readers?
12. Is the readability level appropriate (according to a readability formula)?

**Usability**

**External Organizational Aids**

1. Does the table of contents provide a clear overview of the contents of the textbook?
2. Do the chapter headings clearly define the content of the chapter?
3. Do the chapter subheadings clearly break out the important concepts in the chapter?
4. Do the topic headings provide assistance in breaking the chapter into relevant parts?
5. Does the glossary contain all the technical terms in the textbook?
6. Are the graphs and charts clear and supportive of the textual material?
7. Are the illustrations well done and appropriate to the level of the students?
8. Is the print size of the text appropriate to the level of student readers?
9. Are the lines of text an appropriate length for the level of the students who will use the textbook?
10. Is a teacher’s manual available and adequate for guidance to the teachers?
FIGURE 2.10 Continued

External Organizational Aids (continued)

- **11.** Are the important terms in italic or boldface type for easy identification by readers?
- **12.** Are the end-of-chapter questions on literal, interpretive, and applied levels of comprehension?

Internal Organizational Aids

- **1.** Are the concepts spaced appropriately throughout the text, rather than being too many in too short a space or too few words?
- **2.** Is an adequate context provided to allow students to determine the meanings of technical terms?
- **3.** Are the sentence lengths appropriate to the level of students who will be using the text?
- **4.** Is the author’s style (word length, sentence length, sentence complexity, paragraph length, number of examples) appropriate to the level of students who will be using the text?
- **5.** Does the author use a predominant structure or pattern of organization (compare-contrast, cause-effect, time order, problem-solution) within the writing to assist students in interpreting the text?

Interestability

- **1.** Does the teacher’s manual provide introductory activities that will capture students’ interests?
- **2.** Are the chapter titles and subheadings concrete, meaningful, or interesting?
- **3.** Is the writing style of the text appealing to the students?
- **4.** Are the activities motivating? Will they make the student want to pursue the topic further?
- **5.** Does the book clearly show how what is being learned might be used by the learner in the future?
- **6.** Are the cover, format, print size, and pictures appealing to the students?
- **7.** Does the text provide positive and motivating models for both sexes as well as for a variety of racial, ethnic, and socioeconomic groups?
- **8.** Does the text help students generate interest as they relate experiences and develop visual and sensory images?

Summary Rating

Circle one choice for each item.

The text rates highest in understandability / usability / interest.
The text rates lowest in understandability / usability / interest.
My teaching can best supplement understandability / usability / interest.
I would still need assistance with understandability / usability / interest.
The third domain, interestability, is intended to ascertain whether features of the text will appeal to a given group of students. Illustrations and photos may have instant appeal; students can relate to drawings and photographs depicting persons similar to themselves. The more relevant the textbook, the more interesting it may be to students.

Experiment with the Readability Checklist by trying it out on a textbook in your content area. Once you've completed the checklist, sum up your ratings at the end. Does the text rate high in understandability, usability, or interestability? Is a low rating in an area you can supplement well through your instruction, or is it in an area in which you could use more help? Also, summarize the strengths and weaknesses of the textbook. If you noted two areas in which you'd still need assistance, this text is unlikely to meet your needs. Finally, decide how you can take advantage of the textbook’s strengths and compensate for its weaknesses.

**FLIP Strategy**

Efforts to directly access student- or reader-based judgment have resulted in a strategy to provide students with the guidelines they need to assess reading tasks (Schumm & Mangrum 1991). Whereas checklists and formulas are designed for teacher use, a strategy such as FLIP (an acronym for friendliness, language, interest, and prior knowledge) is designed to engage the reader in estimating the level of difficulty of a given source or textbook. With teacher guidance, perhaps using an overhead projector and the think-aloud technique, students actually walk through FLIP and consider these factors:

*Friendliness:* How friendly is my reading assignment? (Students look for text features such as index, graphs, pictures, summaries, and study questions.)

*Language:* How difficult is the language in my reading assignment? (Students estimate the number of new terms.)
Interest: How interesting is my reading assignment? (Students look over the title, headings, pictures, etc.)

Prior knowledge: What do I already know about the material covered in my reading assignment? (Students think about the title, heading, summary, etc.) (pp. 121–122)

Figure 2.11 illustrates how Kristen Hecker uses a FLIP strategy to help her third graders assess the level of difficulty of *Under the Sea*, a book about different life forms living in the ocean, by Claire Llewellyn (1991). First, she asks the students to look at the pictures in the text and share what they think the book will be about with two of their classmates. Then, she helps the whole class examine the tools used by the author in the book: glossary, table of contents, and

![Figure 2.11 A FLIP for Third-Grade Science](image-url)
Looking Back

Assessing students and texts is a process of gathering and using multiple sources of relevant information for instructional purposes. Two major approaches to assessment prevail in education today: a formal, high-stakes one and an informal, authentic one. Pressure from policy makers and other constituencies has resulted in the adoption of curriculum standards specifying goals and objectives in subject areas and grade levels in most states. Hence, student performance on state-mandated tests must also be considered by teachers who need to make instructional decisions based on their students’ content literacy skills, concepts, and performance.

An informal, authentic approach is often more practical in collecting and organizing the many kinds of information that can inform decisions, including (1) students’ prior knowledge in relation to instructional units and text assignments, (2) students’ knowledge and use of reading and other communication strategies to learn from texts, and (3) assessment of materials. The use of portfolios, careful observation and documentation of students’ strengths and weaknesses as they interact with one another and with content-specific material, sheds light on the why as well as the what in teaching and learning.

In this chapter, the key terms, major purposes, and legitimate uses of standardized tests were presented. Contrasts were drawn between portfolios and testing. As teachers engage learners in a process of portfolio assessment, they make adaptations appropriate for their subject matter and consider issues that have been raised about using portfolios. Suggestions for assessing students’ background knowledge included interviews, pretesting, and instructionally based strategies. Interpreting interviews, surveys, scales, and observations, and developing rubrics help in assessing and students’ self-assessing behaviors and views. For insights into how students interact with text material and a measure of performance on the reading materials used in a course, teacher-made content area reading inventories were suggested.

Assessing the difficulty of text material requires both professional judgment and quantitative analysis. Text assessment considers various factors within the reader and the text, the exercise of professional judgment being as useful as calculating a readability formula. Teachers, therefore, must be concerned about the quality of the content, format, organization, and appeal of the material. We supplied three types of procedures for assessing text difficulty: readability formulas, the cloze procedure, and readability checklists.

Despite efforts to assess students and to scaffold instruction in ways that facilitate text learning, some students will continually struggle with literacy. Even though the best
readers and writers may struggle in certain situations, the struggling student often has given up on literacy as a way of learning. Chapter 3 takes a closer look at the literacy needs of the struggling reader and writer. As you read about struggling readers and writers, focus your attention on the role of explicit instruction in the development and use of literacy strategies.

## Minds On

1. You are planning for the new school year in a district whose scores on the statewide proficiency tests have not shown much improvement. A text has been chosen; you have a wide range of auxiliary reading and viewing materials from which to choose. As long as you don't stray from the district’s curriculum standards, you may use any assessment strategies to meet the needs, abilities, and interests of your students. Outline your plan for instructional assessment.

2. For keeping records, most portfolios of student work include a cover page, one that reflects the teacher’s philosophy of assessment. With a group of classmates, select a content area and design a cover page that reflects your vision of authentic assessment.

3. Imagine that you are a new teacher reviewing the required text you will be using in the fall. Initially, you find the book fascinating, and you are certain it will excite many of your students. Yet after analyzing the work, you discover that its readability appears to be above the reading level of most of your students. How might you use this text effectively?

4. Readability formulas are predictive measures. How do predictive measures differ from performance measures in helping you determine how difficult reading materials will be for your students?

## Hands On

1. In groups of three, turn to the “state standards” section of this chapter. How has your state (or region or province) taken action to put standards for curriculum content in place? Describe any recent revisions in process or testing procedures that may affect your local school district. Rewards? Consequences?

2. Develop an observation checklist for the assessment of reading and study behavior in your content area. Compare your checklist with those developed by others in the class for similar content areas. What conclusions might you draw?
3. Each member of your group should locate one sample of text material on the same topic from these sources: an elementary content area text, a secondary content area text, a newspaper, and a popular magazine. Determine the readability of a sample passage from each by using two different readability formulas. Compare your findings by using two additional readability formulas. What conclusions can you draw from the comparison?

4. Two members of your group of four should be designated as observers. The other members should collaboratively attempt to solve the following mathematics problem:

Calculate the surface area of a cylinder that is 12 inches long and 5 inches in diameter.

Note any observations that you believe might be useful in assessing the group’s performance. What types of useful information do observations like these provide?