9

Assessment of Reading, Writing, and Mathematics

Student Learning Outcomes

After completing this chapter, you should be able to do the following:

- Distinguish between formal and informal types of academic assessment instruments.
- Explain the purpose(s) that each assessment can potentially serve in the classroom.
- Identify multiple methods for assessing student performance in reading, mathematics, and writing.
- Develop informal assessment tools that measure student performance in academic areas.
- Interpret data from academic assessments.
- Use assessment data to guide instruction and interventions.

INTRODUCTION: ACADEMIC ASSESSMENT

Teachers assess student performance in all academic content areas starting from the earliest formal educational settings. In 2001, the No Child Left Behind Act (NCLB) mandated the use of high-stakes testing to measure student proficiency and hold schools accountable. The Individuals with Disabilities Education Improvement Act (IDEIA, 2004) aligned special education policy to NCLB and required that the majority of students receiving special education services participate in the same high-stakes assessments as their typically developing peers. The remaining students participate in an alternate assessment. The current age of accountability emphasizes the use of assessments in the classroom. Assessment of academic achievement, however, is not limited to formal test instruments. Teachers can use a variety of tools to measure academic achievement.

There are a multitude of ways to assess student progress that do not cause stress for students who struggle with academic tasks or are plagued by test anxiety. This chapter explains many of the ways to assess reading, mathematics, and writing and describes the purposes for the different assessment methods. The chapter also details the procedures for interpreting data and using the data to guide instruction and interventions. Finally, it discusses general academic achievement assessments.

READING

The National Reading Panel (National Institute of Child Health and Human Development, 2000) was asked to review extant research on teaching reading in K–3 to determine methods that led to improving reading proficiency. The panel examined more than 100,000 studies and determined that reading can be broken into five component parts. Figure 9-1 summarizes the five parts and the primary skills that can be measured within each component. The components build on each other, and proficiency in each is necessary to attain the ultimate goal of reading: to understand what one is reading.

The first component of reading is *phonemic awareness*. This set of skills focuses on an understanding of the smallest meaningful part of the spoken word: the phoneme, or sound that a letter makes. In phonemic awareness instruction, the focus is entirely on the sounds; the student is not learning the letters of the alphabet yet. Phonemic awareness is typically the focus of very early literacy instruction. Students learn the sounds that are included in words, learn to blend and segment phonemes, and learn how to change phonemes to make new words. This skill is the precursor to decoding unfamiliar words.

Students must have a good understanding of phonemes in order to move on to the next component, *phonics*. Phonics is the understanding of the relationship between phonemes and graphemes (the smallest meaningful part of the written word). Students receiving phonics instruction work to forge relationships between the letters of the alphabet and the sounds each letter makes. This component is where the majority of decoding instruction occurs. When students have mastered phonics, they are typically able to decode most unfamiliar words presented to them. Phonics instruction typically occurs through late elementary school; after that, students are expected to be able to decode independently.

Phonemic awareness and phonics can be difficult for some students with linguistic differences and/or various auditory-related impairments. For example, students with otitis

FIGURE 9-1
Measures of Reading Skills

Phonemic awareness: understanding sounds that make up words
- Segmenting phonemes
- Blending phonemes
- Rhyming phonemes

Phonics: understanding the relationship between phonemes and graphemes
- Letter-sound identification
- Decoding skills

Fluency: reading with appropriate rate, accuracy, and prosody
- Words read correctly per minute
- Use of appropriate expression and rhythm during oral reading

Vocabulary: understanding the meaning of words
- Definitions
- Relationship of morphemes to word meanings

Comprehension: understanding the meaning of text
- Literal and inferential questions
- Identification of main idea and supporting details
- Recalling details from text
- Using metacognitive strategies

media (chronic middle ear infections) or auditory-processing issues may find it more difficult to make concrete connections between letters and sounds. When these connections are not solid, decoding becomes difficult, and other areas of reading (fluency and comprehension) are affected. Students with linguistic differences may struggle to learn letter sounds that are not found in their primary language. Thus, they may struggle to remember letter–sound associations, particularly if there is a dearth of opportunities to practice decoding and receive specific feedback on their progress.

The *fluency* component begins a transition period for students. Fluency requires students to make full use of their phonemic awareness and phonics skills as they read passages at an appropriate rate, making few errors and using appropriate prosody. Prosody is the expression good readers infuse into their oral reading. Good readers use an appropriate rate; they do not read too quickly or too slowly. They have developed automaticity in applying rules for decoding words and make few errors as they read. They are able to insert emotion in the right places, as opposed to reading in a monotone voice. The ability to read fluently allows students to be able to read more complex academic textbooks independently. During the late elementary period, students make the transition from "learning to read" to "reading to learn." Reading instruction shifts from a focus on decoding and fluency to a focus on learning strategies to improve vocabulary and comprehension of the text. Students are expected to independently read and gain knowledge from content textbooks, which are often written above the grade level of the reader. If students are still struggling with decoding, they will not be able to read fluently, nor will they be able to focus on learning strategies for vocabulary and comprehension.

The fourth component of reading is *vocabulary*. Students learn about the etymology of words as well as appropriate use of prefixes and suffixes. They learn strategies that teach them how to determine meanings of unknown words. This component is necessary because it facilitates the final component of reading, *comprehension*. Comprehension, as stated previously, is the ultimate goal of reading. Readers must understand the text with which they are presented. In order to do that, they must be able to decode words, read text fluently, and understand the vocabulary in the passage. To handle more complex texts, readers need metacognitive strategies to help them monitor their understanding and allow them to pull main ideas from the text.

All of these components of reading can be assessed in multiple ways. Some instruments measure general reading skills, whereas others are component specific. Teachers can assess students' performance in the individual reading components to determine if there is a problem in a specific area, to form skill-based reading groups, to guide instruction by meeting students at their current performance level, or to monitor student progress. It is important to make sure that the right instrument is selected based on the purpose for conducting the assessment.

Screening Instruments

Screening instruments are typically administered to all students at the beginning of the academic year. The purpose of a screening instrument is to identify students who seem to be struggling with reading in general. Screening instruments typically do not provide specific information about strengths and weaknesses; that is the job of more specific diagnostic instruments. Screening instruments can be individually administered or done electronically in large or small groups. They typically have very short administration times because they are designed to merely identify struggling readers rather than diagnose a specific problem. These instruments often are considered informal assessments; however, if the data from a screening instrument will be used to make instruction decisions for a student, it is important for the instrument to have high reliability and validity. The screening instrument should be administered by people who are trained to do so according to the instrument's instructions. Proper administration of an assessment instrument allows the instrument to measure what

it is intended to measure (i.e. validity) and provide results that accurately reflect the performance of the student. Benchmark assessments are a particular type of screening instrument. These assessments are administered 3–4 times per school year to determine which students are making satisfactory progress, need support, or are struggling. Again specific areas of strength or weakness are not identified, only that there is a problem.

Marcia Invernizzi, Laura Justice, Timothy Landrum, and Keonya Booker (2004/2005) identified four criteria that teachers can use to evaluate potential screening instruments. First, good reading screening instruments should evaluate early literacy skills in phonological awareness, alphabet knowledge, print awareness, concept of word, and grapheme–phoneme correspondence. These skills are all crucial for success in reading. Early detection of difficulties in these areas and subsequent intervention has a positive impact on student performance. If the screening instrument does not measure these four areas, it will be less effective in identifying children who struggle with early literacy (Justice, Invernizzi, & Meier, 2002). Secondly, Invernizzi et al. also contended that good reading screening instruments are sensitive. The purpose of a reading screening instrument is to discriminate between students who are and who are not having problems with reading; thus, it is crucial that the instrument be sensitive enough to student performance to be able to do just that. The third characteristic of a good reading screening instrument is that it must be efficient and easily administered. One of the main characteristics of any screening instrument is that it does not take a great deal of time or effort to administer. Teachers typically administer these assessments to all students. So in order to be effective, a screener must be quick and easy to give. The final characteristic of a good reading screening assessment is that it is technically adequate. In other words, the instrument must have adequate reliability and validity. It must consistently measure what it is intended to measure. The resources section of this chapter contains a link to a website that has compiled a database of reading screening instruments, along with relevant information pertaining to cost, skills measured, technical adequacy, and target grade levels.

Diagnostic Assessment

Whereas screening instruments are intended to identify students who are at risk for problems in reading, diagnostic assessment instruments are intended to identify the specific strengths and weaknesses of the performances of those at-risk students. Teachers can garner a great deal of useful data from the administration of diagnostic reading instruments. The data are used to create homogeneous reading groups, develop individualized education program (IEP) goals and objectives, and drive instruction through the provision of lessons that meet students where they are currently performing. The data from diagnostic reading instruments can be used to assist in the differentiation and modification of the general education curriculum to better meet the individualized needs of students with exceptionalities.

One type of reading diagnostic is an informal reading inventory (IRI), which includes graded word lists and passages used to measure students' decoding, fluency, and comprehension. Graded word lists are lists of words that students would typically learn at specific grade levels. For example, the lower end of graded word lists are more focused on sight words, so the preprimer list might include words such as *see, go, and,* and *can.* More advanced graded word lists include vocabulary such as *decided, silent,* and *certainly* on an upper elementary list and *linear, inept*, and *galore* on a high school list. Graded passages are similar, except the words are in a passage rather than a list (i.e., words taken out of context). As with any type of assessment, it is important to ensure the appropriateness of the instrument to fulfill the purpose set by the teacher.

Rona Flippo, Dari-Ann Holland, Mary McCarthy, and Elizabeth Swinning (2009) developed a set of considerations for selecting an IRI. Teachers must first consider the

content of the IRI. They must determine which skills are assessed by the IRI, whether the IRI is available in other languages to meet the needs of students who do not use English as a primary language, and which grade levels and ranges are included in the IRI. The passages included in the IRI are another area for consideration. Teachers should determine the source of the passages, the type of text (narrative vs. expository text), the length of the passages, and the sensitivity to cultural and linguistic diversity. Comprehension measurement should be considered by examining the reliance on students' prior knowledge for passage comprehension, the complexity of the comprehension questions asked after each passage (the inclusion of literal, inferential, and critical thinking questions), the inclusion of pictures or other visual aids, the sufficiency of questions accompanying each passage, and the inclusion of metacognitive questions. Ease of administration is another area that is important when deciding which IRI to use. Teachers should examine the quantity and type of data collected by the IRI, the existence of parallel forms, the criteria for determining reading level (independent, instructional, and frustration levels), and the explanation of the included subtests. Finally, teachers should consider the ease of data interpretation. Consider the sufficiency of instruction for data interpretation and whether the IRI provides suggestions for linking the data to instruction. Consider whether the data provided by the instrument will be worth the time and effort involved in administering the instrument to the student. Keep in mind that every teacher who sets out to choose an IRI for the classroom has his or her own specific purpose in mind; thus, what one teacher deems sufficient or appropriate for his or her purpose may differ considerably from another teacher's definition of sufficient or appropriate.

Teachers can determine the reading level of a student using IRI passages that are written at specific grade levels. The highest level at which the student is able to read a passage with 95% or better accuracy is typically deemed the *independent level*. Materials written at this level would be appropriate for silent reading activities because the student can decode most of the words fluently and can focus on comprehending what he or she has read. The highest level at which the student reads between 90%–95% accuracy is considered the *instructional level*. Materials written at this level should be used to teach decoding and comprehension strategies to the student. The text is challenging for the student but not overwhelming. Finally, the lowest level at which the student reads with less than 90% accuracy is considered the *frustration level*. Text at this level is very difficult for the student to decode, and he or she could be overwhelmed by the challenge. At this level, the student focuses too much on decoding, leaving minimal energy for comprehending the text.

Some teachers prefer to use more formal reading diagnostic instruments instead of IRIs. The main reason for this preference is the technical adequacy of diagnostic instruments. Many IRIs do not have strong reliability or validity, so teachers prefer to use a standardized instrument. The law requires that an instrument be technically adequate if the resulting data is going to be used to help determine the presence of a learning disability (Torgesen, 2006). These formal assessments, however, should only be administered when they will clearly provide additional information that cannot be gained through other means of assessment, such as examination of work samples, teacher observation, and student interviews. Figure 9-2 describes some of the instruments available to measure various skills in reading.

Reading diagnostic assessments are not appropriate to administer to all students for a number of reasons. First, they are always administered in an individualized setting; thus, they are more time consuming to administer. They are also lengthier than screening instruments; this means that the time spent administering diagnostic assessments can take away from a considerable amount of instructional time. It is more efficient to only administer these tests to students who are significantly struggling with reading and demonstrate that they are not able to make considerable progress even when presented with interventions.

FIGURE 9-2 Reading Assessment Instruments

Woodcock Reading Mastery Test (WRMT), Third Edition

Population: Ages 4 6–79 11 years; Grades K–12

Format: Individual administration

Time: 15–45 minutes

Skills measured: Phonological Awareness, Listening Comprehension, Letter/Word Identification, Rapid Automatic Naming, Oral Reading Fluency, Word Attack, Word Comprehension, Passage Comprehension

Source: **http://www.pearsonassessments.com**

Gray Oral Reading Test (GORT), Fourth edition

Population: Ages 6.0–18.11 years

Format: Individual administration

Skills measured: Rate, Accuracy, Fluency (Rate and Accuracy Combined), Comprehension, and Overall Reading Ability

Source: **http://psychcorp.pearsonassessments.com**

Gates-MacGinitie Reading Tests (GMRT), Fourth edition

Population: Kindergarten through adult

Format: Group administration

Skills measured: Literacy Concepts, Phonological Awareness, Letters, Letter–Sound Correspondence, Listening Comprehension, Word Decoding, Word Knowledge, Reading Comprehension

Source: **http://www.riversidepublishing.com**

Dynamic Indicators of Basic Early Literacy Skills, Sixth Edition (DIBELS™)

Population: Grades K–6

Format: Individual administration

Skills measured: Initial Sound Fluency, Letter Naming Fluency, Phoneme Segmentation Fluency, Nonsense Word Fluency, Oral Reading Fluency, Oral Retelling Fluency, Word Use Fluency

Time: 10–15 minutes per student

Other information: DIBELS™ includes benchmark assessments to be administered to the whole class three times per year and progress monitoring components used weekly for students identified as at risk.

Source: Buros Mental Measurement Yearbook 16, EBSCO database

Progress-Monitoring Assessments

To review the reading assessment process to this point, screening instruments determine which students are at risk for problems in reading, and diagnostic assessments determine these students' specific strengths and weaknesses to help teachers plan effective instruction. Once that instruction is implemented, teachers must use assessment to determine whether or not the instruction is effective. The purpose of progress-monitoring assessments is to help teachers adapt their instruction to meet the individualized levels and needs of the students in the class (Fuchs & Fuchs, 2003).

How does a teacher determine which progress-monitoring tools are effective? Pamela Stecker, Erica Lembke, and Anne Foegen (2008) identified three characteristics of

effective progress-monitoring tools. First, the tool must be sensitive to student change. Because one of the purposes of progress monitoring is to use the data to drive instruction, the data must be able to determine when students are not on target to meet the goal for the specified time period. A second characteristic is that the tool be educationally meaningful. The progress-monitoring tool must measure what it says it will measure and provide data that can be interpreted and used to guide instruction. The final characteristic is cost effectiveness. The financial, labor, and time costs of administering the instrument cannot outweigh the benefits of the information provided by the instrument. The data must provide the necessary information on which instruction decisions can be made or it will not be worth it for the teacher to sacrifice time and effort to administer, score, and interpret the instrument and related data.

One effective progress-monitoring tool is a formative assessment called curriculum-based measurement (CBM™). CBM™ was designed to be used in data-based decision making (Deno, 1989) and examines 1) whether a student is performing at an expected level and 2) whether the student is on target to achieve a set goal by the end of a specific time period (Roehrig, Duggar, Moats, Glover, & Mincey, 2008). CBM™ has been proven to be technically adequate (Baker & Good, 1995; Deno, 1985; Fuchs & Deno, 1994). It uses brief assessments called probes to evaluate student proficiency in a particular academic area. A key characteristic of CBM™ is the standardized administration procedures that allow for data to be graphed over time. The graphed data can be used to differentiate instruction for students. Andrea Capizzi and Lynn Fuchs (2005) found that teachers who used CBM™ data planned instruction that was tied to student needs and had increased student gains. CBM™ is particularly effective in a Response to Intervention (RtI) model because it efficiently indicates whether interventions used in Tiers 2 and 3 are having a positive impact on student performance (Fuchs & Fuchs, 2005; Griffiths, VanDerHeyden, Parson, & Burns, 2006).

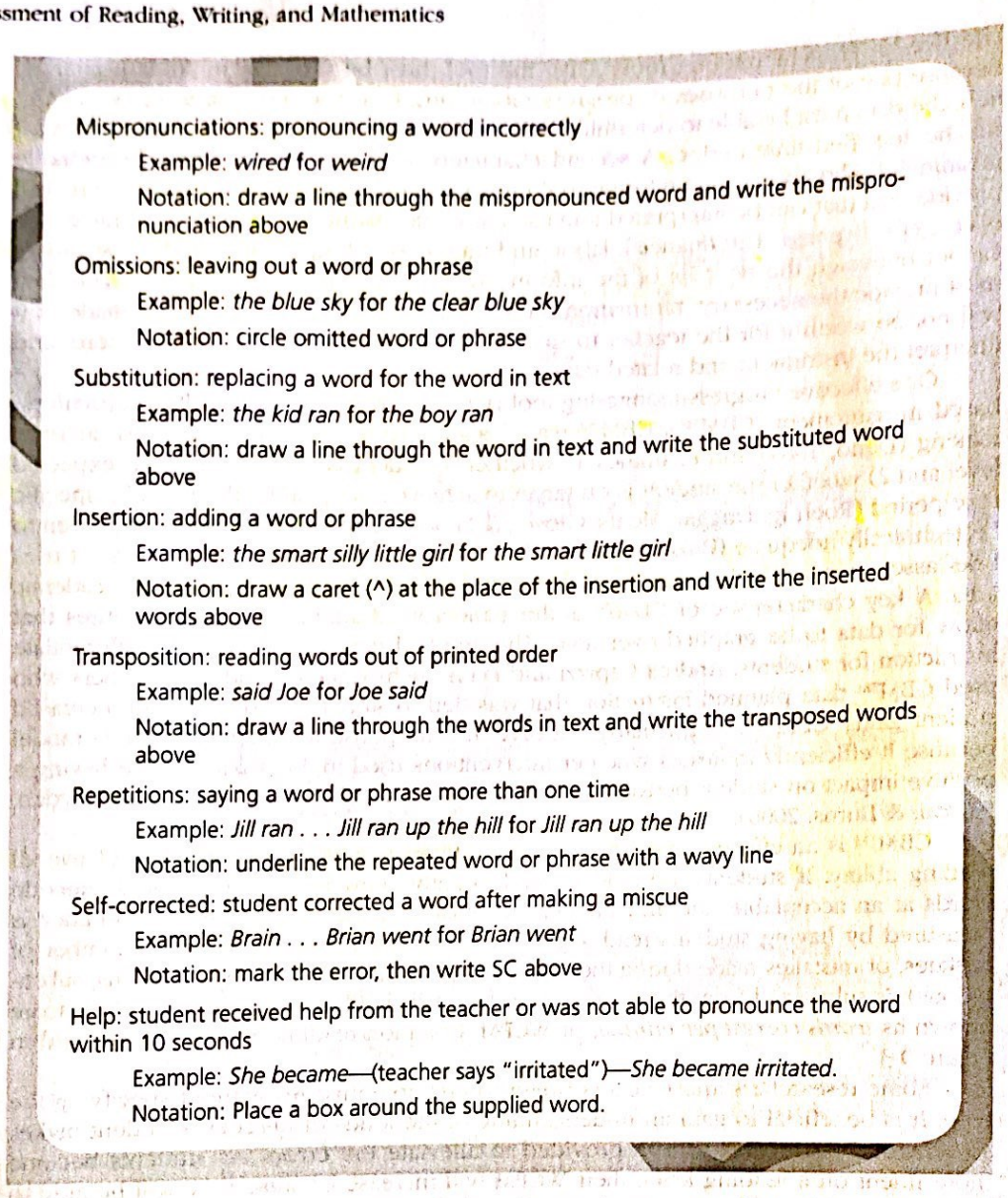
CBM™ is an effective measure of reading fluency, which is an indicator of overall reading ability. If students are able to read fluently, they are typically able to decode words at an acceptable rate and understand what they are reading. Reading fluency is measured by having students read a passage aloud for 1 minute. The total number of miscues, or mistakes made during the reading (e.g., mispronunciations, omissions, substitutions), is subtracted from the total number of words read in 1 minute to obtain a score known as *words correct per minute,* or WCPM. Common reading miscues are defined in Figure 9-3.

Some researchers and teachers prefer to just note miscues without specifying the type. It is beneficial to gain an understanding of the kinds of miscues a student makes so remedial instruction can be provided to alleviate the errors. As students become more fluent on a reading level, their WCPM will increase because they will be able to read more words total and make fewer miscues. It is important to remind students that the passage they are reading is on the level at which they are expected to perform at the end of the specified time period; thus, they do not have to perform at a high level at the beginning. Their fluency will increase as they receive instruction over the course of the time period.

Another type of progress-monitoring tool is the maze assessment. This tool measures silent reading fluency and is a good indicator of overall reading ability (Shin, Deno, & Espin, 2000). The maze assessment takes a passage written on the level of expected performance at the end of a specific amount of time and replaces every *n*th word with a blank and three word choices. Students are expected to choose the word that best fits in the blank (Stecker et al., 2008). An example of a maze tool can be found in Figure 9-4.

Word recognition can also be measured using informal assessment measures. Students are provided with a list of words written on a specific level or that are functional in

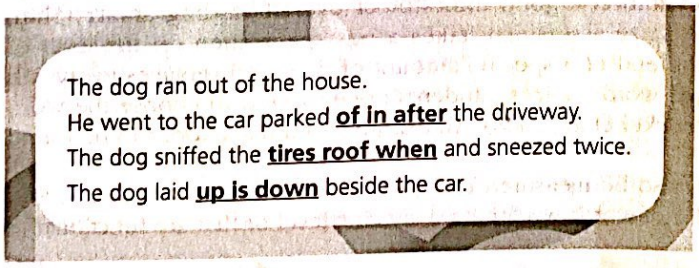
FIGURE 9-3 Common Reading Miscues



nature. Students are expected to read the words on the list without hesitation. The goal of word recognition is to build up the students' sight word vocabularies—that is, the words they can automatically recognize on sight.

In sum, screening instruments are used to determine which students are at risk for problems in reading. Diagnostic instruments are used to identify these students' specific

FIGURE 9-4 Maze Assessment Example



strengths and weaknesses in order to plan effective, individualized instruction. Progress-monitoring instruments are used to evaluate the effectiveness of the instructional methods or intervention for the student. CBM™ and maze assessments are effective tools to monitor student performance on oral reading and silent reading fluencies, respectively.

CHAPTER EXERCISE: ASSESSING READING ABILITY

Marquis, a fourth grader, has been struggling with reading in the past few weeks. His teacher administered a series of diagnostic measures to determine which areas are weaknesses for Marquis. Following is the results from his assessments:

Reading fluency probe

- *Second-grade passage:* read with 96% accuracy; answered 5 out of 5 literal and inferential comprehension questions; miscues were mispronunciations and omissions
- *Third-grade passage:* read with 93% accuracy; answered 3 out of 5 literal and inferential comprehension questions; miscues were mispronunciations and omissions
- *Fourth-grade passage:* read with 75% accuracy; answered 1 literal and 0 inferential comprehension questions out of 5 total; miscues were mispronunciations, omissions, and substitutions

Word recognition probe

- *Second-grade word list:* read with 100% accuracy
- *Third-grade word list:* read with 92% accuracy
- *Fourth-grade word list:* read with 80% accuracy

Phonics subtest on informal reading inventory

- Able to recognize short vowel sounds and all consonants
- Struggled with long vowel sounds when the word ended with an *e*
- Struggled with diphthongs, particularly when they were in positions other than the onset of a word

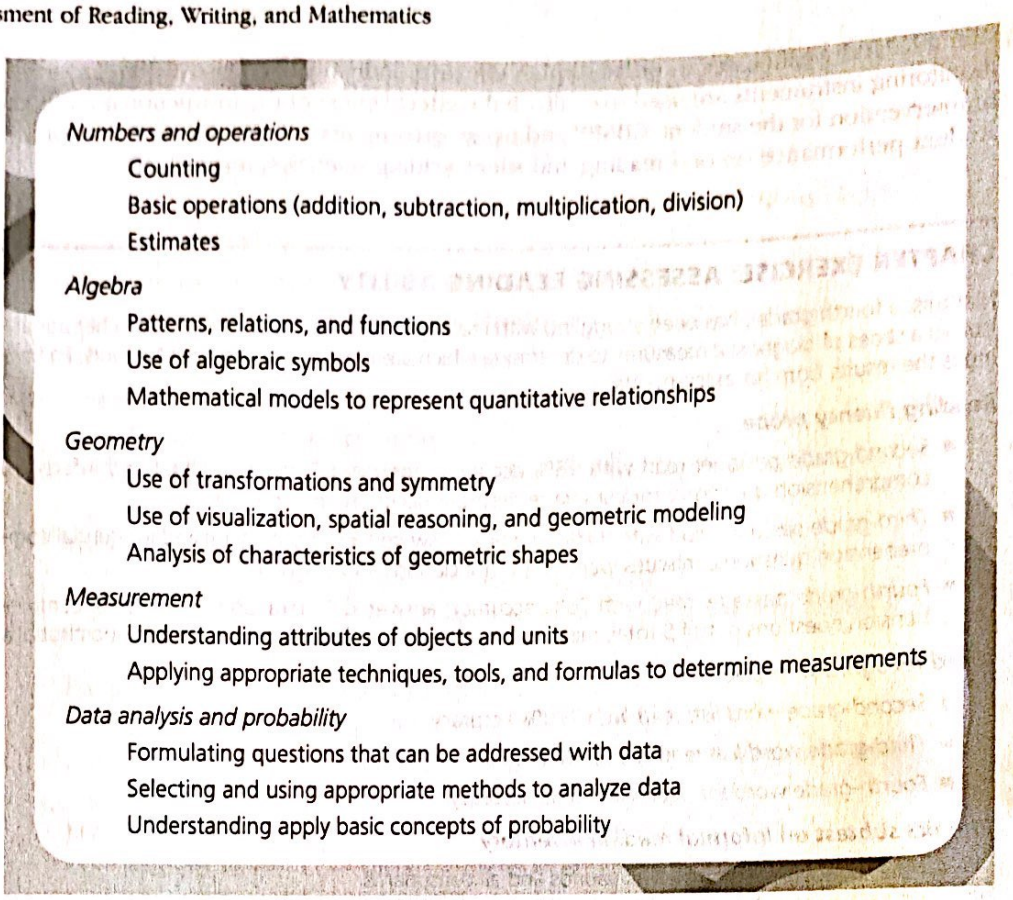
1. Write a brief description of the frustration, instructional, and independent levels of Marquis's performance.
2. Make a brief recommendation about areas that Marquis's teacher should emphasize.

MATHEMATICS

The National Council of Teachers of Mathematics (NCTM) is the professional organization that developed a comprehensive set of mathematics standards that are incorporated in instruction and state academic content standards across the country. These standards include skills that fall into five strands: Numbers and Operations, Algebra, Geometry, Measurement, and Data Analysis and Probability (NCTM, 2000). This organization believes that assessment must be a critical component of instruction and that when assessment is incorporated into instruction, it significantly contributes to all students' learning (NCTM, 2000). As with reading, there are screening instruments, diagnostic assessments, and progress-monitoring tools for mathematics.

Complex mathematical problem solving is based on the knowledge of foundation skills. Number sense is one component of mathematics that must be mastered in order for students to be successful in more advanced mathematics classes. Number sense involves magnitude comparison; students can compare quantities or numbers and quantify the differences. Another foundational component of mathematics is the use of counting strategies. For example, when adding two groups together, the student can take the quantity of the larger

FIGURE 9-5 Areas for Mathematics Assessment



group and "count on" the quantity of the smaller group to get the total. For example, when adding 2 + 3, the student can take the larger quantity (3) and count on two more numbers to get the answer of 5. Students who do not have a good understanding of counting strategies will start at 1, count to 3, and then add the additional two numbers to get to 5. The use of counting strategies allows students to efficiently solve problems without expending a tremendous amount of energy to find the solution. The ability to retrieve mathematics facts is another critical component that is needed for mathematics success. Having a level of automaticity relative to basic mathematics facts allows students to focus more on the algorithm and less on fact recall, making their effort more effectual. Ability to solve word problems is also an indicator of future success in mathematics. In fact, children appear to find word problems easier to solve than number sentences or equations (National Research Panel, 2001). Another important component necessary for mathematics success is numeral recognition, or being able to recognize and name the numerals. Although it is not strictly a mathematical skill per se, it does indicate the potential for students to crack the code of the mathematics language. Figure 9-5 indicates some areas of mathematics that can be assessed.

Screening Assessments

Screening instruments can measure the fluency of early numeracy skills and critical foundational skills that lead to future success in mathematics. Benjamin Clarke and Mark Shinn (2004) used a researcher-designed screening instrument to measure number identification, quantity discrimination, and missing number skills in first-grade students. The skills were measured in the fall and correlated to a spring measure on predictive validity with significant correlations. Those skills were able to predict future success in mathematics.

Screening instruments in mathematics should measure early numeracy skills in the pre-kindergarten and early elementary years. These skills include, but are not limited to, counting, comparison of quantities, and completing number sequences. In the middle elementary years, students should be screened for basic fact fluency using CBM™ techniques. The goal is to have automaticity with basic facts, so the fluency measure should have a time limit. As with reading screening instruments, the purpose of mathematics screening instruments is to determine which students are struggling with mathematics.

Diagnostic Assessment

Leanne Ketterlin-Geller and Paul Yovanoff (2009) identified response analysis as one method of diagnostic assessment used in schools. Response analysis essentially requires the teacher to examine student responses to relevant items sets. Teachers can gather this kind of information in multiple ways. First, they can use a think-aloud process with students. Teachers ask students to talk them through a specific mathematics problem or task. They can have students talk them through the problem prior to solving it or as they are solving it. As teachers listen to the students identifying the steps to be done, they can analyze the students' responses to determine if there are any misunderstandings of the algorithm. If the teacher does identify errors in the response, he or she should take steps to remediate and/or reteach the skill, strategy, or algorithm as soon as possible.

Another way to analyze student response is to conduct an error analysis. In error analysis, the teacher examines student work for patterns in the errors that might indicate various misunderstandings. For example, if a student consistently answers 9 times 5 as 46, this could be an issue where the student learned the fact incorrectly. Figure 9-6 describes the type of errors and provides examples.

1. *Fact error:* Student has not learned the fact correctly.

$$\begin{array}{r} 23 \\ \times\ 4 \\ \hline 93 \end{array}$$ (student answered 4 × 3 as 13 instead of 12)

2. *Operation error:* Student did not check the sign before computing the problem.

$$\begin{array}{r} 15 \\ -\ 7 \\ \hline 22 \end{array}$$ (student added instead of subtracting)

3. *Procedural error:* Student did not do the algorithm correctly.

$$\begin{array}{r} 89^{1}2 \\ -\ 645 \\ \hline 257 \end{array}$$ (student did not follow the rule for regrouping; she did not change the 9 to an *8* after borrowing 10)

4. *Careless error:* There is no consistent pattern; student is not checking work.

$$\begin{array}{r} 47 \\ \times\ 12 \\ \hline 95 \\ +\ 470 \\ \hline 565 \end{array} \qquad \begin{array}{r} 71 \\ \times\ 23 \\ \hline 213 \\ +\ 1420 \\ \hline 1633 \end{array}$$

In the first problem, the student multiplied 7 × 2 and got 15. In the second problem, the student multiplied 7 × 2 and got 14, so we know the student did not learn the fact incorrectly. It is likely a careless error.

FIGURE 9-6 Error Analysis Examples

FIGURE 9-7
Mathematics Assessment Instruments

KeyMath, Third Edition

Population: Ages 4:6–21:11 years; Grades K–12

Format: Individual administration

Time: 30–90 minutes

Skills measured: Basic Concepts (Conceptual Knowledge), Operations (Computational Skills), Applications (Problem-Solving)

Source: http://psychcorp.pearsonassessments.com

Test of Early Mathematics Ability, Third Edition

Population: Ages 3:0–8:11 years

Format: Individual administration

Skills measured: Number Skills, Number-Comparison Facility, Numeral Literacy, Mastery of Number Facts, Calculation, and Concept Understanding

Source: http://www.proedinc.com

Comprehensive Mathematical Abilities Test (CMAT)

Population: Ages 7:0–18:11 years

Format: Individual administration

Time: 30 minutes to 2 hours

Skills measured: Addition; Subtraction; Multiplication; Division; Problem Solving; Charts, Tables, and Graphs; Algebra; Geometry, Rational Numbers; Time; Money; and Measurement

Source: http://www.proedinc.com

The evaluation of students' mathematical language ability should be part of the diagnostic assessment process. Mathematics involves a great deal of content-specific vocabulary; some of those words have both mathematical and nonmathematical definitions. For example, the word "negative" can mean harmful or bad, but it also has a mathematical meaning of less than zero. It is important to determine how much students understand the language of mathematics. Many high-stakes tests expect students not only to solve mathematical problems but also to explain concepts. These explanations indicate whether students have more than a superficial understanding of the concepts. Part of the diagnostic assessment should examine student knowledge of mathematical vocabulary, such as matching terms and definitions. Another part should examine student ability to apply mathematical vocabulary in a written explanation; asking students to write a math journal entry that explains their understanding of a specific concept would be an appropriate assessment task. Figure 9-7 lists some of the assessment instruments that can be used to measure achievement in mathematics.

Progress Monitoring

Progress monitoring in mathematics fulfills a number of purposes and is conducted in a variety of ways. One very important purpose is to fulfill the requirements of the IEP. The progress-monitoring assessment should be tailored to the goals and objectives identified on the IEP; thus, the probes that are used in this type of CBM™ progress monitoring would be skill specific. All the problems on the probe involve the use of the same mathematical skill. Another purpose of progress monitoring is to determine the progress made in learning the skills to be taught in a specific period of time. The probes used for this purpose include a variety of problem types; the problems are a representative sample of the curriculum to be taught during the

12	38
× 14	× 18
48	244
+ 120	+ 380
168	624

The first problem is correct; in the second problem, the student neglected to regroup from multiplying the two digits in the ones column (8 × 8). If the scoring were done using percent correct, the student would only get 50%. Using digits correct, however, there is a possible total of 8 digits in Problem 1 (48, 120, 168), 9 digits in Problem 2 (244, 380, 624) and a cumulative total of 17 possible digits correct. The student in this case got all 8 digits for Problem 1, but only 5 of the 9 digits in Problem 2. For both problems, the student earned a score of 13 out of 17 digits. The digits correct score reflects a more sensitive measure of student performance and would better indicate student growth over time.

FIGURE 9-8 Difference Between Percent Correct and Digits Correct

specified time period. In both types of progress monitoring, the teacher needs to ensure consistency in the procedures used to administer the probes. The directions must be stated to the students the same way, the same amount of time must be allotted for each administration, and scoring must be done the same way for each administration of the probes. When scoring, it is useful to count the number of digits correct in the problem as opposed to calculating the percent of problems correct. Digits correct is a more sensitive indicator of student growth. An illustration of the difference between these two ways of scoring is shown in Figure 9-8.

When using CBM™ to monitor progress, it is important to graph the data from each probe that is administered. The first few probes should be administered prior to the start of instruction to measure what students already know. This is called the *baseline*. The mean or median score of the baseline probes will help you set an appropriate goal for students to reach at the end of the specific period of time (e.g., a grading period). Draw a *goal line* from the mean or median baseline score to the expected performance level to indicate the goal. Each time a probe is administered, the score is plotted on the graph. The *trend line* is the line that connects all the plotted scores; this indicates the trend of student performance. Any time an intervention or instruction is changed in any way, a *phase-change* or *intervention line* (i.e. vertical line) must be drawn on the graph to indicate a change. An example follows in Figure 9-9.

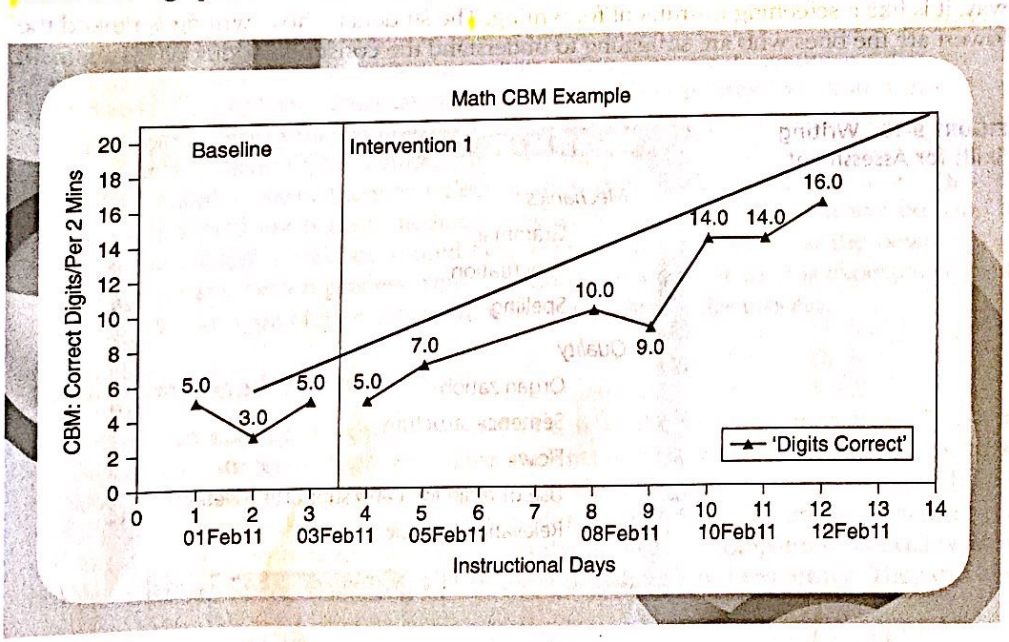


FIGURE 9-9 Math Curriculum-Based Measurement (CBM™) Graph Example

The data from the CBM™ graph can be used to drive instruction. During the intervention, compare the last three consecutive data points on the trend line to the goal line. If the three points are above the goal line, the student has mastered the skill. You have the choice to raise the goal or start teaching the next skill set. If the three points are clustered around the goal line, the intervention or instruction is working. The student is on target to meet the goal, and you should continue without making a change. If the three points are below the goal line, the student is not making satisfactory progress toward the goal. Go back to the last point that was on the goal line, and reteach the material from that point on. An intervention line should be drawn to indicate that there has been a change in instruction; continue to monitor the last three consecutive data points.

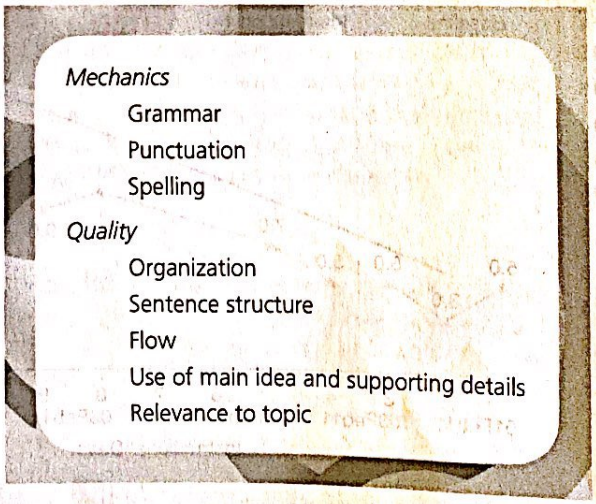
WRITING

Writing is now emphasized in all academic subjects. The reason for this increased infusion of writing is state proficiency assessments. Consequently, states have added short-response, extended-response, and essay questions to their assessments. These questions require students to provide an in-depth explanation of a specific concept to demonstrate their understanding. Therefore, teachers need to instruct students in the content vocabulary for each academic subject area. Without an understanding of this vocabulary, students may find it difficult to formulate their explanations. Students are also assessed on their general writing skills, such as being able to write a business letter on a particular topic using the appropriate format. Figure 9-10 describes some writing skills that can be assessed. Christine Espin, Jacalyn Weissenburger, and Betty Benson (2004) have identified three ways of assessing student writing: holistic, primary trait, and analytic scoring.

Holistic Scoring

Holistic scoring asks the scorer to rank the writing in relationship to other students' work. It relies on the scorer's first impression of the work sample and is the traditional method of scoring student writing. Student work is rank ordered within the larger group. This kind of scoring process does not allow for specific data about student knowledge. In a way, it is like a screening instrument for writing. The students whose writing is ranked the lowest are the ones who are struggling to understand the concepts. Additional assessment

FIGURE 9-10 Writing Skills for Assessment



may be necessary to determine these students' strengths and weaknesses and ways to remediate their weak points.

Primary Trait Scoring

Primary trait scoring is similar to holistic scoring in that the writing sample is rank ordered. The teacher, however, does not compare the sample to other students' work; instead, the sample is compared to a predetermined set of criteria and categorized based on alignment with those criteria. The use of rubrics with writing samples is considered a type of primary trait scoring. The rubric may use points or categories such as *excellent, satisfactory, needs improvement,* and *unsatisfactory*. Each category has descriptions of a particular trait that describes that level of performance. For example, if a category on the rubric was relative to the appropriate use of grammar and spelling, the description under *excellent* might be that the writer made no more than two errors in grammar and spelling throughout the paper. The description under *satisfactory* might be that the writer made no fewer than three and no more than five errors in grammar and spelling on the paper. If the rubric is using points for the categories, writers that earn the top rating in a category might earn 4 points, whereas writers earning a *satisfactory* rating might only receive 3 points. Rubric descriptions and categories vary depending on the writing purpose (Espin et al., 2004).

Analytic Scoring

Analytic scoring is similar to primary trait scoring in that a rubric is used; however, the same rubric is applied regardless of the writing purpose. This kind of scoring typically is used with high-stakes tests and classroom portfolio assessments. It is easier to train scorers on this method of scoring because the same standards are applied consistently.

Portfolio Assessment

Writing is a particularly effective area to apply portfolio assessment techniques. Portfolios have a great deal of flexibility associated with them and can fulfill a variety of assessment purposes. First, determine the purpose for the portfolio because the contents of the portfolio, selection criteria, and evaluation methods may differ based on the established purpose. The purpose may be to celebrate student accomplishment, meet predetermined criteria, follow student progress over time, explore student thought processes as students move through the creation process, or something else. Next, establish what will be included in the portfolio. The contents can range from only "best" work to a mixture of brainstorming sketches through finished copies and everything in between. Establish who will choose the items to be included. If possible, students should have some choice in what is included to allow them to reflect on their own work and provide a rationale for which work they think is best, mediocre, and so forth. Any rubrics that will be used to evaluate the student portfolios should be presented to the students at the beginning of the portfolio construction process. Students should be aware of what is expected of them for this particular type of assessment from the beginning of the process.

Curriculum-Based Measurement

CBM™ can be applied to writing as well as it is to the other content areas. Teachers choose a story starter or topic and have students write for a specific amount of time. Expectations should be made clear so that the students can organize their thoughts before they begin writing and write with a clear beginning, middle, and end. Teachers can use the writing sample to measure such things as the number of complete sentences written, number of words spelled correctly, and number of correct word sequences. They can also

use writing probes to examine student error patterns. This allows teachers to understand the process used by students during the writing task and to see any misunderstandings that are occurring. Teachers can then go back and remediate as necessary through the use of minilessons, writing circles, writer conferences, peer teaching, cooperative learning, or other instructional practices.

CASE STUDY

Kyra is a 10-year-old third grader. She was adopted when she was 18 months old from a Russian orphanage. Although she has a loving family that wants only the best for her, Kyra has been having trouble in school for many years. She has struggled with her academic classes and experienced a great deal of frustration because she does not learn new skills as quickly as her peers. She repeated kindergarten because of the problems she was experiencing, but she still has difficulty mastering content. Each night, she spends at least 3 hours working on assignments that she was not able to finish in school: reading her science, reading, and social studies textbooks; completing her writing assignments; completing her mathematics homework; reading from her chapter book; and finishing her language arts worksheets. Her mother sits at the kitchen table each night to help her do the work, but by the time everything is finally finished, they are both exhausted, frustrated, and close to tears. Kyra's parents have explained these difficulties to Kyra's teachers each year, yet homework time does not get any easier, and no one seems to have a solution. Kyra's mother has finally had enough and sets up a conference with Kyra's current teacher.

Kyra's teacher, Ms. Martinez, is a third-year teacher in the district. She genuinely cares about all of her students and wants them to succeed. She has established a routine of frequent assessment in her classroom and monitors student progress throughout the year. She has helped students set up personal graphs so that they can see their progress in each subject. Student have input in the goals on which they are working. Ms. Martinez closely monitors students she sees struggling so that she can differentiate her instruction to meet their individual needs. She has noticed Kyra struggling this year and has been trying some different interventions to help her, particularly in reading.

When Kyra's mother comes in for the conference, Ms. Martinez is prepared with Kyra's performance graph. She is able to show the graph to Kyra's mother and explain the different interventions that she has tried as well as their effects on Kyra's performance. Kyra's mother is impressed because none of the other teachers have been able to show this kind of information to her; however, she is still concerned because none of the interventions have raised Kyra's performance to meet her personal goal.

Imagine you are Kyra's teacher. You have tried the following Tier 2 interventions with no success: small-group remedial instruction in decoding, addition, and subtraction; additional practice on weekly spelling words; and more frequent progress monitoring. What is your next move to help Kyra? What will you tell her mother?

COMPREHENSIVE INSTRUMENTS

There are instruments available to measure academic achievement in all content areas. Specific instruments have been identified in their respective content areas; however, there are also comprehensive academic achievement instruments. These tests have subtests that measure reading, mathematics, writing/spelling, and other academic content areas. Figure 9-11 identifies some of these instruments. As a special educator, it is important for you to be knowledgeable of these tests and to aware with their general administration requirements. These tests are often used in formal student evaluations and skill progress checks.

Woodcock-Johnson Test of Achievement, Third Edition

Population: Ages 2–90+ years

Format: Individual administration

Time: about 5 minutes per test

Skills measured: Letter/Word Identification, Reading Fluency, Passage Comprehension, Computation, Math Fluency, Math Reasoning, Spelling, Writing Fluency, Writing Samples, and other skills in an extended battery

Source: **http://www.riversidepublishing.com**

Kaufman Test of Educational Achievement, (KTEA), Second Edition

Population: Ages 4:6–25:0 years (Comprehensive Form); Ages 4:6–90+ years (Brief Form)

Format: Individual administration

Time: Comprehensive Form: Pre-K–K = 30 minutes; Grades 1–2 = 50 minutes; Grades 3+ = 80 minutes; Brief Form: Ages 4:6–90 years) 15–45 minutes

Skills measured: Phonological Awareness; Nonsense Word Decoding, Word Recognition Fluency, Reading Comprehension, Math Concepts and Applications, Math Computation, Written Expression, Spelling, Listening Comprehension, Oral Expression, and others

Source: **http://www.pearsonassessments.com**

Weschler Individual Achievement Test (WIAT), Third Edition

Population: Ages 4:0–50:11 years

Format: Individual administration

Time: Varies

Skills measured: Word Reading (phonological awareness and decoding skills); Reading Comprehension; Numerical Operations; Math Reasoning; Spelling; Written Expression; Listening Comprehension; Oral Expression

Source: **http://www.pearsonassessments.com**

FIGURE 9-11 Comprehensive Academic Achievement Instruments

CHAPTER REFLECTION

Assessment plays a critical role in all academic areas. It is important to ensure that students are not only meeting the goals and objectives on their IEPs but also receiving effective instruction to meet their individualized needs. The only way to do this is to apply appropriate assessment techniques. As teachers, we cannot truly know what students have mastered and what they are struggling with unless we gather data using the right tool. Then, we must use that data to facilitate student success. How would you defend your practices if someone says there is too much assessment in schools today? Or that assessment puts too much pressure on students?

CONCLUSION

Screening tools are designed to help teachers determine which students need interventions to help them succeed in academic areas. When work samples, interviews, and observations do not provide enough data, teachers may need to administer diagnostic assessments to gather specific information about the students' strengths and weaknesses in order to plan effective instruction with an appropriate level and skill set. Once an effective instruction is in place, teachers must continue to monitor progress with assessment in order to meet IEP requirements and evaluate the impact of the instruction on student performance. Teachers can use data from formative assessment tools such as CBM™ to guide their instruction and prevent wasted instructional time by introducing interventions that meet the individualized needs of students who struggle with certain academic areas.

PROGRESS CHECK

Review and answer the following questions; these queries are designed to help provide a self-check of the material and concepts that have been covered in the chapter.

1. What types of assessment can be used in reading, mathematics, and writing?
2. Why would a teacher conduct assessments in academic areas?
3. How can assessment data be used to drive instruction?

RESOURCES

The Southwest Educational Development Laboratory (lrs.ed.uiuc.edu) allows teachers to search for reading diagnostic tests. The authors of the website have compiled a database of reading diagnostic assessments that can be searched by various criteria. Each entry in the database includes pertinent information about the instrument. The Southwest Educational Development Laboratory has extensively researched the instruments included in the database. This is a valuable resource for educators.

Just Read, Florida! (**www.justreadflorida.com**) has compiled a chart of reading diagnostic assessments that met specific criteria for reliability and validity. Their website effectively summarizes necessary information about a number of instruments so that teachers can identify one that will meet their specific needs.

Rubistar4Teachers (**rubistar.4teachers.org**) is a website that allows teachers to create their own custom rubrics. There is a database with many rubrics created for various purposes; however, the rubrics can also be customized. This resource is valuable because it is so flexible in meeting the needs of teachers.