DOCUMENT RESUME

ED 443 231 EC 307 925

AUTHOR	Elliott, Stephen N.
TITLE	Educational Assessment and Accountability for All Students: Facilitating the Meaningful Participation of Students with Disabilities in District and Statewide Assessment Programs.
INSTITUTION	Wisconsin State Dept. of Public Instruction, Madison.
ISBN	ISBN-1-57337-079-7
PUB DATE	2000-00-00
NOTE	114p.; Written with Jeffery P. Braden. Foreword by John T. Benson.
AVAILABLE FROM	Publication Sales, Wisconsin Department of Public Instruction, Drawer 179, Milwaukee, WI 53293-0179; Tel: 800-243-8782 (Toll Free); Web site: http://www.dpi.state.wi.us.
PUB TYPE	Guides - Non-Classroom (055)
EDRS PRICE	MF01/PC05 Plus Postage.
DESCRIPTORS	Academic Achievement; *Academic Standards; Accountability; *Disabilities; *Educational Assessment; Elementary Secondary Education; Evaluation Methods; Guidelines; Inclusive Schools; Outcomes of Education; *State Programs; *Student Evaluation; *Student Participation
IDENTIFIERS	Individuals with Disabilities Educ Act Amend 1997; *Testing Accommodations (Disabilities); Wisconsin

ABSTRACT

This guide provides information about the assessment and inclusion of all students in statewide and district assessment programs. In particular, it focuses on tactics for including students with disabilities in assessment to achieve a more complete picture of student learning and educational accountability. It is designed to help Wisconsin educators become familiar with the state's academic content standards and knowledgeable of the general content of tests in the Wisconsin Student Assessment System, so that they can actualize the requirements of the recently reauthorized Individuals with Disabilities Education Act and the potential of standards-based education for all students. In addition, the book provides detailed information on the state's testing guidelines, the valid use of testing accommodations and alternate assessments, and how to communicate these assessment results to educational stakeholders. Specific chapters include: (1) "Educational Assessment Today"; (2) "Characteristics of Good Assessment"; (3) "Understanding and Using the Wisconsin Student Assessment System"; (4) "Facilitating the Participation of All Students in Assessments"; and (5) "Best Practices in Assessment Programs for Educational Accountability." Appendices include standards for teacher competence in educational assessment of students, guidelines for testing procedures, and a code of fair testing practices in education. (Chapters include references.) (CR)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 443 231

EDUCATIONAL ASSESSMENT AND ACCOUNTABILITY FOR ALL STUDENTS

SCOPE OF INTEREST NOTICE

The ERIC Facility has assigned this document for processing to:

In our judgment, this document is also of interest to the Clearinghouses noted to the right. Indexing should reflect their special points of view.

Facilitating the Meaningful Participation of Students with Disabilities in District and Statewide Assessment Programs

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

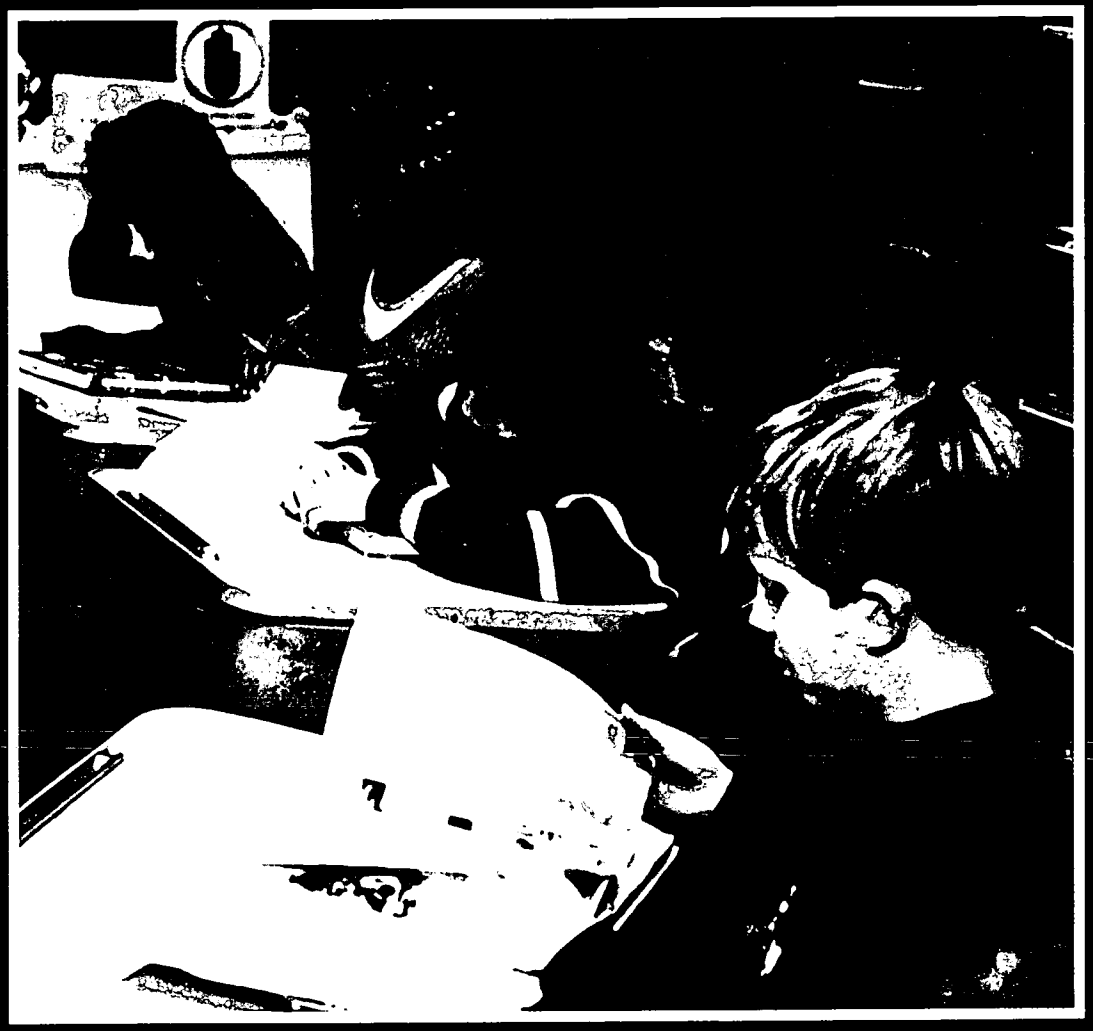
☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Doyle

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)



Wisconsin Department of Public Instruction

BEST COPY AVAILABLE

E 179 25

ERIC

SCOPE OF INTEREST NOTICE

The ERIC Facility has assigned this document for processing to: EC

In our judgment, this document is also of interest to the Clearinghouses noted to the right. Indexing should reflect their special points of view. TM

Educational Assessment and Accountability for All Students

Facilitating the Meaningful Participation of Students with Disabilities in District and Statewide Assessment Programs

Stephen N. Elliott, Ph.D.

With
Jeffery P. Braden, Ph.D.

Department of Educational Psychology
and
Wisconsin Center for Education Research
University of Wisconsin-Madison

Wisconsin Department of Public Instruction
Madison, Wisconsin

This publication is available from:

Publication Sales
Wisconsin Department of Public Instruction
Drawer 179
Milwaukee, WI 53293-0179
(800) 243-8782
www.dpi.state.wi.us

Bulletin No. 00193

© February 2000 Wisconsin Department of Public Instruction

ISBN-1-57337-079-7

Special thanks to CTB/McGraw-Hill for permission to reprint:
Figures 3.2, 3.3, 3.4, 3.6, 3.7, 3.8, 3.9, 3.10, 3.11,
4.5, 4.6, and Box 4.1

The Wisconsin Department of Public Instruction does not discriminate on the basis of sex, race, religion, age, national origin, ancestry, creed, pregnancy, marital or parental status, sexual orientation, or physical, mental, emotional, or learning disability.

Printed on
Recycled Paper

Table of Contents

Preface

This book is about the assessment and inclusion of all students in statewide and district assessment programs. In particular, it focuses on tactics for including students with disabilities in assessments to achieve a more complete picture of student learning and educational accountability.

As stressed throughout the book, assessing all students is an important and sometimes challenging undertaking. It requires knowledge of testing practices, test content, legal guidelines, and technical aspects of tests, as well as a clear understanding of students' learning objectives and instructional programs. If Wisconsin educators are going to actualize the requirements of the recently reauthorized Individuals with Disabilities Education Act (IDEA 1997) and the potential of standards-based education for all students, then all educators will need to be armed with a solid understanding of assessment fundamentals and details about assessment practices. This book is designed to help educators become familiar with our state's academic content standards and knowledgeable of the general content of tests in the Wisconsin Student Assessment System (WSAS). In addition, the book provides detailed information on the state's testing guidelines, the valid use of testing accommodations and alternate assessments, and how to communicate these assessment results to educational stakeholders. As a result of reading *Educational Assessment and Accountability for All Students* and talking with colleagues about assessment activities like those required by the WSAS, you will be prepared to facilitate the meaningful participation of all students in statewide and district assessments.

Stephen N. Elliott, Professor
University of Wisconsin-Madison

Foreword

All children, including children with disabilities, deserve the fullest educational experience our schools in Wisconsin can provide. This includes the right to be involved in the general curriculum and, to the maximum extent possible, meet the same challenging expectations that have been established for all children.

The State of Wisconsin has established a statewide assessment system based on model academic content standards to determine progress of children in meeting those high expectations. Children with disabilities must be included in statewide and districtwide assessments with individual modifications and accommodations as needed, or through alternate assessment if necessary. Their participation assists district staff and parents to judge whether the child' academic performance is improving, just as the participation of nondisabled children does.

The Department of Public Instruction recognizes there may be unique challenges in involving children with disabilities in the assessment system. Dr. Stephen Elliott's work in the area of assessment of children with disabilities has been recognized at the national level. I believe this publication will be helpful to district staff as they work to involve such children in a meaningful, positive manner.

John T. Benson
State Superintendent of Public Instruction

Educational Assessment Today

1

Virtually everybody values high levels of student achievement. Consequently, teachers and other educational professionals are expected to document student achievement and provide periodic summaries of educational progress to students, parents, and fellow educators. The process of documenting and reporting information about student achievement is dependent on good assessments and a method of communicating the results of these assessments so that they are meaningful.

Assessment is NOT a new activity for teachers. Most teachers engage in a wide range of assessment activities daily. For example, let's look into the classroom of Mary Flores, a fourth grade teacher, with an eye toward the various assessment activities she undertakes during the course of a typical day.

Mary arrived at school, as usual, 30 minutes before the first bus. She readied her room for the day's activities by writing the work schedule on an overhead transparency and briefly organizing her lesson notes, and then went to meet her students as they came streaming into the building at 8:15. During the course of the day, she

- *recommended John spend extra time each night this week reviewing his multiplication facts;*
- *called on Sylvia twice even though she had not volunteered to answer questions about the social studies unit;*
- *scored and assigned grades to her students' spelling tests;*
- *referred Jared to the school psychologist for evaluation because of his persistent learning difficulties in math and science;*
- *stopped her planned English lesson halfway through the period to review the previous day's lesson because several students seemed confused;*
- *assigned homework in math and social studies, but not English;*
- *reviewed learning objectives for the forthcoming statewide assessment in mathematics and then made some minor adjustments in her lesson plans to include two days to do some sample test items;*
- *held a lunch-time conference with the parents of a student with a disability to discuss the possible use of testing accommodations to facilitate his inclusion in the forthcoming state and district assessments;*
- *gave a quiz in science covering two chapters and a field trip experience;*
- *listened to oral book reports from half of her students and then provided them with feedback about each presentation;*
- *made notes to herself about some key words and important concepts in science with which students were struggling during a class discussion on rocks; and*
- *wrote three short essay questions and outlined model answers to each in preparation for the next week's end-of-unit test in social studies.*

As illustrated by the vignette of Mary Flores, **educational assessment is an information gathering and synthesizing process for the purpose of making decisions about students' learning and instructional needs.** Common assessment methods for most teachers include self-constructed tests or quizzes, interviews or oral questioning, classroom observations, behavior rating scales, classroom projects, and commercially published tests.

Today, with the advent of standards-based educational reforms and changes in laws concerning the assessment of all students, many educators involved in the assessment of student achievement need more advanced knowledge of assessment tools and practices. In particular, educators need more knowledge about the use and interpretation of standardized group achievement tests as they

BEST COPY AVAILABLE

apply to all students because of the increased consequences associated with such tests in statewide assessment programs. Thus, this book has been written to advance teachers' understanding of assessment, in particular large-scale assessments, and of ways to facilitate the inclusion of ALL students in assessments that are being used as the primary method for increasing educational accountability to the public.

Why Assess Students?

Teachers and parents obviously want students to learn and excel in school. Consequently, assessments are necessary to determine if students are learning and developing competencies required for success later in life. Educators have observed that most students work harder and are more attentive when they think they are going to be held accountable for what they are studying. In other words, when students know they will be assessed on the subject matter they are being taught, they tend to study harder and learn more. Thus, for some students, knowing they are going to be assessed has important attentional and motivational consequences. It is widely recognized that for some students, tests can be a source of anxiety as well as exciting opportunities to demonstrate what one knows. Tests and assessments also can be sources of anxiety for educators. So why give tests and create statewide assessment systems? Tests play a major role in the lives of most students and teachers, and we use them to

- measure student achievement;
- evaluate students' acquisition and degree of mastery of important skills;
- provide information to guide instructional practices;
- evaluate the effectiveness of instructional practices; and
- monitor educational systems for public accountability.

To adequately achieve each of the various purposes listed above, educators must use different types of tests and related assessment practices. Before getting too far into our examination of the various assessment practices educators use and the information resulting from these practices, it is important to have a good understanding of key assessment terms and fundamental assessment principles that should guide wise use of tests and assessment results.

Let's Communicate: Some Key Terms

Effective communication about tests and educational assessment in an era of standards-based reform requires us to carefully define assessment, testing, and measurement, as well as terms associated with standards, including content standards, performance standards, and proficiency standards.

By **assessment** we mean the process of gathering information about a student's abilities or behavior for the purpose of making decisions about the student. A teacher can use many tools or methods to assess a student, such as paper-and-pencil tests, rating scales or checklists, interviews, observations, and published tests. Thus, assessment is more than testing.

Testing is simply one procedure through which we obtain evidence about a student's learning or behavior. Teacher-constructed tests, as well as commercially published tests, have and will continue to play a major role in the education of students. Such tests are assumed to provide reliable and valid means to measure students' progress. A test is a sample of behavior. It tells us something, not everything, about some class or type of behavior. Well-designed tests provide representative samples of knowledge or behavior.

To **measure** means to quantify or to place a number on a student's performance. Not all performances demonstrating learning can or need to be quantified (for example, art or musical exhibitions). The science of measurement in itself includes many important concepts—validity, reliability, standard scores—for teachers and others responsible for assessing students.

Educational assessment today is occurring within a context of educational change commonly referred to as standards-based reform. Wisconsin is one among many states that has embarked upon standards-based reform. Three types of standards are central to Wisconsin's reform efforts. First, **content or academic standards**. These are general statements that describe *what students should understand and be able to do* in various content areas such as English language arts, mathematics, science, and social studies. Subsumed within each content standard are **performance standards**, which are defined as specific statements of expected knowledge and skills necessary to meet a content standard requirement at a particular grade level. Thus, performance standards indicate *how students can*

show what they understand and can do (see Figure 1.1). Finally, **proficiency standards** are descriptive categories that describe the degree to which performance standards have been attained. In Wisconsin, there are four levels of proficiency used to describe how well a student has done on a test that is designed to measure most of the state's content standards (see Figure 1.2).

We will say much more about educational assessment within a standards framework throughout this book. Thus, it is important you have a good understanding of the previous six key terms before reading further.

Guiding Principles

Large-scale assessment is a puzzling activity to many teachers. Historically, such assessments have not been aligned with standards, used different tests about every three years, and were not associated with any significant consequences. Times are changing and periodic large-scale assessments are becoming an important part of educational accountability. Therefore, it may be useful to keep the following fundamental assessment principles in mind when you are discussing or using achievement tests to evaluate your students.

Principle No. 1: Standards First, Then Testing. When states and school districts set out to reform their education systems, it is important that they follow a logical sequence of events. First, they should set goals for each educational system. Second, they must adopt content standards that specify what children should know and be able to achieve. Third, they must adopt curricula and select instructional materials which enable teachers to help their students meet the standards. Finally, they should develop assessments to measure students' progress toward meeting the standards. In other words, "assessments should follow, not lead, the movement to reform our schools. Only then can we build and use new tests that accurately measure students' progress toward meeting standards." (Kean, 1998, p. 2)

Many of the desired skills and much of the information educators value today are part of the *content and performance standards* the state has developed in the areas of reading, mathematics, language arts, writing, science, and social studies. The use of results from tests that validly assess what all students know and can do in these content areas is a major component of a common *accountability system* for students receiving instruction in either a regular education or special education classroom. Information about all students' educational performance lies at the core of any educational accountability system. Only with public reporting on these performances can policy makers and educators make informed decisions to improve education for all students. At this point in time, results of students' performances on achievement tests have become the most frequently used indicator for accountability purposes. Thus, involving all students in assessment systems like the Wisconsin Student Assessment System (WSAS) is an important aspect of an inclusive education and it is essential to educational accountability.

Principle No. 2: Tests Measure Educational Progress; They Don't Create It. The central purpose of any test is to provide accurate and reliable information, not to drive educational reform. Some people have suggested that tests alone can create higher levels of educational achievement, but it is important to realize that new assessment systems cannot cure ailing education systems. Tests do not create better students—good teachers and good schools do.

Meaningful information resulting from tests, however, can help teachers do their jobs better. From a teacher's perspective, the primary purpose of assessment is to gather information about students' performances to make decisions about how and where the students should be instructed. Therefore, to the degree that teachers are knowledgeable about assessment, they increase the likelihood of making good decisions about the students in their classrooms. In essence, effective teaching boils down to good instruction, good assessment, and using each to do the other better (Witt, Elliott, Daly, Gresham, and Kramer, 1998).

Principle No. 3: No Single Test Does Everything; Use Multiple Measures and Repeated Measurements. Most educators realize that no single test can serve all the possible purposes for testing. A variety of tests or multiple measures are necessary to provide educators with a comprehensive view of what students know and can do. This should not be surprising given the array of learning expectations we have for students—we want them to be able to read, write, communicate orally, use technology, do research, calculate, conduct experiments, and understand and solve social problems. Some of these skills or competencies could be meaningfully assessed

Figure 1.1

Sample Mathematics Content and Performance Standards from Wisconsin's Academic Standards

MATHEMATICS

D. MEASUREMENT

CONTENT STANDARD

Students in Wisconsin will select and use appropriate tools (including technology) and techniques to measure things to a specified degree of accuracy. They will use measurements in problem-solving situations.

Rationale: Measurement is the foundation upon which much technological, scientific, economic, and social inquiry rests. Before things can be analyzed and subjected to scientific investigation or mathematical modeling*, they must first be quantified by appropriate measurement principles. Measurable attributes* include such diverse concepts as voting preferences, consumer price indices, speed and acceleration, length, monetary value, duration of an Olympic race, or probability of contracting a fatal disease.

PERFORMANCE STANDARDS

BY THE END OF GRADE 4 STUDENTS WILL:

D.4.1 Recognize and describe measurable attributes*, such as length, liquid capacity, time, weight (mass), temperature, volume, monetary value, and angle size, and identify the appropriate units to measure them

D.4.2 Demonstrate understanding of basic facts, principles, and techniques of measurement, including

- appropriate use of arbitrary* and standard units (metric and US Customary)
- appropriate use and conversion of units within a system (such as yards, feet, and inches; kilograms and grams; gallons, quarts, pints, and cups)
- judging the reasonableness of an obtained measurement as it relates to prior experience and familiar benchmarks

D.4.3 Read and interpret measuring instruments (e.g., rulers, clocks, thermometers)

D.4.4 Determine measurements directly* by using standard tools to these suggested degrees of accuracy

- length to the nearest half-inch or nearest centimeter
- weight (mass) to the nearest ounce or nearest 5 grams
- temperature to the nearest 5°
- time to the nearest minute
- monetary value to dollars and cents
- liquid capacity to the nearest fluid ounce

D.4.5 Determine measurements by using basic relationships (such as perimeter and area) and approximate measurements by using estimation techniques

with a group-administered, paper-and-pencil test requiring brief answers, while others would require more individualized assessments with direct observations by a teacher and the production of a product or detailed report. In addition, it is also a sound practice to assess important skills or competencies at least twice to gain confidence in the assessment results.

Principle No. 4: Valid and Reliable Test Scores Are Important. For assessment results to be useful, the subject matter examined should be similar to what has been emphasized during instruction and students' responses must be measured and scored accurately. In the words of testing experts, an assessment must be valid and reliable. Tests that are used to make important educational decisions must meet rigorous technical standards for producing accurate and valid information.

The concepts of test score validity and reliability are quite abstract for most people and seemingly important only to the experts who construct tests. And yet almost every student with whom we have ever worked will express concerns about a test that doesn't appear to measure what he or she has been taught, or which results in inconsistent scores for two or more students who have produced the same response. Students care about the quality of tests and the meaning of the resulting scores even if they don't understand the technical concepts of reliability and validity. Most educators and parents also care about the quality of tests, especially if important educational decisions such as promotion or graduation are based on them. Consequently, we will say quite a bit about the concepts of reliability and validity (in Chapter 2), especially in the context of inclusive assessment practices.

High Standards for All Students

You probably have read or heard colleagues speak about "High Standards for All Students," and have no doubt wondered, "Is this possible?" Few educational movements have been so clearly identified by a single rallying cry as the standards-based reforms now dominating the nation's education policy agenda (McDonnell, McLaughlin, and Morison, 1997). Central to the standards-based reform efforts is the belief that we can improve overall educational quality by setting clear and high academic standards and expecting schools to teach and students to learn according to those standards.

Four common elements seem to characterize this reform across the country. First, there is a focus on student achievement as the primary measure of school success. Second, this reform emphasizes setting challenging academic standards that specify the knowledge and skills students should acquire and the levels at which

Figure 1.2

General Proficiency Levels Used to Describe Student's Performance on the Statewide Knowledge and Concepts Examinations

Advanced	Proficient	Basic	Minimal Performance
Distinguished in the content area. Academic achievement is beyond mastery. Test score provides evidence of in-depth understanding in the academic content area tested.	Competent in the content area. Academic achievement includes mastery of the important knowledge and skills. Test score shows evidence of skills necessary for progress in the academic content area tested.	Somewhat competent in content area. Academic achievement includes mastery of most important knowledge and skills. Test score shows evidence of at least one major flaw in understanding the academic area tested.	Limited in the content area. Test score shows evidence of major misconceptions or gaps in knowledge and skills basic to progress in the academic content area tested.

BEST COPY AVAILABLE

they should demonstrate mastery of that knowledge. Third, there is a desire to extend the standards to all students, including those for whom the learning expectations have been traditionally low. Fourth, and one of the main concerns of this book, reform efforts rely heavily on achievement testing to spur change and to monitor its impact. Consequently, personnel in the Wisconsin Department of Public Instruction, just like those in most states across the country, have been developing frameworks for educational standards, state assessments, and accountability systems. Concurrent with the standards-based education reform efforts, there have been changes in federal law concerning students with disabilities and their involvement in all statewide and districtwide assessment programs. Thus, the goals of most standards-based reforms are to: (a) specify in the form of academic and performance standards the knowledge and skills all students will be expected to demonstrate at selected times during their education; (b) encourage educators to align their curriculum and instruction so as to facilitate students' opportunities to acquire the knowledge and skills competencies; (c) develop or purchase valid tests or other methods for assessing the extent to which all students achieve these knowledge and skills competencies; and (d) communicate annually with the public, using proficiency standards to report how well students are performing with respect to identified knowledge and skills competencies. These are challenging goals, but not unrealistic.

Perhaps one of the most significant challenges for all of us in education is to establish high academic standards and document the results of all students' education against these standards across statewide or districtwide assessment systems. A particularly vexing part of this challenge is the meaningful participation of students with disabilities in one accountability system with all other students. Given that a significant number of students with disabilities and limited English proficiency historically have been excluded or exempted from large-scale assessments, substantial efforts will be needed to achieve an accountability system that truly includes all students. For example, participation rates for students during the past several years in statewide assessments—such as the Wisconsin Reading Comprehension Test at third grade and the WSAS Knowledge and Concepts Test at grades 4, 8, and 10—have ranged from a low of 41 percent to a high of 95 percent. Many of the students who did not participate were students with disabilities or limited English proficiency.

There are several possible reasons for the lower-than-desired participation rates of students with disabilities in our statewide assessments. These include

- a perception that the tests are not relevant;
- a desire to "protect" these students from another frustrating testing experience;
- a concern that these students will lower the school's mean score in each content area;
- the fact that some parents do not want their son or daughter spending time taking a test that they don't understand or value; and
- the belief that guidelines for administering a standardized achievement test prohibit, or at least limit, what can be changed without jeopardizing the validity of the resulting test score. Many educators have been admonished, "Don't mess with the test," and so are confused about what can and cannot be changed within a test.

If educators and other educational stakeholders who aspire to "high standards for all students" are to have a meaningful picture of how well students are learning and applying valued content knowledge and skills, all students need to be assessed periodically. The absence of students with disabilities from our state and district assessments will result in unrepresentative mean scores and norm distributions, reinforced belief that students with disabilities cannot do challenging work, and undermined inclusion efforts for many students who can benefit from the same instruction as their nondisabled peers.

Testing students, making decisions about including students with disabilities in assessment programs, and implementing assessments so they are valid can be challenging activities requiring teachers' active involvement. As we noted earlier, large-scale assessments (like the Knowledge and Concepts Examination) that are used for system-wide accountability may be a bit puzzling for some teachers. This is an understandable state of mind due to a number of "pieces to the accountability puzzle" (see Figure 1.3) and some new legal requirements concerning students with disabilities. We already have introduced many of the pieces in the accountability puzzle, and in fact have written this entire book around the key topics highlighted by this puzzle metaphor. At this time it is enough

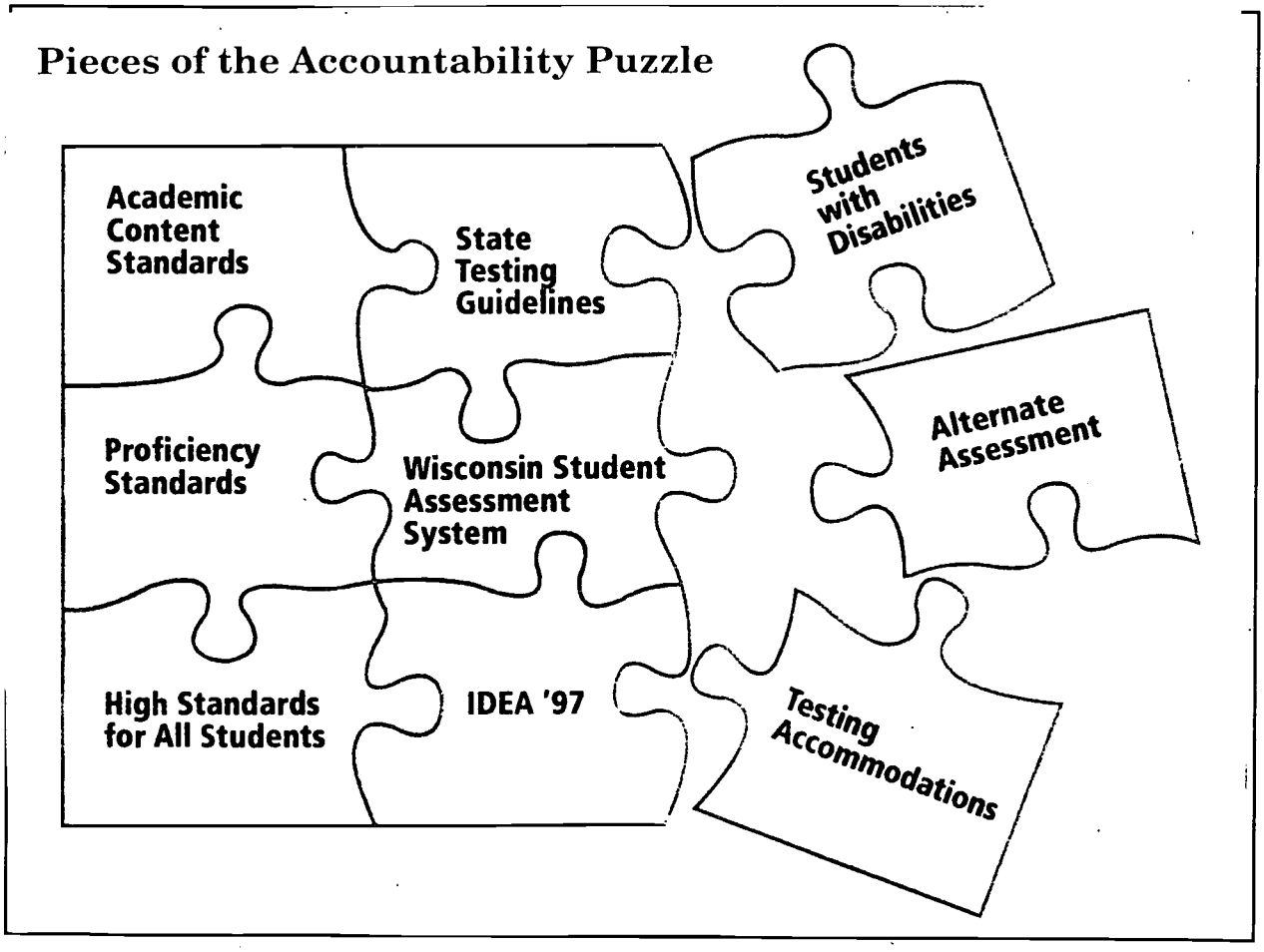
to simply familiarize yourself with the twelve topics identified in the puzzle. Over the course of reading this book, however, you will learn more about how these pieces fit together and result in a big assessment picture.

Teachers Have Standards, Too: Professional Roles and Responsibilities for High Quality Student Assessments

Though we have only begun an examination of assessment of student achievement, it should be very evident that teachers must be knowledgeable assessment agents, capable of using a variety of techniques to describe students' learning and to communicate with students, parents, and others about such learning. Accordingly, the American Federation of Teachers (1990, p. 1) believes that "assessment competencies are an essential part of teaching and that good teaching cannot exist without good student assessment." As a result of these beliefs, educators representing the American Federation of Teachers, the National Council on Measurement in Education, and the National Education Association wrote a set of seven standards for teacher competence in student assessment. A brief listing of these standards follows (see Appendix A for a complete copy of *Standards for Teacher Competence in Educational Assessment of Students*):

- Teachers should be skilled in *choosing* assessment methods appropriate for instructional decisions.
- Teachers should be skilled in *developing* assessment methods appropriate for instructional decisions.
- Teachers should be skilled in *administering, scoring, and interpreting the results* of both externally-produced and teacher-produced assessment methods.

Figure 1.3



BEST COPY AVAILABLE

- Teachers should be skilled in *using assessment results* when making decisions about individual students, planning teaching, developing curriculum, and improving schools.
- Teachers should be skilled in *developing valid pupil grading procedures* that use pupil assessments.
- Teachers should be skilled in *communicating assessment results* to students, parents, other lay audiences, and other educators.
- Teachers should be skilled in *recognizing unethical, illegal, and otherwise inappropriate assessment methods and uses of assessment in formation.*

The enactment of these standards for competencies in educational assessment requires a range of activities by teachers prior to, during, and after instruction. For example, assessment activities prior to instruction involve teachers (a) clarifying and articulating the performance outcomes expected of students, (b) understanding students' motivations and creating connections between what is taught and tested and the students' world outside of school, and (c) planning instruction for individuals and groups of students that is aligned with what will be tested.

Assessment-related activities occurring during instruction involve (a) monitoring student progress toward instructional goals, (b) identifying gains and difficulties students experience in learning and performing, (c) adjusting instruction to better meet students' learning needs, (d) giving contingent, specific praise and feedback, and (e) judging the extent to which students have attained instructional outcomes.

Finally, the assessment-related activities occurring after instruction include (a) communicating strengths and weaknesses based on assessment results to students and parents, (b) recording and reporting assessment results for school-level analysis, evaluation, and decision making, (c) analyzing assessment information before and during instruction to understand each student's progress and to inform future instructional planning, and (d) evaluating the effectiveness of instruction and related curriculum materials.

To close this section on teachers' roles in assessment, you might find it interesting to examine a review study by Robert Hoge and Theodore Coladarci (1989) concerning research on the match between teacher-based assessments of student achievement levels and objective measures of student learning. As a rationale for their work, they noted that while many decisions about students are influenced by a teacher's judgments of the student's academic functioning, historically there seems to be a widespread assumption that teachers generally are poor judges of their students' academic abilities.

Hoge and Coladarci identified 16 studies that were methodologically sound and featured a comparison between teachers' judgments of their students' academic performance and the students' actual performance on individualized achievement tests. They found generally high levels of agreement between teachers' judgmental measures and the standardized achievement test scores. The correlations ranged from a low of .28 to a high of .92, with the median being .66. (Note: A perfect correlation would be 1.00.) The median correlation certainly exceeds the validity coefficients typically reported for psychological tests.

In a recent replication of this research on the accuracy of teacher judgments, Demaray and Elliott (1998) found that teachers accurately predicted 79 percent of the items that a diverse sample of students actually completed on a standardized achievement test of reading and mathematics. The teachers in this study were virtually equally adept at predicting the achievement of students with high ability and students with below average ability.

Collectively, the research on teachers' ability to judge students' academic functioning has an important practical implication: teachers, in general, can provide valid performance judgments of their students. This result is comforting and shouldn't be surprising given the number of hours teachers have to observe their students' performances. The results, however, don't say tests are unnecessary, as some teachers suggest. In order to provide meaningful information to many educational stakeholders, we will continue to need periodic achievement test results for all students, as well as teacher judgments.

Assessment is Communication!

We started this book with a dictionary-like definition of assessment. That is, *assessment is an information gathering and synthesizing process for the purpose of making decisions about students' learning and instructional needs.* We have stressed throughout this chapter that communication is a central part of, and perhaps the primary reason for an assessment. In education we want to

communicate how well students are learning to a wide array of people, including students themselves, parents, administrators, legislators, and fellow teachers. If we are going to be successful in our communication efforts, teachers must have a strong command of assessment knowledge. Without this knowledge, communication with the public and our fellow educators about student learning in the context of widely held standards will be far less meaningful and effective. In summary, think of assessment as a communication activity—one that is rich with feedback and opportunities to tell a story about student achievement and educational effectiveness!

References

Demaray, M.K., and S.N. Elliott. "Teachers' Judgments of Students' Academic Functioning: A Comparison of Actual and Predicted Performances." *School Psychology Quarterly* 13 (1998), pp. 8-24.

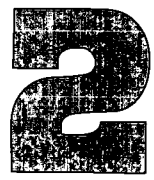
Hoge, R.D., and T. Coladarci. "Teacher Based Judgments of Academic Achievement: A Review of the Literature." *Review of Educational Research* 59 (1989), pp. 297-313.

Kean, M. *Education Assessment: A Primer for School Boards.* Monterey, CA: CTB/McGraw-Hill, 1998.

McDonnell, L.M.; M.J. McLaughlin; and P. Morison, eds. *Educating One and All: Students with Disabilities and Standards-Based Reform.* Washington, DC: National Academy Press, 1997.

Witt, J.C., et al. *Assessment of At-Risk and Special Needs Children, 2nd Edition.* Boston: McGraw-Hill, 1998.

Characteristics of Good Assessment



Good educational assessments yield "good" scores. Educational assessments come in many forms, including traditional multiple-choice tests, observations of students' work samples, and extended responses or performances. And as emphasized in the previous chapter, they serve a variety of purposes. But regardless of the type of assessment or its purpose, all good assessments should possess the characteristics of *validity*, *reliability*, and *usability*. For many readers, these are familiar terms commonly associated with tests and testing. And yet their meaning is not well understood. Too many readers will automatically assume we are about to present advanced statistics and some esoteric measurement concepts that have little to do with their teaching lives. Such an assumption is wrong! Instead, this chapter focuses on very practical concepts that are central to assessing students and using the results of any assessment with confidence. In this short but important chapter we define and discuss three characteristics of good assessments and provide some guidelines for using this information when you select or construct your own assessments.

All educators occasionally will have to explain the significance of their assessments, especially large-scale assessments mandated by a school district or the state. The involvement of students with disabilities in such assessments likely will stimulate even more inquiries about the validity and reliability of the resulting scores if testing accommodations or an alternate assessment have been used. Therefore, knowledge of validity, reliability, and usability are important in the delivery of effective assessment services.

Before examining these three key assessment concepts, let's establish how we typically use achievement tests and the resulting test scores. Basically, an achievement test is given once or possibly twice a year to a group of students with the intent of providing a score for each student that is indicative of his or her knowledge or ability in a given subject matter. *The resulting test scores are useful or good to the extent that the test (a) measures what the students have been studying in their classes and (b) the resulting scores are accurate.* To the extent that the test measures subject matter content that is different from what students have been studying, students' test scores become less meaningful as indicators of their achievement and less useful in guiding teachers' future instructional efforts. Likewise, if the students' answers do not result in a test score that can be determined consistently and accurately, teachers' confidence in the score is lessened.

In summary, we tend to find achievement tests useful when they are representative of what students have been taught and when they yield consistent, accurate scores. When these conditions have been met, we are more comfortable or confident making inferences from the resulting test score about students' classroom performances. When academic standards (like some of those in our state content standards) have influenced classroom instruction, then it is logical to also consider a possible relationship between students' test scores and such standards. That is, it is reasonable to use test scores in a subject matter area as evidence of the degree to which students have acquired the knowledge and skills specified in content standards. The next chapter, in which we focus on the Wisconsin Knowledge and Concepts Examination, will examine further the relationship or alignment among standards, tests, and instruction. For now, examine Figure 2.1 to get a picture of the connections and associated inferences between a student's test score and his or her classroom performances in mathematics, as well as the relationship between both of these and academic standards in mathematics. The inferred connections among these elements of the education system may be logical, but they are only meaningful if the resulting test scores are valid! In order for us to make sound inferences about students' achievement, it is critical that tests like that used in the WSAS yield valid test scores.

Validity

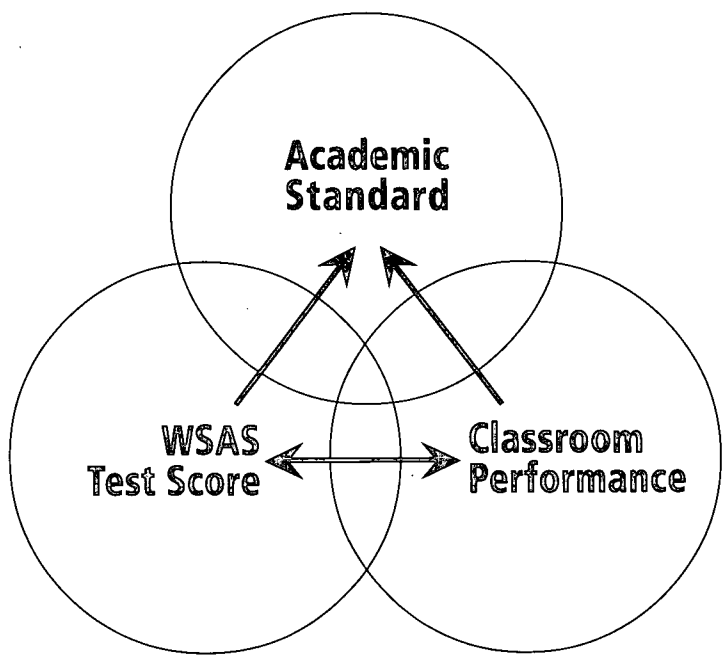
When you test a student in basic mathematics, you are testing a *sample* of that student's mathematical knowledge and skills. From the resulting test score, you make an inference about the student's ability to add, to subtract, etc. Your inference depends on the truthfulness or meaning of the test—its validity. *Validity refers to the adequacy and appropriateness of the interpretations made from assessments, with regard to a particular use.* Of all the essential characteristics of a good test, none surpasses validity. If a test is not valid for the purpose used, it has little or no value. For example, if a test designed to measure academic achievement in geography or history has questions that are phrased in difficult language, it probably does not test geography or history as much as it does reading. The test does not do a good job of measuring what it primarily claims to measure. Validity is specific. That is, a test may be valid for one purpose and not the others. For example, to administer a spelling test for the purpose of determining a student's achievement in grammar is very likely to be invalid.

Traditionally, test developers have talked about three major kinds of validity: content validity, criterion-related validity, and construct validity. A test has *content validity* if it adequately samples behavior that has been the goal of instruction. Does the test adequately represent the material that was taught? Testing a minor portion of a unit on *Hamlet* after stressing the unity of the total play greatly diminishes content validity. Determining whether a test has content validity is somewhat subjective. It usually is established when subject-matter experts and experienced teachers agree that the content covered is a representative sample of the knowledge and skills in the tested domain of knowledge and skills.

A test is said to have *criterion-related validity* if its results parallel some other external criteria. Thus, test results are similar or not similar to another sample of a student's behavior (some other criterion for comparison). If students do well on a standardized reading test that measures many

Figure 2.1

Pictorial Illustration of the Desired Relationships Among Academic Standards, Test Scores on WSAS, and Students' Classroom Performances



aspects of reading, they likewise should do well in completing and understanding geography and history assignments. Some people refer to this type of validity as predictive validity because a score from one assessment is being used to make predictions about a performance on another assessment that occurs later.

A test has *construct validity* when the particular knowledge domain or behavior said to be measured is actually measured. For example, a teacher may claim that his or her test measures application of mathematical concepts and not just mathematical computations. Therefore, a review the test should reveal that large portions of the items require students to apply results of mathematical computations using mathematical concepts correctly. To further substantiate that the test measures the application of mathematical concepts, one could look for agreement between the test results and other evidence from students' classroom activities and work samples. Construct validity is a complex issue and increasingly is coming to refer to the entire body of information about what a test measures. As you can see in our example of the assessment of mathematical applications, decisions about construct validity require information about the content of the test and the degree to which the test results relate to other measures of the same construct.

It makes no sense to prepare or select a test designed to measure something other than what has been taught if you want the results to affect instruction and provide information about student learning. As an example, we don't measure a student's height using a bathroom scale. Therefore, teachers and others should work hard to ensure that a test measures what it is designed to measure. When it does, we say it has good construct validity.

Factors influencing validity

Numerous factors can make assessment results invalid for their intended use. Some are obvious and avoidable. For example, no teacher would think of measuring knowledge of mathematics with a social studies assessment. Nor would it be logical to measure problem-solving skills in fourth grade mathematics with an assessment designed for eighth graders. In both instances, the assessments would yield invalid results.

Some of the factors that influence validity are subtle. A careful examination of test items or assessment tasks will indicate whether the assessment instrument appears to measure the subject matter content and the mental functions that the teacher is interested in measuring. However, several factors may prevent or interfere with the test items or assessment tasks functioning as intended. When this happens, the validity of the interpretations of the assessment results is diminished. Linn and Gronlund (1995) identified a list of 10 factors inherent in a test or the assessment itself that can interfere with valid results. These factors are listed and briefly described in Figure 2.2.

Factors involved in the administration and scoring of a test also may affect the validity of test results. With classroom assessments, factors such as insufficient time, unfair aid to individual students, cheating, and inaccurate scoring can lower validity. When using published tests, failure to follow the standard directions and time limits, giving students unauthorized assistance, and unreliable scoring contribute to lowering the validity of the results. Factors associated with changes in the administration of a test and the validity of the resulting scores are central to the use of testing accommodations with students with disabilities. Consequently, many teachers who administer assessments to all students will be confronted with decisions concerning the validity of the results for students with disabilities who received accommodations in the administration of a particular test or assessment.

Factors associated with students' responses to test items or assessment tasks can also affect the validity of the results. As Linn and Gronlund (1995) observed, some students may be bothered by emotional problems that interfere with their performance on a test. Others may be frightened or anxious in a testing situation and unable to respond as they normally would in daily classroom situations. Still other students may not be motivated to put forth their best effort. We are also aware that some students with disabilities may need accommodations in the response format or method for reporting answers to test items. These and other factors that change students' responses to an assessment can distort results and consequently lower validity.

Evidence of validity

Evidence of the validity of a score on a test or an assessment instrument generally takes two forms: (a) how the test or assessment instrument "behaves" given the content covered, and (b) the effects of using the test or assessment instrument.

Questions commonly asked about a test's "behavior" concern its relation to other measures of a similar construct, its ability to predict future performances, and its coverage of a content domain. Questions about the use of a test typically focus on the test's abilities to reliably differentiate individuals into groups and to guide teachers' instructional actions with regard to the subject matter covered by the test. Some questions also arise about unintended uses of a test or an assessment instrument. For example: Does use of the instrument result in discriminatory practices against various groups of individuals? Is the test used to evaluate others, such as parents or teachers, whom it does not directly assess? These questions concern a relatively new area of validity referred to as consequential validity (Green, 1998; Messick, 1989), which is discussed in greater detail in the final chapter of this book.

Criteria for evaluating the validity of tests and related assessment instruments have been written about extensively (Linn and Gronlund,

Figure 2.2

Inherent Test Factors that Influence Validity

1. **Unclear directions.** Directions that do not clearly indicate to the student how to respond to the tasks and how to record the responses will tend to reduce validity.
2. **Reading vocabulary and sentence structure too difficult.** Vocabulary and sentence structure that are too complicated for the students taking the assessment will result in the assessment's measuring reading comprehension and aspects of intelligence, which will distort the meaning of the assessment results.
3. **Ambiguity.** Ambiguous statements in assessment tasks contribute to misinterpretations and confusion. Ambiguity sometimes confuses the better students more than it does the poor students.
4. **Inadequate time limits.** Time limits that do not provide students with enough time to consider the tasks and provide thoughtful responses can reduce the validity of interpretations of results. Rather than measuring what a student knows about a topic or is able to do given adequate time, the assessment may become a measure of the speed with which the student can respond. For some content (e.g., a typing test), speed may be important. However, most assessments of achievement should minimize the effects of speed on student performance.
5. **Inappropriate level of difficulty of the test items.** In norm-referenced tests, items that are too easy or too difficult will not provide reliable discrimination among students and will therefore lower validity. In criterion-referenced tests, the failure to match the difficulty specified by the learning outcome will lower validity.
6. **Poorly constructed test items.** Test items that unintentionally provide clues to the answer will tend to measure the students' alertness in detecting clues as well as mastery of skills or knowledge the test is intended to measure.
7. **Test items inappropriate for the outcomes being measured.** Attempting to measure understanding, thinking skills, and other complex types of achievement with test forms that are appropriate only for measuring factual knowledge will invalidate the results.
8. **Test too short.** A test is only a sample of the many questions that might be asked. If a test is too short to provide a representative sample of the performance we are interested in, its validity will suffer accordingly.
9. **Improper arrangement of items.** Test items are typically arranged in order of difficulty, with the easiest items first. Placing difficult items early in the test may cause students to spend too much time on these and prevent them from reaching items they could easily answer. Improper arrangement may also influence validity by having a detrimental effect on student motivation. This influence is likely to be strongest with young students.
10. **Identifiable pattern of answers.** Placing correct answers in some systematic pattern (e.g., T, T, F, F or A, B, C, D, A, B, C, D) will enable students to guess the answers to some items more easily; and this will lower validity.

(source: Linn and Gronlund, 1995)

1995; Witt et al., 1998). A joint committee of the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education developed a comprehensive list of standards for tests that stresses the importance of construct validity and describes a variety of forms of evidence indicative of a valid test. These new *Standards for Educational Psychological Testing* (American Psychological Association, 1985) include valuable information for educators involved in testing diverse groups of students.

Key aspects of validity

Many test users and consumers of test-based information struggle with the relatively abstract concept of validity and its importance to the meaningful use of tests or assessments. Rest assured, it is the single most important characteristic of good assessment information. Keep in mind the following key aspects of validity noted by leading measurement experts (Airasian, 1994; Linn and Gronlund, 1995):

- Validity is concerned with the general question, "To what extent will this assessment information or test score help me make an appropriate decision?"
- Validity refers to the decisions that are made from assessment information, not the assessment approach or test itself. It is not appropriate to say, "This assessment information is valid" unless you also say for what decisions or groups it is valid. Keep in mind that assessment information valid for one decision or group of students is not necessarily valid for others.
- Validity is a matter of degree; it does not exist on an all-or-nothing basis. Think of assessment validity in terms of categories: highly valid, moderately valid, and invalid.
- Validity involves an overall evaluative judgment. It requires an evaluation of the degree to which interpretations and uses of assessment results are justified by supporting evidence. Educators also must consider assessment results in terms of the consequences of those interpretations and uses.

Although validity may be the most important characteristic of a good assessment, it is by no means the only characteristic you should understand. Consumers of test results also want the results to be reliable, so let's examine what reliability means with respect to test scores.

Reliability

A test is reliable to the extent that a student's scores are nearly the same on repeated measurements. It is characterized as reliable if it yields consistent scores. Suppose, for example, that a teacher has just given an achievement test to her students. How similar would the students' scores have been had she assessed them yesterday, or next week, or in a couple of months? How would the students' scores have differed if she had selected a different sample of tasks to test? How much would the scores have differed if another person scored the test? These are the types of questions with which reliability is concerned.

Remember, assessment results merely provide a limited measure of performance obtained at one point in time. Some error always exists in any test or assessment since fluctuations in human behavior are not totally controllable, and the test itself may contain possibilities of error. As errors in measurement increase, the reliability of a test decreases. Unless an assessment can be shown to be reasonably consistent over different occasions, different raters, or with different samples of tasks from the same subject matter, we can have little confidence in the results.

Carefully note the relationship and distinction between reliability (consistency) and validity (meaningfulness). A valid test must be reliable, but a reliable test need not be valid. In other words, *reliability is a necessary but not sufficient condition for validity*. For example, giving an algebra test to first or second graders will produce consistent results, but the results are not meaningful for six-year-olds. Thus, the test would be reliable, but not valid.

Reliability is primarily statistical, but please don't let that turn you off to learning more about it. It is important if you are going to be involved in using test results, and essential if you are ever going to design and conduct an alternate assessment for a student with a severe disability. The logical analysis of an assessment will provide little evidence concerning the reliability of the resulting scores. To evaluate the consistency of scores assigned by different raters, two or more raters must score the same set of student performances. Similarly, an evaluation of the consistency of scores obtained in response to different forms of a test or different collections of performance-based assessment tasks requires the administration of both test forms or collections of tasks to an appropriate group of students. Whether the focus is on

inter-rater consistency or the consistency across forms or collections of tasks, consistency may be expressed in terms of shifts in the relative standing of students in the group or in terms of the amount of variation to be expected in a student's score. We report consistency in the case of inter-rater judgments or across forms of a test by means of a correlation coefficient. In the case of the expected amount of variation in a given student's test score, however, we report consistency by means of a statistic called the standard error of measurement. Both of these methods of expressing reliability are widely used and educators responsible for communicating the results of assessments should understand them.

Correlations can range between +1.0 and -1.0, where +1.0 indicates perfect agreement between the magnitudes of the scores for the same individual. The case of a test-retest approach to reliability is illustrated in Figure 2.3. Given that most teachers do not repeatedly administer a test, alternative methods of estimating the reliability of a test, such as internal consistency, must be used. The latter method uses a slightly different formula for calculating a reliability coefficient (referred to as coefficient alpha). Regardless of the method for quantifying the reliability of a test, most experienced users of teacher-constructed tests consider reliability coefficients in the +.80 or higher range to be essential. Many published tests have reliability coefficients in the +.90 range.

The *standard error of measurement* (SEM) is an estimate of the variation expected in a student's score if the student is given the same test over and over. The amount of variation in the scores is directly related to the reliability of the assessment procedures. Low reliability is indicated by large variations in the resulting scores, and high reliability by little variation in the scores.

It is impractical to repeatedly administer the same test to a student. Fortunately, however, it is possible to estimate the amount of variation in the resulting scores. This estimate of the variation in scores is the SEM. The calculation of the

Figure 2.3

Example of How to Calculate Test-Retest Reliability

Test-Retest Reliability of a Kindergarten Screening Test

	Number of Answers Correct	
Student (N)	**Test (X)**	**Retest (Y)**
1	9	10
2	7	6
3	5	1
4	3	5
5	1	3

$$r_{xy} = \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{\left[N\Sigma X^2 - (\Sigma X)^2\right]\left[N\Sigma Y^2 - (\Sigma Y)^2\right]}}$$

$$r = \frac{150}{\sqrt{(200)(230)}}$$

$$r = .70$$

SEM for a test is beyond the scope of this book (but if you are really interested in how it is done, see Appendix B), and besides, most manuals of published tests provide specific standard errors of measurement. All you need to do is be able to apply your knowledge of SEMs when interpreting a student's test results. It is a wise practice to interpret a test score as a band of scores (which most people call a confidence band) rather than as a specific score. The next chapter will provide more details about SEMs and confidence bands.

Factors influencing reliability

Although teachers seldom find it possible or useful to calculate reliability coefficients or SEMs, they should be cognizant of factors that can influence assessment results. Two such factors are the number of items or tasks on a test and the objectivity of the scoring of the items or tasks.

In general, the larger the number of tasks on an assessment, the higher the reliability will be, because a longer assessment will provide a better sample of the knowledge and skills being measured. In addition, the scores are less likely to be distorted by chance factors.

Objectivity of an assessment refers to the degree to which equally competent scorers obtain the same results for the same students. Most of the published tests educators use are high in objectivity, and are often scored by machines or highly trained scorers. In general, tests featuring selected-response items can be scored more reliably than constructed-response items. Concerns about the reliability of scores, frequently voiced as issues of bias or fairness, often have been used to argue against the use of complex constructed-response type tasks on achievement tests. However, with training it is possible to get highly reliable scores for written essays or performance tasks with multiple parts.

Key aspects of reliability

We can conclude our examination of reliability, then, by saying that unless a test is reasonably consistent on different occasions or with different samples of the same behavior, we can have very little confidence in its results. A variety of factors, some concerning the student taking the test and others inherent in the test's design and content, can affect the reliability of a test. Student characteristics affecting a test's reliability include guessing, test anxiety, and practice in answering items like those on the test (Witt et al., 1998). Characteristics that can influence reliability include a test's length (longer tests are generally more reliable), homogeneity or similarity of items (more homogeneous tests are usually more reliable), and time allotted (speed tests are typically more reliable than unbound tests).

In conclusion, when considering the reliability of any test or assessment process, keep the following three points in mind:

- Reliability refers to the stability or consistency of assessment information, not the appropriateness of the assessment information collected.
- Reliability is a matter of degree; it does not exist on an all-or-none basis. It is expressed in terms of degree: high, moderate, or low reliability.
- Reliability is a necessary, but not sufficient, condition for validity. An assessment that provides inconsistent results cannot be relied upon to provide useful information. If important educational decisions are to be made from a test, the resulting score(s) must be highly reliable.

Usability

So far we have argued that good assessments should measure what they say they measure and that the measurements must be consistent—that is, good assessments are valid and reliable. Good assessments also must be usable. This may seem like an obvious point, but educators should not overlook it when designing or selecting an assessment, particularly when the assessment involves a large number of children. For example, in our statewide assessment system, the WSAS, over 200,000 students are eligible to take the Wisconsin Knowledge and Concepts Examination each year. Thus, issues concerning ease of administration, interpretation and application, time required to administer the test, and cost should be weighed against alternative ways of getting the same information and the resulting consequences.

Unlike the concepts of validity and reliability, there is no general set of guidelines or statistical indices used to determine the usability of a test or assessment program. A wide array of variables influence decisions about usability, and often they are the subject of hot debate. However, the more closely an assessment is aligned with what is taught in classrooms, the fairer and more usable it "feels" to the teachers and students on the front lines. Figure 2.4 illustrates the concept of alignment between what is taught, what is tested, and what is valued in the system. Alignment of

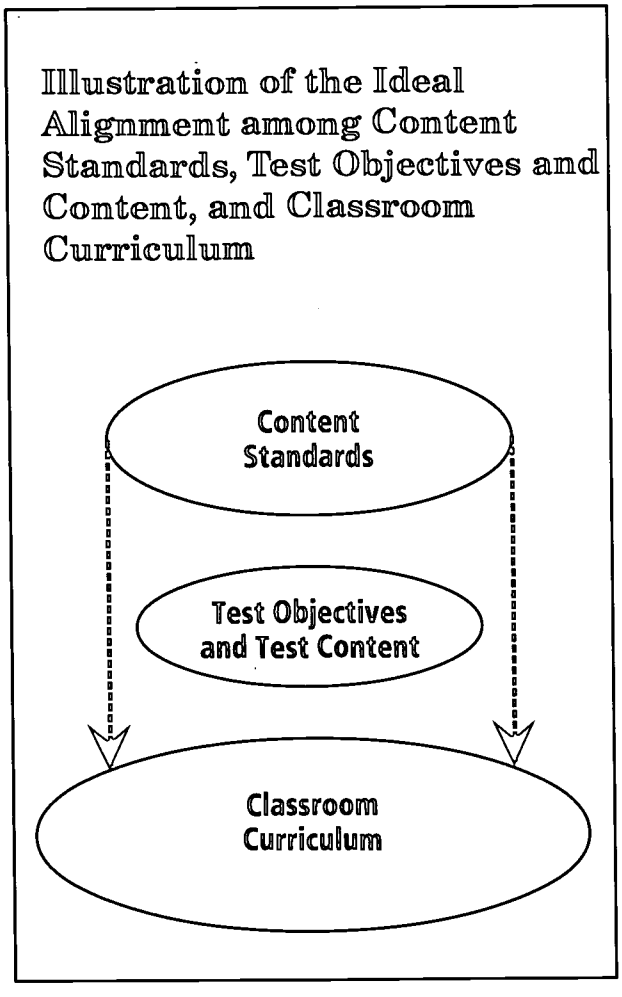
ERIC

BEST COPY AVAILABLE

assessments with classroom instruction and state standards is an essential usability factor. The next chapter covers more about this important concept.

Another key usability issue concerns how the results of an assessment are communicated. When results are stated in understandable terms to most consumers, but especially teachers, it increases the likelihood that they will facilitate teachers' instructional efforts and advance an understanding of their own abilities for students and their parents. An example of this is to report scores as proficiency levels or categories. Related to *how* results are communicated is the issue of *when* results are communicated. For feedback of any kind to be useful, it must occur close in time to the performance of interest. Far too often, test results—particularly those from large-scale assessments—come months after the testing event occurred and with little time to focus on remediation efforts.

Figure 2.4



Applying Knowledge of Good Assessments to Your Work

As emphasized in this chapter, good assessments are valid, reliable, and usable. Many educators have translated this "holy trinity" of measurement to mean that a test must measure what it says it measures and do so in a way that is practical and results in consistent scores. This is an acceptable translation, but perhaps a bit of an oversimplification of the judgments required of persons involved in using an assessment. Recall that validity is not an all-or-none characteristic of an assessment, but matter of degree. Also remember that reliability is a necessary but not sufficient condition of validity. Ultimately, a statement about the validity of an assessment involves an evaluative judgment of the degree to which interpretations and uses of the assessment results (scores or proficiency statements) are justified.

To make decisions about the degree to which an assessment yields valid results, it is useful to ask four questions:

The Content Question. How well does the sample or collection of assessment tasks *represent* the domain of tasks to be measured? For most teachers this question is answered by reviewing copies of tests and comparing the items to what they teach. The greater the similarity, the more confidence they have that the test measures what they value.

The Test-Criterion Relationship Question. How well do students' performances on the assessment *predict* future performances or *estimate* current performances on some valued measure of the knowledge and skills other than the test itself? For most teachers, this question is answered by comparing the assessment results with another measure of performance, such as classroom tests or summary observations by the teacher. The greater the similarity between the test and teachers' other criterion of performance, the more confidence teachers have in the test scores.

The Construct Question. How well can teachers interpret performance on the assessment as a meaningful measure of the knowledge and skills the assessment purports to measure? For most teachers, answers to this question will be out of reach, because it requires establishing the meaning of the assessment by experimentally

determining what factors influence students' performances. Many educators will fall back on their review of the content and test-criterion relationships as evidence that the test measures a specific construct. Construct validation takes place primarily during the development of a test and is based on an accumulation of evidence from many sources. If one is using a published test or assessment program to measure a particular construct such as mathematical reasoning or reading comprehension, then one will find the necessary evidence on the construct validity of the instrument included in a technical manual.

The Consequences Question. How well does use of the assessment results accomplish the intended purposes of the assessment and avoid unintended effects? If an assessment is intended to contribute to improved student learning, the consequences question becomes deceivingly simple: "Does it?" In trying to answer this question, teachers typically pose many more questions. For example, "What impact does the assessment have on teaching? What are the possible negative, unintended consequences of the use of the assessment results?" As you can see, there is no short or easy answer to the consequences question. Nevertheless it is worthwhile to address it. In fact, it is often the first question many educators ask when confronted with a large-scale assessment program. We will revisit the topic of consequential validity in the last chapter of this book, after you have had a chance to learn more about the intended uses of the WSAS and the use of testing accommodations for students with disabilities.

Next to validity, reliability is the most important characteristic of a good assessment. Reliability provides the consistency that makes validity possible and it indicates the degree to which various kinds of generalizations are reasonable. High reliability is essential when test results are going to be used to make final decisions that concern individual students and have lasting consequences. Under these conditions, the tests or assessments used should have a very small standard error of measurement and one should be able to readminister and rescore them to establish the consistency of the score(s), especially if a student's original score is below a critical cut-point. Lower reliability is tolerable when the test results are used to make reversible decisions of relatively minor importance, and when the decision is confirmable by other data.

Finally, it is not enough to have tests or assessments that yield valid and reliable scores. The tests or assessments also must be usable. That is, persons with limited assessment training must be able to administer them and the tests must be constructed to allow a wide range of students to participate in the assessment. Of course, time and costs are also important usability factors, as is the ease of interpretation. Ultimately, issues of usability influence validity; that is, if educators do not use an assessment as designed, they are unlikely to achieve the intended purpose of the assessment.

Many readers of this book will be working with students with disabilities and trying to facilitate their meaningful involvement in state and district assessment programs. As a result, they will find themselves having to make a number of decisions about the validity of assessment results. Specifically, when students need testing accommodations, teachers will be expected to select and implement accommodations that do not invalidate test results. The use of a testing accommodation, in fact, is intended to enhance the validity of the test score for the student with a disability. In addition, when a student cannot meaningfully participate in the regular assessment (e.g., Wisconsin Reading Comprehension Test or Wisconsin Knowledge and Concepts Examinations) given to the majority of students, teachers and their fellow IEP team members will be responsible for conducting an alternate assessment. In many cases, teachers will construct these alternate assessments for an individual student. The alternate assessments, however, still will need to be valid and reliable. Consequently, knowledge of the characteristics of a good assessment is critical to using test results and to facilitating the meaningful participation of all students in large-scale assessment programs.

In conclusion, issues pertaining to decisions about validity of test results start before a test is given, are ongoing after a test is completed, and are always relative to the stated purpose of the test. As you can see, the typical and seemingly straightforward question, "Is the test valid?" requires some technical knowledge to answer and is actually inappropriately worded. A better question, and one you should be equipped to address, is, "Is the test a good test?"

References

Airasian, P.W. *Classroom Assessment, 2nd Edition.* Boston: McGraw-Hill, 1994.

American Psychological Association. *Standards for Educational and Psychological Testing.* Washington, DC: American Psychological Association, 1999.

Green, D. R. "Consequential Aspects of the Validity of Achievement Tests: A Publisher's Point of View." *Educational Measurement: Issues and Practice,* Summer (1998), pp. 16-19.

Linn, R.L., and N.E. Gronlund. *Measurement and Assessment in Teaching, 7th Edition.* Englewood Cliff, NJ: Merrill/Prentice Hall, (1995).

Messick, S. "Validity." In *Educational Measurement, 3rd Edition.* Ed. R.L. Linn. New York: Macmillan, 1989.

Witt, J.C., et al. *Assessment of At-Risk and Special Needs Children, 2nd Edition.* Boston: McGraw-Hill, 1998.

Understanding and Using the Wisconsin Student Assessment System (WSAS)

3

This chapter will provide you with an understanding of the Wisconsin Knowledge and Concepts Examinations (WKCE). Understanding the WKCE will help you do two things: first, you can better align curriculum, instruction, and assessment to improve outcomes. Second, knowledge of the WKCE content and results will help you decide whether, and how, to include students with disabilities in the examinations.

This chapter has three sections. First, it will explain why the WKCE is important. Second, it will describe the types of results reported on the examinations. Third, the chapter gives you an opportunity to apply your knowledge of the WKCE to understanding sample outcomes. It concludes with some common questions and answers regarding the WKCE.

Why Assess?

Although some educators embrace assessment, most view state assessments as a necessary evil—or just plain evil. Educators' antipathy for mandated assessment is understandable, as such programs often are forced on them by external agencies, and may be used for many purposes that educators do not embrace, such as rating school districts or determining student promotion and graduation. However, there are two reasons why educators engage in assessment programs. The first is that you should; the second is that you must.

Why you should assess. Effective schools coordinate three features to enhance educational success: Curriculum, Instruction, and Assessment (CIA). It is essential for schools to carefully align curriculum, instruction, and assessment to enhance the performance of individual students. When the three are not aligned, schooling is less effective and incomplete. When these three factors are aligned, more students understand what is expected, teachers understand what to teach, and schooling is more effective. Assessment of student progress is an essential catalyst for aligning curriculum and instruction. One might say good assessment functions as a "CIA agent"—that is, assessment stimulates alignment of curriculum and instruction to insure student learning. Figure 3.1 illustrates appropriately and inappropriately aligned curricula.

There are at least two ways to align curriculum, instruction, and assessment. The first method is to delineate curricula so that teachers, administrators, parents, and students understand its scope and intent. When curricula are clearly delineated, then teachers can select instructional practices to promote the outcomes specified in the curricula. The final step in this method of instructional alignment is assessment. That is, after educators specify the curriculum students are to master, and provide instructional activities to promote mastery, they must assess students' performance on curricular objectives. This last step tells educators the degree to which they have been successful. It also informs students, parents, and the community at large of the effectiveness of schooling. Wisconsin educational standards are intended to stimulate this "top-down" alignment process. That is, by telling the public, educators, and students the content students are to master at various stages of educational progress, the state intends to stimulate local school districts and educators to align their curricula and instructional practices in order to achieve state standards. The WKCE measures how well schools do in helping children meet state standards; assessment stimulates accountability to insure curriculum-instruction alignment.

However, there is another way to align curriculum, instruction, and assessment. This method for alignment begins with assessment. That is, educators start by assessing student performance. Although beginning at the end of the CIA appears illogical, it can be a powerful way for educators to take control of student learning. In fact, the most

effective school reform typically begins with educators clarifying the outcomes they desire from students rather than beginning with curricula and teaching practices (Newmann, Marks, and Gamoran, 1995). Thus, educators can use the assessment of educational outcomes as a starting point, rather than an end point, for CIA alignment. Understanding the results of the WKCE can help educators achieve that alignment, and, in turn, better educational outcomes for students.

Why you must assess. Even if one does not accept the need for assessment as an essential ingredient in CIA alignment, Wisconsin educators must assess ALL students. The state legislature mandates that children in fourth, eighth, and tenth grades take the WKCE, and that the results of these assessments be reported to the public. Beginning in 2002, educators also will use this assessment in part to determine promotion to the next grade.

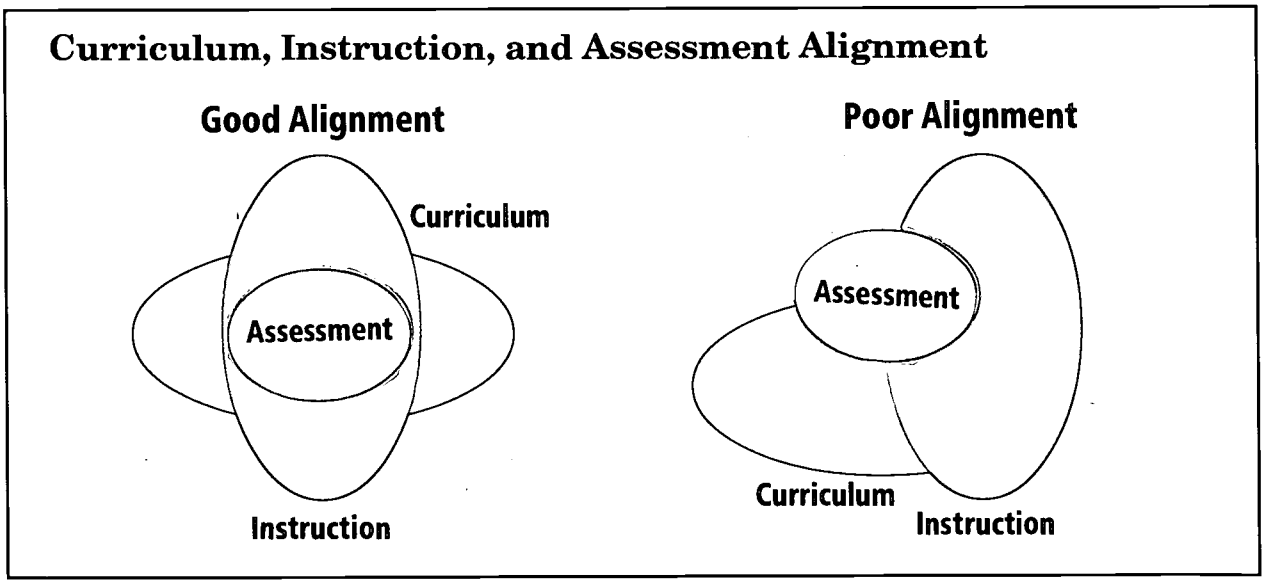
Whereas state mandates require assessment of school children, federal special education mandates require participation of all students in districtwide and statewide assessment programs when possible. To meet that goal, states must include the majority of students with disabilities in district and state tests or an alternate assessment. Thus, not only does Wisconsin mandate inclusion of students in the fourth, eighth, and tenth grade WSAS processes, state and federal guidelines mandate participation of students with disabilities in WSAS whenever possible. One goal of this chapter is to enhance your knowledge of WKCE content so that you can make effective decisions about the inclusion of students with disabilities in these examinations.

WSAS Structure and Content

Before explaining WSAS results, it is a good idea to take a moment to describe the structure of the WSAS. First, the WSAS is not the same as the WKCE. The WSAS is, as the name implies, an assessment system or process for describing student performance on Wisconsin's education standards. Originally, the system was to have three components: (a) the Knowledge and Concepts Examinations, (b) state-developed performance assessments, and (c) district-administered and maintained assessments. District assessments include student portfolios, districtwide assessments such as those either developed or purchased by districts to supplement WSAS, and alternate assessments customized to assess the progress of students with exceptional educational needs.

The Wisconsin State Legislature funded only the Wisconsin Knowledge and Concepts Examinations; consequently, many educators assume that the WSAS *is* the WKCE, as that examination is the only enacted element of the WSAS. Although it is not clear that the state will ever fully implement the original framework of examinations, performance assessments, and district assessments, it is a useful structure for drawing a distinction between the WSAS (in which *all* students must participate), and the WKCE (in which at least 98 percent of students should participate).

Figure 3.1



Wisconsin buys the WKCE from test publishers; we do not make our own. Currently, Wisconsin purchases the examinations—called the TerraNova or "new ground" tests—from a test publisher (CTB/McGraw-Hill) to assess Wisconsin students at fourth, eighth, and tenth grades. These tests cover some of the academic standards in reading, language arts, science, social studies, and mathematics. Each subject-matter examination has about 40 items in each domain. The exception is the reading and language arts domains, which are jointly assessed by about 70 items and a written language sample, or student essay.

Two features of items are critical to understanding tests: item format (how the item asks a question), and item content (what the item asks about). Wisconsin's Department of Public Instruction (DPI) selected the TerraNova in part because its items demand that students construct a response, not just select one from an array of choices. Although some items are multiple choice, many are short answer, correcting a passage, or essay responses.

Item content, or, "what the test tests" also is important to understanding tests. The content of test items relates to academic objectives. For example, a language arts objective might be "analyze text," which is shown by drawing conclusions, inferring relationships, and identifying theme and story elements; a mathematics objective might be "data analysis, statistics, and probability," which is shown by analyzing, interpreting, and evaluating data, and applying concepts and processes of data analysis, statistics, and probability to real-world situations. The TerraNova items used for the WKCE are designed to assess seven reading/language arts objectives, nine mathematics objectives, seven science objectives, and four social studies objectives at each (fourth, eighth, and tenth grade) level. Figure 3.2 lists the academic objectives in the reading and language arts domain the current examinations assess using TerraNova.

One other aspect of item content is the thinking skill that is demanded from the student to answer the question. Another reason the DPI selected TerraNova is that it contains many items that require students to organize, analyze, synthesize, and evaluate—not just to recognize and regurgitate. The thinking skills tapped by TerraNova, in order of increasing complexity, are

- gathering information,
- organizing information,
- analyzing information,
- generating ideas,
- synthesizing elements, and
- evaluating outcomes.

Figures 3.3 and 3.4 provide examples of WKCE examination content from the TerraNova tests. As you look at these examples, ask yourself three questions:

- What kind of response does the item require from the student?
- What academic objective or skill does the item require from the student?
- What thinking skill does the item demand from the student?

By asking yourself these questions, you will understand how these examinations integrate item response formats, academic skills, and thinking skills into the assessment of Wisconsin students. You can find more detail on TerraNova item formats, academic content, and thinking skills in *"Teacher's Guide to TerraNova"* (CTB/McGraw-Hill, 1997).

Understanding WSAS Results

Types of Results

When students take the WKCE, their performance is determined by the number and difficulty of the questions they correctly answer. However, reporting the number of correct answers to parents and students is not very useful. For example, saying your child got 30/40 items correct does not tell you much about how well your child did. If the test was very difficult, the 30/40 might represent an exceptionally good performance; if the test was exceptionally easy, 30/40 might be failing. Likewise, if the standard for accuracy is 50 percent, 30/40 is good; but if the standard is 90/100, 30/40 is poor.

To understand a test score, you need to know two things: how the score compares to other students' scores, and how the score compares to a given performance standard. Scores that tell how a child does relative to other children are called "norm-referenced" scores. Scores telling how a child does relative to a performance standard are called "criterion-referenced" scores. Neither type of score is sufficient to explain performance. Knowing a racer finished fifth in a 10K race, or knowing a salesperson sold $250,000 in products one month, is only part of the story. You need to know the racer's time to fully understand whether the racer ran well or just was matched against weaker runners; likewise, $250,000 in sales may be good, average, or poor relative to other salespeople. Both norm- and criterion-referenced re-

ports are necessary to understand an individual's score; neither type of report is sufficient.

For example, consider the following reports given to parents of two school children. The first parent might be told her child is in the 90th percentile relative to other children in the United States. That statement conveys how well her child scored relative to other children taking the same test. However, that report does not convey what her child knows how to do. It only conveys the child's relative, or normative, position on the test. The second parent might be told that his child understands 42 sight words. This statement conveys information about how well his child has mastered some important pre-reading skills. This helps tell him what the child has learned, but it does not tell him where the child is relative to others of the same age or level of education. That is, the second parent might not know if 42 sight words is a good performance or a poor performance

Figure 3.2

Reading/Language Arts Objectives and Skills

The Reading/Language Arts part of the test measures the skills—reading comprehension, language expression, vocabulary, and reference skills—that are essential for effective communication. Directions, passages, and test questions are linked by themes that provide context and stimulate interest.

Comprehension items focus on the central meaning of a passage rather than on surface details. Items reflect the reading process by moving from initial understanding through interpretation, and on to evaluation and application.

Essential language usage skills such as verb tense, subject-verb agreement, and basic sentence formation are measured as are sentence-combining and paragraph-writing skills. Listed below are the Reading/Language Arts Objectives measured by the Wisconsin Knowledge and Concepts Examinations. The objective statements in italics indicate the processes measured by short- answer (constructed-response) items only.

02 Basic Understanding

- Demonstrate understanding of the literal meaning of a passage through identifying stated information, indicating sequence of events, and defining grade-level vocabulary.
  -Write responses to questions requiring literal information from passages and documents.

03 Analyze Text

- Demonstrate comprehension by drawing conclusions; inferring relationships such as cause and effect; and identifying theme and story elements such as plot, climax, character, and setting.
  -Write responses that show an understanding of the text that goes beyond surface meaning.

04 Evaluate and Extend Meaning

- Demonstrate critical understanding by making predictions; distinguishing between fact and opinion, and reality and fantasy; transferring ideas to other situations; and judging author purpose, point of view, and effectiveness.
  -Write responses that make connections between texts based on common themes and concepts; evaluate author purpose and effectiveness, and extend meaning to other contexts.

05 Identify Reading Strategies

- Demonstrate awareness of techniques that enhance comprehension, such as using existing knowledge, summarizing content, comparing information across texts, using graphics and text structure, and formulating questions that deepen understanding.
  -Write responses that interpret and extend the use of information from documents and forms, and that demonstrate knowledge and use of strategies.

07 Sentence Structure

- Demonstrate an understanding of conventions for writing complete and effective sentences, including treatment of subject and verb, punctuation, and capitalization.
- Demonstrate an understanding of conciseness and clarity of meaning in combining two sentences.

08 Writing Strategies

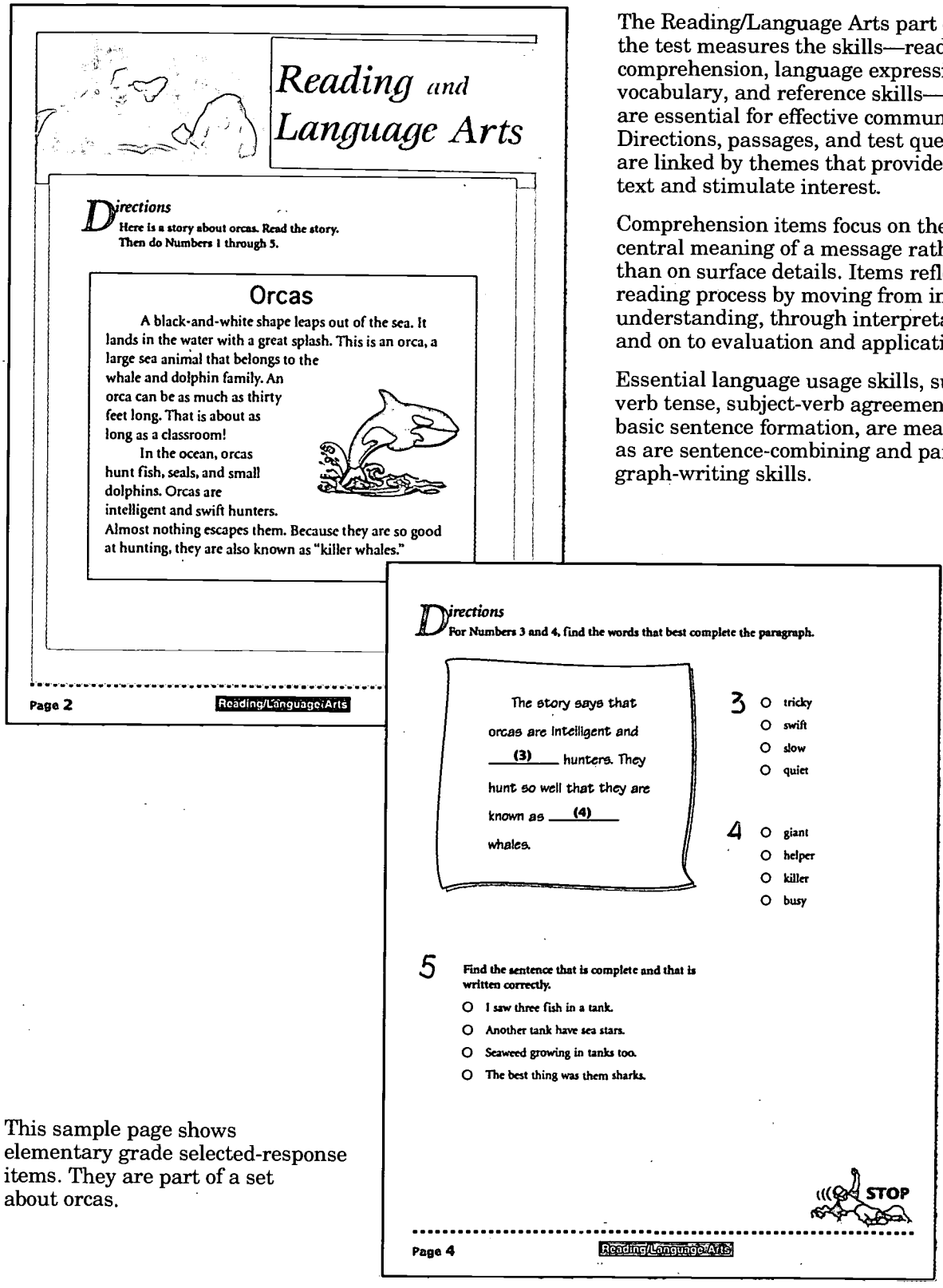
- Demonstrate knowledge of information sources, outlines and other pre-writing techniques.
- Demonstrate an understanding of the use of topic sentences, concluding sentences, connective and transitional words and phrases, supporting statements, sequencing ideas, and relevant information in writing expository prose.

09 Editing Skills

- Identify the appropriate use of capitalization, punctuation, nouns, pronouns, verbs, adjectives, and adverbs in existing text.
  -Demonstrate knowledge of writing conventions and sentence structure through identifying and correcting errors in existing text and in text written by the student.

Figure 3.3

Items in Student/Parent Pre-Test Guide for WSAS Grade 4



Reading and Language Arts

Directions
Here is a story about orcas. Read the story. Then do Numbers 1 through 5.

Orcas

A black-and-white shape leaps out of the sea. It lands in the water with a great splash. This is an orca, a large sea animal that belongs to the whale and dolphin family. An orca can be as much as thirty feet long. That is about as long as a classroom!

In the ocean, orcas hunt fish, seals, and small dolphins. Orcas are intelligent and swift hunters. Almost nothing escapes them. Because they are so good at hunting, they are also known as "killer whales."

Page 2 Reading/Language Arts

Directions
For Numbers 3 and 4, find the words that best complete the paragraph.

The story says that orcas are intelligent and ___(3)___ hunters. They hunt so well that they are known as ___(4)___ whales.

3
- O tricky
- O swift
- O slow
- O quiet

4
- O giant
- O helper
- O killer
- O busy

5 Find the sentence that is complete and that is written correctly.
- O I saw three fish in a tank.
- O Another tank have sea stars.
- O Seaweed growing in tanks too.
- O The best thing was them sharks.

STOP

Page 4 Reading/Language Arts

This sample page shows elementary grade selected-response items. They are part of a set about orcas.

The Reading/Language Arts part of the test measures the skills—reading comprehension, language expression, vocabulary, and reference skills—that are essential for effective communication. Directions, passages, and test questions are linked by themes that provide context and stimulate interest.

Comprehension items focus on the central meaning of a message rather than on surface details. Items reflect the reading process by moving from initial understanding, through interpretation, and on to evaluation and application.

Essential language usage skills, such as verb tense, subject-verb agreement, and basic sentence formation, are measured, as are sentence-combining and paragraph-writing skills.

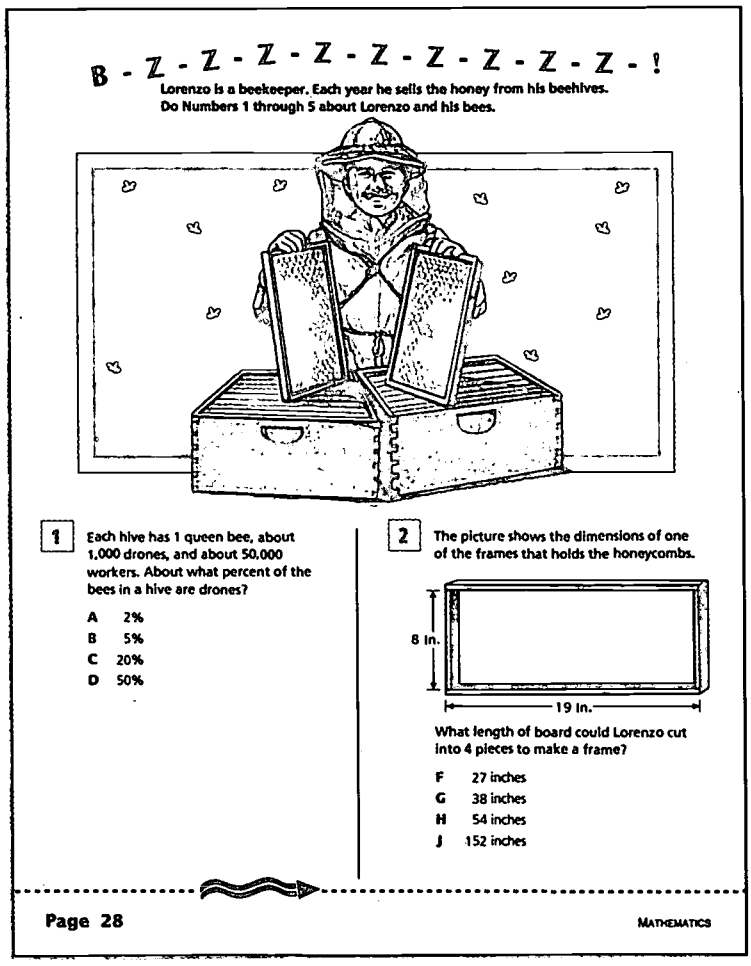
relative to other children. Just as the parent who received the first report does not understand what the child can do, the parent who received the second report does not understand where the child is relative to others. Thus, both types of information are necessary to explain a child's score.

Norm-Referenced WSAS Scores

Many norm-referenced scores are reported by the WKCE. The most basic norm-referenced score is a student's rank or standing. The statement, "My child finished fifth on a test" is likely to prompt congratulations, and a question: "How many others took the test?" However, ranks are cumbersome when large numbers of children are involved. For example, learning your child tied for 14,458th with 229 other children around the country in a year when 36,422 children took the test might tell you exactly how your child compares to the norm group, but it is not easy to understand. Consequently, norm-referenced scores are reported in ways that allow you to understand a child's rank or standing without knowing the number of people who took the test. The most popular types of norm-referenced scores are percentiles, normal curve equivalents, standard scores, and stanines. Each of these is explained in the following paragraphs.

Figure 3.4

Items in Student/Parent Pre-Test Guide for WSAS Grades 8 and 10


B - Z - Z - Z - Z - Z - Z - Z - Z - Z - !

Lorenzo is a beekeeper. Each year he sells the honey from his beehives. Do Numbers 1 through 5 about Lorenzo and his bees.

1 Each hive has 1 queen bee, about 1,000 drones, and about 50,000 workers. About what percent of the bees in a hive are drones?

A 2%
B 5%
C 20%
D 50%

2 The picture shows the dimensions of one of the frames that holds the honeycombs.

8 in.
19 in.

What length of board could Lorenzo cut into 4 pieces to make a frame?

F 27 inches
G 38 inches
H 54 inches
J 152 inches

Page 28 MATHEMATICS

Mathematics test questions allow students to use different strategies and take different paths to find the solutions. The test taps broad mathematical knowledge, beginning with computation and estimation and followed by applied mathematics and number theory. Many questions call for critical thinking, reasoning, and mathematical problem solving. The use of real-world settings and engaging art helps to involve students.

Calculators are not used in the first section of Part 1 and are optional for the rest of the mathematics test. Test questions have been designed and adequate time is provided so that using a calculator will not offer any particular advantage.

This sample page shows two selected-response questions for middle school. These questions are from a set about beekeeping.

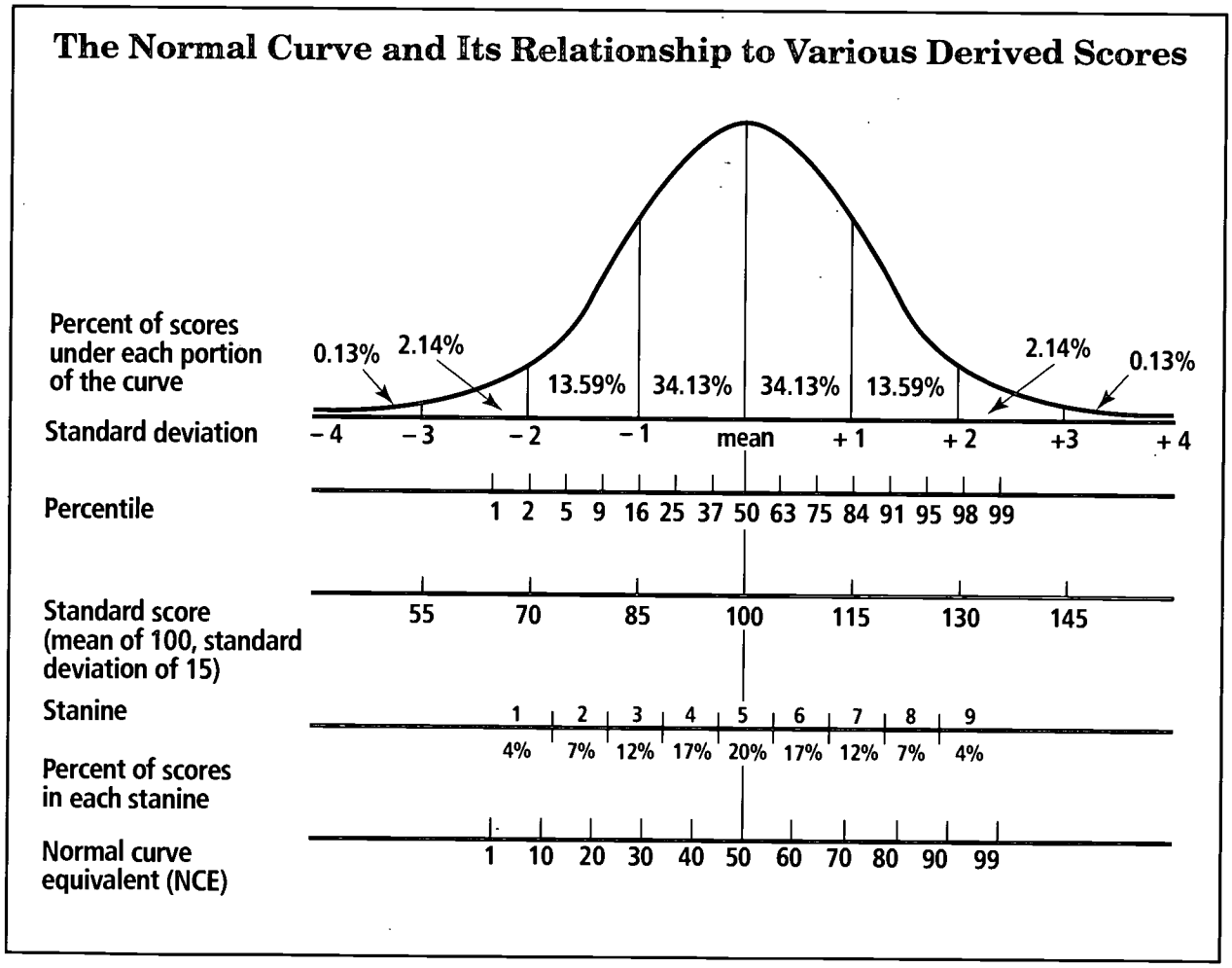
BEST COPY AVAILABLE

Percentiles. Percentiles are a norm-referenced score between 1 and 99. A percentile represents the proportion, or percentage, of children who scored equal to or worse than the child. A child at the 25th percentile is a child whose score was equal to, or better than, 25 percent of the children who took the test. Usually, percentiles are reported relative to a national normative group that represents the demographic characteristics of the United States. In the past, WKCE scores also were reported relative to other Wisconsin children, so that a student's score would be reported as a national percentile and as a state percentile. This is no longer the case; percentiles now are reported only relative to the national norm group.

Figure 3.5 shows a typical distribution of scores on a test. The figure is designed to show low test scores on the left-hand side of the figure, and high test scores on the right hand side. As the line moves from bottom to top, more students are indicated. Therefore, the small space between the bottom line of the figure and the curve at the left end means few students had very low scores. Likewise, the small space between the curve and the bottom line means few students had very high scores. The large space between the curve and the line in the middle of the figure means lots of students have average scores.

Norm-referenced scores are shown in lines below the curve. Notice that, although percentiles are convenient and easily understood, they are not equally spaced. For instance, the difference between children at the 45th percentile and those at the 50th percentile is smaller than the difference between children at the 94th percentile and those at the 99th percentile. In fact, the gap between the first and second percentile is about equal to the difference between the 37th and 50th percentiles. Thus, percentiles give rank order, but they are insensitive to how far apart children are.

Figure 3.5



Normal Curve Equivalents. A normal curve equivalent (NCE) is a two-digit score also between 1 and 99. However, a normal curve equivalent is an equal interval scale. It defines how well a child scores relative to the middle of the norm group, and does so in equal units. The middle, or the mean or arithmetic average, of the norm group is set to a score of 50. (Much like the Celsius scale arbitrarily sets 0 to the freezing point of water, and 100 to the boiling point, the NCE scale arbitrarily sets the midpoint of a distribution to 50.) The average spread of individuals about this mean is set to be 21.06. Therefore, a child whose NCE is 30 is about one standard deviation below the mean of 50. A child whose NCE is 85 is about 1.75 standard deviations above the mean. Normal curve equivalents are more consistent than percentiles for describing a child's position relative to the norm group, because NCEs are equally spaced. That is, the difference between NCEs of 30 and 35 is the same as the difference between 50 and 55, or 85 and 90 (see for yourself by looking at Figure 3.5). They are better than percentiles, because they reflect position in the norm group using equal units across scores (i.e., NCEs provide rank order *and* distances between scores). However, normal curve equivalents are not widely understood. Thus, NCEs are used mostly by professionals to understand children's scores relative to a normative group.

Standard Scores. Another way to reflect student scores relative to the norm group is with standard scores. These scores are essentially the same kind of scores as NCEs, but they set the midpoint and standard deviation of the distribution to different values. This is similar to the differences in Celsius and Fahrenheit scales; they each have different values for the freezing point of water (0 versus +32 degrees), and have different spacing between degrees (one degree on the Celsius scale is nearly two degrees on the Fahrenheit scale). Most standard scores fix the mean to 100, whereas NCEs fix it to 50; the standard deviation of most standard scores is fixed to be 15, versus 21.06 for NCEs. A quick glance at Figure 3.5 shows how the scales compare in describing position on the normal curve. The WKCE and most other group achievement tests use the NCE scale to describe score position, whereas most intelligence and individually-administered achievement tests use the standard score scale to describe score position. The reason for this is strictly habit. Just as you can translate degrees Fahrenheit to degrees Celsius, you can translate scaled scores to NCEs, and vice versa, using simple algebra.

Stanines. Stanines are yet one other way to show a score in a form that expresses rank and relative distance between scores. Instead of dividing up the range of scores from 1-99 (as NCEs do), or from 55-145 (as standard scores do), stanines divide the range of scores into nine equal, or standard, units. (This division is actually how stanines got their name: standard + nine = stanine.) This method simplifies the task of reporting where students are in the distribution, but there is a cost. The intervals between stanines are fairly crude, and so stanines are less precise descriptions for where students fall than either NCEs or standard scores. Note that the distance between stanines is constant, except that the lowest (1) and highest (9) stanines are open-ended.

Grade Equivalents. If you use grade equivalents, you may wonder why they are included in this section. Don't grade equivalents describe where a student's score falls in the curriculum? Doesn't a grade equivalent of 3.2 mean a child has mastered the curriculum up to the second month of third grade? Isn't a fourth grader who earns a grade equivalent of 6.8 about 2 to 3 years ahead of curricular expectations? Don't you love a series of rhetorical questions? The answer to all these questions is "No!"

Grade equivalents have nothing to do with grade-level expectations or with mastery. A grade equivalent is merely the midpoint of a distribution of scores for children in a given grade. To say a score is at the 4.3 grade level is to say the score was equal to the average score for a group of fourth graders who took the test in the third month of the year (i.e., 4 [grade year] + .3 [month] = 4.3). Grade equivalents are median scores defined so that half of the children in a given grade group will score below the equivalent, and half of them will score above the equivalent. In other words, half of all children in the nation are below grade level (and, by definition, half are above grade level). No matter how well or poorly our nation's schools educate children, half of all children will be below grade level. Grade equivalent scores are easily misunderstood—that is, most people think they reflect criterion-referenced scores, or mastery of academic subject matter by grade. Therefore, do not use grade equivalent scores to communicate student progress. The potential for misunderstanding outweighs the potential benefit of

understanding. Describe scores relative to a norm using percentiles, NCEs, standard scores, or stanines—avoid using grade equivalents, because they deceive your audience into thinking about curricular comparisons rather than norm comparisons. For this reason, parents or guardians whose children take the WKCE no longer receive grade equivalents. The final word: Just say "No" to grade equivalents!

Criterion-Referenced WSAS Scores

Criterion-referenced scores describe a student's performance relative to a given standard. The Knowledge and Concepts Examinations provide four types of criterion-referenced scores: percentages, Objective Performance Indexes, scale scores, and proficiency levels.

Percentages. A percentage is the proportion of items a student answered correctly out of the total number of items in the test. Percentages range from 0-100 percent, and are calculated by adding the number of items correct, divided by the total number of items, times 100. Percentages are not percentiles! A student may have 80 percent correct on a set of items. If the test is difficult, 80 percent could be a very good score and could result in the student being in the 99th percentile when compared to others who took the same test. If the test is easy, 80 percent correct could be a poor score, resulting in the student being in the 1st percentile when compared to other students who also took the test. Percentages are criterion-referenced, because they reflect performance against an absolute (0-100 percent), not normative, standard.

Objective Performance Index. These scores (called OPIs) estimate the percentage of items a student would get correct in a test in which all items measure the same academic objective or skill. That is, items measuring similar skills within the WKCE are grouped together to measure the academic objectives described in the previous section on TerraNova test content. If there are five items measuring a specific skill (e.g., measurement skills in mathematics), and the student answered four of the items correctly, the student's OPI would be near 80 percent (i.e., 4/5 x 100). The reason the OPI may not be exactly 80 percent is that different items are weighted more or less strongly in estimating the OPI, based on their item characteristics.

OPIs, like percentages, range from 0-100 percent. However, they are grouped into three categories on the WKCE. Each category captures a range of scores, and is associated with a symbol. These categories are:

- **Mastery** ◉ (75-100 percent) OPIs in this range suggest the student has mastered the skill.
- **Partial Mastery** ◒ (50-74 percent) OPIs in this range suggest the student has partially, but not completely and reliably, mastered the skill.
- **Non-mastery** ○ (0-49 percent) OPIs in this range suggest the student has not mastered the skill.

Because OPIs estimate student mastery of specific curricular skills, they are the most useful WKCE score for planning instruction. That is, you could review individual students' scores to identify specific academic strengths and weaknesses. Likewise, you might review class averages to determine those skills children have learned and those skills they have not yet mastered, to decide which skills you teach well and which need more instructional attention. It is important to look at two things when considering class-wide results: (a) the mean, or average, OPI, and (b) the percentage of children in the class who fall below the mastery level. For example, a class average might be 76 percent (indicating mastery), yet as many as half the students in the class may fall below mastery level on that skill.

Scale Scores. These scores are difficult to understand, yet they form the basis of all other scores—including state proficiency levels. Therefore, it is important to understand scale scores and how they can be used.

To illustrate the concept of scale scores, imagine a curriculum arranged in a line, with one end representing absolutely no knowledge and the other end representing complete mastery of the domain. Imagine that you put mileposts (like those found on interstate highways) along this line, starting with 0 at the end representing no knowledge, and 900 at the end representing mastery. If you had a test in which items were linked to these mile markers, you could estimate how "far" children had progressed in the curriculum from their responses to test items. In fact, this is essentially what the Knowledge and Concepts Examinations do to yield scale scores. They link specific items to points in the curriculum, and "place" the child along the continuum from 0 to 900.

Where are children when they enter school on this curricular "highway"? We estimate that most children begin kindergarten at roughly the 400-450 mile marker: they have learned nearly half of a curriculum by the time they begin school. Most students have acquired oral language, concepts of numeration, understanding of basic social units, classification skills, and the like before entering kindergarten. Thus, the lowest scale scores typically reported by WKCE will be in the high 400s; the highest scale score reported on the WKCE is 899. The examinations cannot mark progress for children at or below preschool levels; a child who gets all of the items wrong will still have an estimated scale score in the mid to upper 400s. Therefore, we cannot use WKCE to assess children who are working to master early developmental skills, such as toilet skills, feeding, or single word oral expression. They require an alternate assessment to demonstrate progress in their curriculum.

Scale scores have many advantages over other scores. First, they describe a child's progress in the curriculum regardless of the level of test. For example, a fourth grader whose WKCE scale score is 580 would be estimated as having the same level of skills as an eighth grader whose scale score is 580, despite their taking two different levels of the examination. Second, they can describe a child's absolute progress in curricula independent of the child's relative standing. For example, a student whose reading scale score from the fourth grade examination is 510 might be at the 30th percentile relative to other fourth graders. When the same student takes the eighth grade WKCE, the student's scale score might be 530—but the student's percentile relative to other eighth graders might be in the 10th percentile. The increase in scale scores shows that the student has made progress in the curriculum, but the drop in percentiles shows the student is not making progress as rapidly as the student's peers. Scale scores provide an absolute, not relative, metric for measuring progress.

The third advantage of scale scores is that they can be used to fix expectations for a given grade level independent of how well other students do on the test. For example, if you were to decide that a scale score of 550 represents what a typical fourth grader should master, you could fix 550 to be a grade-level expectation. It would be statistically possible to have every fourth grader in the nation be at or above this scale score level. Unlike grade equivalents (which rise or fall with the performance of the norm group so that 50 percent of children are always above or below grade level), scale scores allow educators to fix a standard for grade-level expectations relative to curricular mastery—not the norm group. This is analogous to definitions of physical fitness, in which you might define fitness as the ability to do 10 pull-ups, 50 sit-ups, and 20 push-ups (i.e., set criterion standards), even though the national averages for number of pull-ups (2) , sit-ups (20), and push-ups (7) might fall below your fitness standards. In fact, Wisconsin educators use scale scores to define grade-level expectations in the form of proficiency levels.

Proficiency Levels. Proficiency levels set grade level expectations for curricular mastery. That is, they define certain points in the curriculum (defined by scale score "mile markers") as goals for tests within a subject matter area (e.g., mathematics). How were proficiency levels set? Printed items, one per page, were put into a book, arranged in order from easiest (i.e., the lowest scale score) on the first page, to hardest (the highest scale score) on the last page. A group of Wisconsin educators (mainly teachers, with a few administrators, school board members, and parents) received three bookmarks. They were told to put a bookmark where they would draw the line between items that represented minimal performance levels for a given grade (fourth, eighth, or tenth). In other words, all items from the first page to the bookmark were at the minimal performance level. They placed the second bookmark where they thought the items increased from basic to proficient, and the third bookmark to separate proficient from advanced items. This procedure was reiterated several times, with opportunities for educators to discuss why they placed their bookmarks where they did. The scale scores corresponding to the placement of the bookmarks recommended by subject matter/grade level teams are the ones currently used to define proficiency levels in Wisconsin. The people who set the proficiency levels, and the activities they used to set them, are described in greater detail by a Wisconsin DPI publication, *Final Summary Report of the Proficiency Score Standards* (DPI, November, 1997), and on the DPI website http://www.dpi.state.wi.us/dpi/oea/.

The final scale scores used to define proficiency scores are summarized in Table 3.1.

Note that teachers identified these scale score levels on the basis of item content, or on what children must do to show they have acquired academic

Summary of Proficiency Categories

WISCONSIN KNOWLEDGE & CONCEPTS EXAMINATIONS

PROFICIENCY CATEGORY SUMMARIES

IN TERMS OF *TERRANOVA* SCALE SCORE

READING	Minimal Performance	Basic	Proficient	Advanced
Fourth Grade	~427 - 599	**600** - 624	**625** - 683	**684** - 797+
Eighth Grade	~498 - 654	**655** - 671	**672** - 717	**718** - 820+
Tenth Grade	~512 - 665	**666** - 693	**694** - 726	**727** - 838+

LANGUAGE ARTS*	Minimal Performance	Basic	Proficient	Advanced
Fourth Grade	~455 - 598	**599** - 630	**631** - *667*	**668** - 763+
Eighth Grade	~502 - 639	**640** - 668	**669** - *706*	**707** - 825+
Tenth Grade	~530 - 666	**667** - 692	**693** – *733*	**734** - 835+

MATHEMATICS	Minimal Performance	Basic	Proficient	Advanced
Fourth Grade	~385 - 580	**581** - 622	**623** - 658	**659** - 788+
Eighth Grade	~487 - 673	**674** - 717	**718** - 749	**750** - 850+
Tenth Grade	~513 - 715	**716** - 743	**744** - 781	**782** - 892+

SCIENCE	Minimal Performance	Basic	Proficient	Advanced
Fourth Grade	~400 - 586	**587** - 618	**619** - 670	**671** - 799+
Eighth Grade	~483 - 661	**662** - 691	**692** - *728*	**729** - 857+
Tenth Grade	~489 - 684	**685** - 717	**718** - *755*	**756** - 893+

SOCIAL STUDIES	Minimal Performance	Basic	Proficient	Advanced
Fourth Grade	~430 - 607	**608** - 626	**627** - 660	**661** - 763+
Eighth Grade	~515 - 648	**649** - 669	**670** - 701	**702** - 803+
Tenth Grade	~530 - 673	**674** - 691	**692** - 720	**721** - 821+

Note. The definition of these four proficiency categories is summarized in Figure 1.2 on page 5.
**Language Arts Cut Scores Revised November, 1998, Cut Scores approved October, 1997.*

SummaryCsRev.doc 12/3/1999

BEST COPY AVAILABLE

skills, not on a statistical basis. It is a rare teacher indeed who knows the WKCE well enough to identify academic content from a scale score alone! Therefore, you might want to better understand the practical meaning of proficiency levels. Here are some activities that can help you become more familiar with proficiency levels and what they mean for your students. These activities take time; you might want to ask your district's in-service/ professional development coordinator to set aside time and support them.

- Take the WKCE at all levels, or at least at the level nearest your grade. Imagine a student you know fairly well, and who represents about the middle range of skill in your classroom, as you take the test. Answer the items as you think that student might. Be sure also to complete a written essay (again, writing as your student might). You can get copies of the examinations given to students from your district assessment coordinator.

- Score your examination. Use the TerraNova Scoring Guide for the level(s) of test you took, and score your responses. Some responses are scored easily, whereas others require judgment. For example, you will have to determine the differences among a one-, two-, and three-point response on a short written answer, and in some cases you will score the same response twice (e.g., once for grammar/style and once for content/meaning). You also will have to score your essay using the six-point scoring framework, or rubric. Pick the example that most closely matches your essay, and write down the number. Then ask a colleague to score the essay. Do not tell the colleague how you scored it. Compare your essay scoring to your colleague's scoring of the same one. If the scores are identical, that is the final score for your essay; if the scores are within one point of each other, "split the difference" by simply adding the two scores and dividing by two. If the scores are more than one point apart, get another colleague to score the essay. Add the two closest scores and divide by two to get the final essay score. Your district assessment coordinator can supply you with all the scoring guides you need, or look at the DPI website for guides.

- Set your own proficiency levels. Get a copy of the *Final Summary Report of the Proficiency Score Standards* (DPI, November, 1997). Read the specific descriptions of proficiency in each academic domain for your grade level (pp. 10-23). Get a list of item difficulties from your district assessment coordinator, so you can rank the items from easiest to hardest (they are not in order on the test!). Using this ranking of items and the content/grade-specific descriptions of proficiency, decide what you think would make the differences between minimal performance, basic, proficient, and advanced levels of achievement for your grade/content area. Check your results against those of other colleagues; discuss your reasons and consider your decisions in light of your discussions. Although it is not possible to directly compare your bookmarks to those used by the DPI, our experience is that Wisconsin educators fall fairly close to the standards used by DPI. Also, teachers are just about as likely to pick standards higher than the current levels as they are to pick standards lower than the current levels.

These exercises will help you better understand the content of the WKCE and how that content is linked to proficiency standards. Knowledge of test content is a necessary, but not sufficient, condition for making informed and effective judgments about what, how, and when to teach material. Also, knowledge of the examinations is essential for deciding whether and how students with disabilities should participate in the WKCE.

Applying Your Knowledge of the Wisconsin Knowledge and Concepts Examinations

Now that you know about the WKCE based on the TerraNova, let's use your knowledge to interpret WKCE results. The WKCE reports results in many ways. This chapter will guide you in interpreting the following reports:

- Individual Profile Report,
- Group Proficiency Level Report,
- Evaluation Summary Report,
- School Record Sheet,
- Writing Frequency Distribution, and
- Objectives Performance Summary.

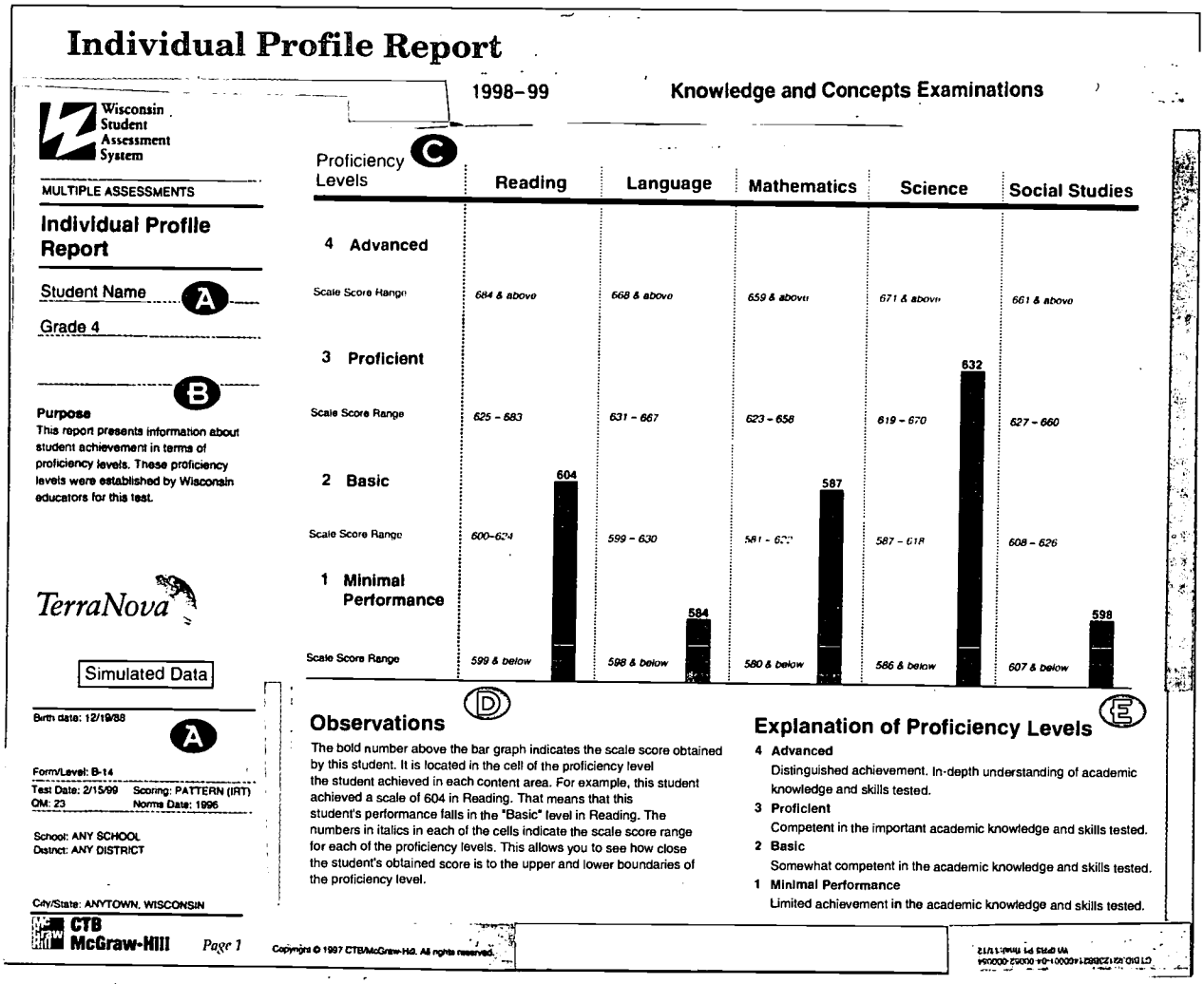
One other type of report (the Item Analysis Summary) is generated for each district. Because it is used primarily by district assessment specialists, and not by teachers, we will not describe it in this chapter.

Individual Profile Report

An example of a fourth grade student's Individual Profile Report appears in Figure 3.6. You will note that the report is two pages. On the first page, the student's proficiency level in five subject matter areas (Reading, Language, Mathematics, Science, and Social Studies) is presented in a graph form. For example, this student's achievement in Reading was at the Basic proficiency level, but the student's achievement in Language is at the Minimal Performance level.

The top section of the report's second page describes the student's results using stanines, scale scores, and national percentiles. Look at Figure 3.6 to see how the student's stanines and percentiles compare to others'. In all areas, the student is above the average for other children taking the test. Compare the student's scale scores to the fourth grade cutoffs in Table 3.1. You will see that the student's scale scores meet or exceed the lowest boundary of the Proficient range in Science (i.e., the scale score of 632 is between 619 and 670). The student's scale scores in Reading and Mathematics are in the Basic proficiency level, whereas the student's scores in Language and Social Studies fall below Basic (i.e., reflect Minimal Performance). Finally, note that the last column of the section reports a National Percentile Range for each of the student's scores. This range uses the estimated likelihood of error in the score (remember, no test is perfect!) to predict where the student's performance actually falls. For example, your best estimate for the student's percentile rank in Language is 53, but you know there is some error in the test, so you would be pretty confident that the student's "true" percentile would fall between the 43rd and 60th percentiles.

Figure 3.6



BEST COPY AVAILABLE

The bottom section of page 2 of the report tells the type of prompt (Informative, Narrative, Descriptive, or Persuasive) given the student. The holistic score of 4.5 tells you one rater scored the essay a 4, and the other scored it a 5, yielding a final score of 4.5 (i.e., (4+5)/2 = 4.5). The descriptions below the holistic score describe the essay quality.

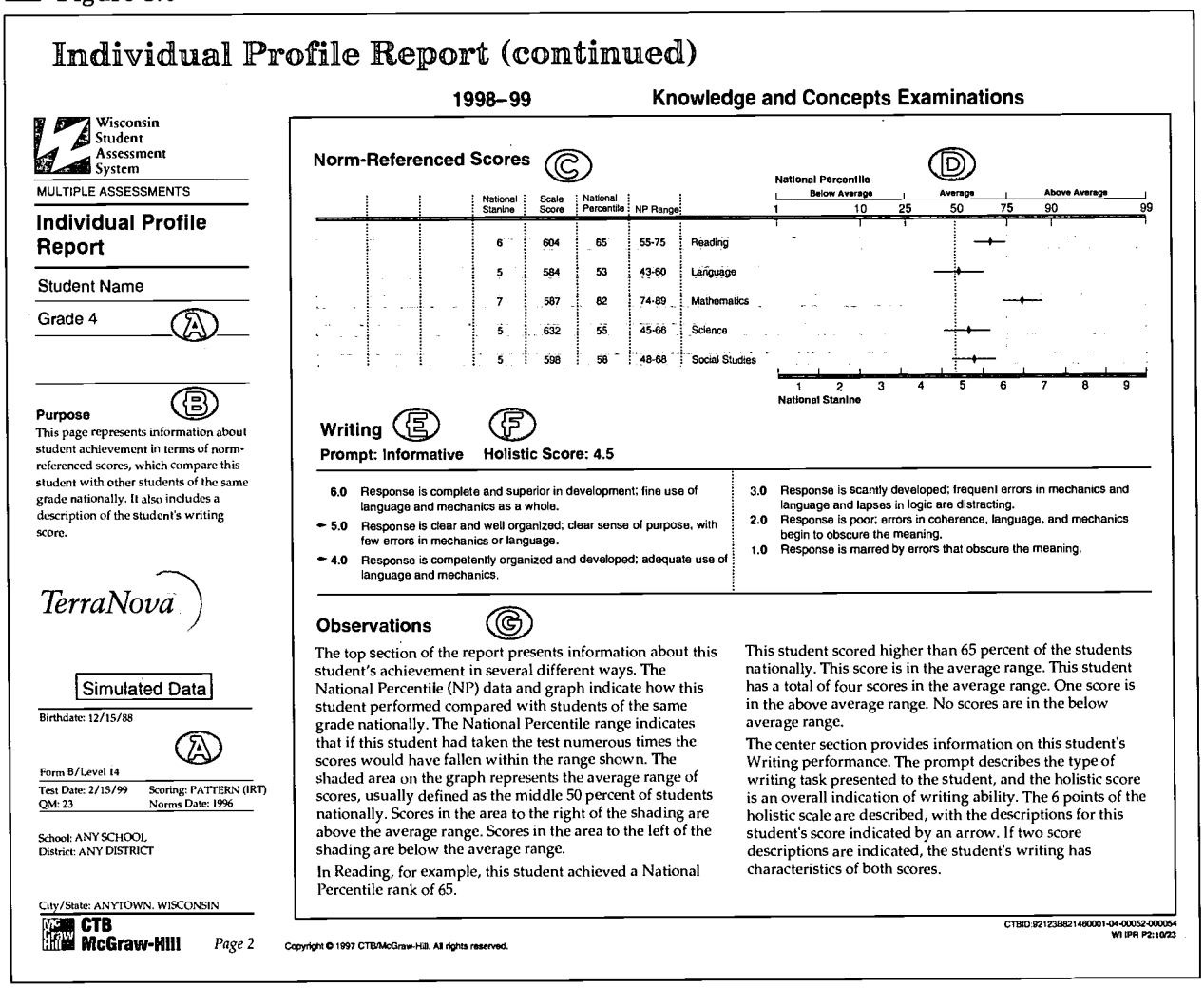
Group Proficiency Level Report

Figure 3.7 presents a Group Proficiency Level Report for a fourth grade class of 30 students. This report describes the proportion of students in each proficiency category for the class, school, district, and state (rows) by subject matter area (columns). Looking at the top row of the second column (Reading), you can see that 27 students (of 30) took the Reading test. Within the Reading domain, 16, or 53 percent, of the students' scores fell in the Minimal Performance range. This compares to 49 percent of scores for fourth graders at that school, 45 percent of fourth graders in the district, and 44 percent of fourth graders across the state. None of the students in this class scored at the Proficient or Advanced level on the Reading test. In contrast, 21 of 30 (70 percent) scored at the Proficient level in Science. For a discussion of how percentages are calculated in each category (see Playing the Percentages on page 37).

Evaluation Summary Report

The Evaluation Summary Report describes the achievement scores for a school or district at fourth, eighth, or tenth grade. Figure 3.8 contains an example of an Evaluation Summary Report for ANY SCHOOL's class of 89 (see lower left hand side of the report) eighth grade (see letter A) students. The top row of results tells the number of

Figure 3.6



Individual Profile Report (continued)

1998–99 Knowledge and Concepts Examinations

Wisconsin Student Assessment System

MULTIPLE ASSESSMENTS

Individual Profile Report

Student Name

Grade 4 (A)

(B)

Purpose

This page represents information about student achievement in terms of norm-referenced scores, which compare this student with other students of the same grade nationally. It also includes a description of the student's writing score.

TerraNova

Simulated Data

Birthdate: 12/15/88

(A)

Form B/Level 14

Test Date: 2/15/99 Scoring: PATTERN (IRT)

QM: 23 Norms Date: 1996

School: ANY SCHOOL

District: ANY DISTRICT

City/State: ANYTOWN, WISCONSIN

CTB McGraw-Hill Page 2

Norm-Referenced Scores (C)

	National Stanine	Scale Score	National Percentile	NP Range
Reading	6	604	65	55-75
Language	5	584	53	43-60
Mathematics	7	587	82	74-89
Science	5	632	55	45-66
Social Studies	5	598	58	48-68

(D)

National Percentile: Below Average | Average | Above Average

1 10 25 50 75 90 99

1 2 3 4 5 6 7 8 9

National Stanine

Writing (E) (F)

Prompt: Informative Holistic Score: 4.5

- 6.0 Response is complete and superior in development; fine use of language and mechanics as a whole.
- 5.0 Response is clear and well organized; clear sense of purpose, with few errors in mechanics or language.
- 4.0 Response is competently organized and developed; adequate use of language and mechanics.
- 3.0 Response is scantly developed; frequent errors in mechanics and language and lapses in logic are distracting.
- 2.0 Response is poor; errors in coherence, language, and mechanics begin to obscure the meaning.
- 1.0 Response is marred by errors that obscure the meaning.

Observations (G)

The top section of the report presents information about this student's achievement in several different ways. The National Percentile (NP) data and graph indicate how this student performed compared with students of the same grade nationally. The National Percentile range indicates that if this student had taken the test numerous times the scores would have fallen within the range shown. The shaded area on the graph represents the average range of scores, usually defined as the middle 50 percent of students nationally. Scores in the area to the right of the shading are above the average range. Scores in the area to the left of the shading are below the average range.

In Reading, for example, this student achieved a National Percentile rank of 65.

This student scored higher than 65 percent of the students nationally. This score is in the average range. This student has a total of four scores in the average range. One score is in the above average range. No scores are in the below average range.

The center section provides information on this student's Writing performance. The prompt describes the type of writing task presented to the student, and the holistic score is an overall indication of writing ability. The 6 points of the holistic scale are described, with the descriptions for this student's score indicated by an arrow. If two score descriptions are indicated, the student's writing has characteristics of both scores.

CTBID:92123882146000I-04-00052-000054
WI IPR P2:10/23

Copyright © 1997 CTB/McGraw-Hill. All rights reserved.

BEST COPY AVAILABLE

students whose scores are included in the summary report. Note that the number varies by subject matter, with only 86 students completing the Reading section, and 89 completing the Mathematics section.

The second major row (letter C) of results lists the arithmetic average, or mean, for many scores, and the average spread of scores about the mean (the standard deviation). Each line in this row is described below.

- The top line provides the mean, or average, NCE for the five subject matter areas. Examples: the mean Reading NCE for this class is 48.0; the mean Science NCE is 51.4. Remember: 50 is the national mean, so all of these scores are close to the national average.

- The second line reports the average spread (standard deviation) around the mean. Examples: the average spread of Reading NCEs around the mean is 13.9; the average spread of NCEs in Mathematics is bigger (19.3). Remember: a representative normal sample would be about 21; standard deviations less than 16 imply students are more alike than would be expected, and numbers greater than 26 suggest students are more diverse than expected.

- The third line reports the national percentile (NP) of the NCE mean. Examples: the mean Science NCE of 51.4 is equal to an NP of 53; the Social Studies NCE mean of 49.9 is equal to an NP of 50. Remember: the average NP is 50 (i.e., an NP of 50 divides the national sample in half, with half scoring lower and half scoring higher).

- The fourth line reports the mean scale score for the group. Examples: the mean scale score for Science (696.7) is lower than the mean scale score for Social Studies (700.5). Remember: scale scores

Figure 3.7

Proficiency Summary by Student Group

Wisconsin Student Assessment System

MULTIPLE ASSESSMENTS

Group Proficiency Level Report

School: ANY SCHOOL

Grade: 4 (A)

(B)

Purpose

This page summarizes the data by proficiency level and content area. Teachers and program administrators may compare all content areas for one level or all levels within one content area.

TerraNova

Simulated Data

Total enrollment: 30

Form/Level: B-14 (A)

Test Date: 2/15/99 Scoring: PATTERN (IRT)
QM: 23 Norms Date: 1998

School: ANY
District: ANY
State: WISCONSIN

City/State: ANYTOWN, WI

CTB McGraw-Hill Page 1

1998–99 Knowledge and Concepts Examinations

(C) Proficiency Levels	(D) Reading 27 Students	Language 27 Students	Mathematics 27 Students	Science 27 Students	Social Studies 27 Students
4 Advanced Distinguished achievement. In-depth understanding of academic knowledge and skills tested.	No. of students 0 % Class 0 % School 5 % District 3 % State 3	No. of students 0 % Class 0 % School 5 % District 3 % State 3	No. of students 0 % Class 0 % School 7 % District 4 % State 4	No. of students 0 % Class 0 % School 8 % District 5 % State 6	No. of students 0 % Class 0 % School 6 % District 4 % State 3
3 Proficient Competent in the important academic knowledge and skills tested.	No. of students 0 % Class 0 % School 4 % District 2 % State 3	No. of students 0 % Class 0 % School 1 % District 1 % State 1	No. of students 2 % Class 7 % School 2 % District 4 % State 4	No. of students 21 % Class 70 % School 30 % District 36 % State 34	No. of students 2 % Class 7 % School 5 % District 7 % State 7
2 Basic Somewhat competent in the academic knowledge and skills tested.	No. of students 11 % Class 37 % School 19 % District 28 % State 26	No. of students 3 % Class 10 % School 7 % District 7 % State 9	No. of students 25 % Class 83 % School 36 % District 43 % State 42	No. of students 6 % Class 20 % School 10 % District 12 % State 12	No. of students 21 % Class 70 % School 27 % District 30 % State 30
1 Minimal Performance Limited achievement in the academic knowledge and skills tested.	No. of students 16 % Class 53 % School 49 % District 45 % State 44	No. of students 24 % Class 80 % School 63 % District 66 % State 62	No. of students 0 % Class 0 % School 30 % District 27 % State 24	No. of students 0 % Class 0 % School 29 % District 25 % State 24	No. of students 4 % Class 13 % School 39 % District 37 % State 35

* Percentages are based on total enrollment, including students for whom no scores are reported, and may not total 100%

Copyright © 1997 CTB McGraw-Hill. All rights reserved.

BEST COPY AVAILABLE

are like yardsticks, so it is possible to compare scores across academic domains and different levels of the test. The fact that the Social Studies NP is lower than the Science NP (even though the scale score is higher) means the national sample finds social studies easier than science.

- The fifth line reports the average spread (standard deviation) of scores around the scale score mean. Examples: the spread of Language scale scores (33.2) is smaller than the spread of Mathematics scale scores (45.4).

The next major section or row of the Evaluation Summary (letter D) divides the group of scores into different sections. The sections are defined by the score that separates the top 10 percent from the rest of the class (i.e., the 90th Local Percentile, or LP); the score separating the top 25 percent (75th LP); the median for the class (50th LP), the score separating the bottom 25 percent (25th LP), and bottom 10 percent (10th LP). This information tells you how scores are spread out—or bunched up—within a class. Within each of these sections, there are three lines reporting results:

- The first line reports the National Percentile (NP) of the LP. Examples: the score defining the top 10 percent of the class (90th LP) is equal to an NP of 91.2 in Language, and 84.3 in Reading. The median class score (50th LP) in Science has an NP of 52.7, and the median Reading score is 41.8. Remember: in a class that exactly reflects the national average, the NP of the 50th LP (median) would be 50 (i.e., the score defining the top 50 percent would be the same for the class and the national average); in classes that score higher than the average, the NP will be over 50, and in classes below the national average, the median NP will be less than 50.
- The second line reports the NCE of the LP. Examples: the NCE of the bottom 10 percent of

Figure 3.8

Evaluation Summary Report

Wisconsin Student Assessment System

MULTIPLE ASSESSMENTS

Evaluation Summary Report

School: ANY SCHOOL (A)

Grade 8

Purpose (B)

This page gives administrators numeric information to evaluate the overall effectiveness of the educational program. This page displays a comprehensive numeric description of your students' achievement. This page is for those who prefer to analyze the data in tabular form.

TerraNova

Simulated Data

Total Enrollment: 89

Test Date: 2/15/99 Scoring: PATTERN (IRT)
QM: 23 Norms Date: 1998

District: ANY DISTRICT

City/State: WISCONSIN

CTB McGraw-Hill Page 1

Copyright © 1997 CTB/McGraw-Hill. All rights reserved.

1998–99 Knowledge and Concepts Examinations

	Reading	Language	Math	Science	Social Studies
Number of Students	86	87	89	86	88
(C) Mean Scores & Standard Deviations					
Mean Normal Curve Equiv.	48.0	52.0	49.5	51.4	49.9
Standard Deviation	13.9	14.9	19.3	15.7	18.2
NP of the Mean NCE	46	54	49	53	50
Mean Scale Scores	696.7	715.3	694.4	696.7	700.5
Standard Deviation	35.2	33.2	45.4	31.0	40.5
(D) Local Percentiles/Quartiles					
90th Local Percentile					
National Percentile	84.3	91.2	89.0	88.3	88.9
Normal Curve Equiv.	71.2	78.3	76.4	75.4	76.2
Scale Score	748.1	765.3	754.4	744.2	756.6
75th Local Percentile Q3					
National Percentile	62.8	75.4	72.3	70.2	74.3
Normal Curve Equiv.	56.9	64.4	62.7	61.0	63.9
Scale Score	719.8	741.5	725.3	719.7	731.3
50th Percentile (median) Q2					
National Percentile	41.8	53.3	54.0	52.7	56.7
Normal Curve Equiv.	45.9	52.0	52.0	50.2	53.3
Scale Score	695.3	719.3	704.0	699.7	711.3
25th Local Percentile Q1					
National Percentile	30.0	36.1	24.8	30.1	26.0
Normal Curve Equiv.	39.0	42.4	35.5	38.9	36.5
Scale Score	678.0	698.2	665.5	678.0	647.0
10th Local Percentile					
National Percentile	12.1	15.5	10.9	13.2	11.0
Normal Curve Equiv.	25.2	29.0	24.3	26.1	24.1
Scale Score	635.3	661.0	639.7	646.9	647.0
(E) National Quarters					
Local/Number 76–99	10	22	19	16	20
Per Quarter 51–75	24	25	26	30	27
26–50	38	26	21	25	20
01–25	14	14	23	15	21
Local/Percent 76–99	11.6	25.3	21.3	18.6	22.7
Per Quarter 51–75	27.9	28.7	29.2	34.9	30.7
26–50	44.2	29.9	23.6	29.1	22.7
01–25	16.3	16.1	25.8	17.4	23.9

BEST COPY AVAILABLE

the class in Mathematics is 24.3; the Science NCE for the top quarter (75th LP) is 61.0.

- The third line reports the scale score of the LP. Example: the 25th LP (bottom quarter of the class) is defined by a Language scale score of 698.2.

The bottom row or section of the report (letter E) tells you how many students had scores within the top, second, third, and bottom quarter relative to national averages.

- The first row of four lines tells the number of students in the class within each national quartile. Examples: 14 students scored in the bottom quartile on the Reading test, whereas 23 students scored in the bottom quartile of the Math test; on the Language test, 22 students scored in the top quartile and 25 scored in the second quartile. Remember: the number of students in any quartile is determined by how well they do on the test, and by the number who took the test.
- The second row of four lines tells the percentage or proportion of students in the class within each national quartile. Examples: 11.6 percent of the class placed in the top quartile in Reading; 25.3 percent of the class placed in the top quartile in Language. Remember: the proportion expected in each national quartile is 25 percent. If the proportion in the top two quartiles is greater than 50

Playing the Percentages

Where do the percentages in a Group Proficiency Level Report come from? The answer is not obvious. To answer the question, look closely at Figure 3.7.

First, note that three, or 10 percent, of fourth graders in this class did not take the WKCE. Reasons for not taking the test might include limited English proficiency, poor attendance, or exclusion due to disabilities. That means that 27, or 90 percent, of the fourth graders in this class took the exams, and three, or 10 percent, did not.

Second, the percentage of students in each category is based on the total number of students enrolled in the class. The total class enrollment comes from the number reported enrolled on the third Friday of the school year. In this example, the enrollment number was 30 students. To calculate the proportion of students in each proficiency level, the report takes the number who scored at that level (e.g., 21 scored at the Proficient level in Science) and divides by the total enrolled (30) to get the proportion of students in the class at the proficient level (70%=21/30).

You might argue that the results underestimate the percentage of children in this class who are in a given proficiency level. For example, you could say that 100 percent of the students who took the Science test scored at (6) or above (21) the Basic proficiency level (i.e., 6 + 21 = 27, or 100% of test-takers). You could say the same for Mathematics. However, the report indicates only 90 percent of the students scored at or above the Basic level in Science and Mathematics. Why isn't it 100 percent?

The answer lies in "playing the percentages." By reporting results as a proportion of students who are enrolled, rather than the proportion of students who took the test, the state is eliminating incentives for excluding students from the WKCE. If the state reported outcomes in terms of the proportion who took the test (rather than total enrollment), it would encourage districts to exclude the lowest-scoring students from the WKCE. For example, if you excluded the six students who scored at the Basic level in Science, plus the three students who did not take the exam, 21 students would score in the Proficient level. That would mean 21/21 students, or 100 percent of those taking the test, would be Proficient! However, only 21 (i.e., 70 percent of the class) actually earned scores at the Proficient or Advanced level. So the percentage of students in each proficiency category is determined by the number of students who earn scores in that category divided by the number of students enrolled in the grade—not by the number of students who took the test. The state reports the percentage at each proficiency level based on the total in the class, rather than the total who took the test, so that districts would not be inadvertently encouraged to exclude students who might score lower than others. Schools have nothing to lose—and perhaps something to gain—by including students in WKCE.

percent, the class is above the national average; if the numbers add to less than 50 percent, the class is below the national average. This is true no matter how many students take the test (25 percent is always expected in each quartile).

School Record Sheet

This document lists each student's scores in each academic domain. Figure 3.9 presents the first (page 1) page of scores from a group of eighth graders, and the last (page 2) page of scores from a group of fourth graders. Each row represents a different student (on page 1), and a final proficiency summary for all students (page 2).

- The first column of the report (letter C) identifies the students by name (omitted on this report), birth date, and the form of the test the students took (Form B, Level 18).
- The second column (letter D) lists the scores reported for each student. They are:

  NP (National Percentile; range 1-99)
  NS (National Stanine, range 1-9)
  NCE (Normal Curve Equivalent; range 1-99)
  SS (Scale Score; range 450-899)
  GE (Grade Equivalent; range preK-12.9+)
  PL (Proficiency Level; 1 = Minimal Performance, 2 = Basic, 3 = Proficient, 4 = Advanced).

- The next column presents each student's Reading score in six different ways (NP, NS, NCE, SS, GE, PL).
- The next four columns present each student's scores in Language, Math, Science, and Social Studies in six different ways (NP, NS, NCE, SS, GE, PL).
- The last column presents each student's holistic writing score (all students responded to the

Figure 3.9

School Record Sheet

Wisconsin Student Assessment System

MULTIPLE ASSESSMENTS

School Record Sheet (A)

School: ANY SCHOOL

Grade 8

Purpose (B)
This report provides a permanent record of test results for students in a class, or some other specified group, and summary data. The results may be used to evaluate individual and group achievement compared to the nation, determine overall performance, and identify areas of strength and need.

TerraNova

Simulated Data

Form/Level: B-18
Test Date: 2/15/99 Scoring: PATTERN (IRT)
QM: 23 Norms Date: 1996

District: ANY DISTRICT

City/State: ANYTOWN, WISCONSIN

CTB McGraw-Hill

Page 1

1998–99 Knowledge and Concepts Examinations

Students	(C) Scores	Reading	Language	Math	Science	Social Studies	Writing Prompt: Informative
STUDENT	NP	88	80	95	64	75	Holistic Score: 4.5
Birthdate: 01/16/85	NS	7	7	8	6	6	
Form: B Level:18	NCE	75	68	85	58	64	
	SS	716	700	748	696	697	
	GE	8.5	9.0	10.0	8.4	9.1	
	PL	3	4	4	2	3	
STUDENT	NP	86	64	96	94	99	Holistic Score: 4.0
Birthdate: 10/20/84	NS	7	6	9	8	9	
Form: B Level:18	NCE	73	57	88	83	99	
	SS	712	682	755	743	774	
	GE	8.6	8.3	10.5	11.1	11.9	
	PL	3	3	4	4	4	
STUDENT	NP	93	91	88	63	82	Holistic Score: 5.0
Birthdate: 07/14/85	NS	9	8	7	8	7	
Form: B Level:18	NCE	81	78	75	57	69	
	SS	728	717	731	695	705	
	GE	10.4	11.1	9.0	8.8	9.2	
	PL	4	4	3	3	3	
STUDENT	NP	59	66	29	43	42	Holistic Score: 3.0
Birthdate: / /	NS	5	6	4	5	5	
Form: B Level:18	NCE	55	58	38	46	48	
	SS	680	684	657	675	667	
	GE	8.1	8.1	7.0	7.4	7.3	
	PL	2	2	1	2	2	
STUDENT	NP	71	99	78	82	98	Holistic Score: 3.5
Birthdate: 05/28/85	NS	6	9	7	7	9	
Form: B Level:18	NCE	62	98	66	69	93	
	SS	692	766	715	718	750	
	GE	9.9	11.8	8.9	8.9	10.8	
	PL	2	4	2	3	4	
STUDENT	NP	83	86	92	80	88	Holistic Score: 5.5
Birthdate: 10/11/84	NS	7	7	8	7	7	
Form: B Level:18	NCE	70	73	80	68	75	
	SS	707	707	740	714	714	
	GE	9.9	8.0	12.1	8.9	9.4	
	PL	3	3	3	3	3	

Individual Scores
NP: National Percentile
NS: National Stanine
NCE: Normal Curve Equivalent
SS: Scale Score
GE: Grade Equivalent

Proficiency Levels
1= Minimal Performence
2= Basic
3= Proficient
4= Advanced

Copyright © 1997 CTB/McGraw-Hill. All rights reserved.

BEST COPY AVAILABLE

Informative prompt), which ranges from 1-6 (see Figure 3.6 on page 34, page 2 of the Individual Score Report, for descriptions of each score).

For example, let's examine the first row of scores.

- The first column tells us the scores to the right are for a student born on January 16, 1985, who took Form B Level 18 of the WKCE.
- The third column tells us the student's scores in Reading were:

  NP (National Percentile): 88
  NS (National Stanine): 7
  NCE (Normal Curve Equivalent): 75
  SS (Scale Score): 716
  GE (Grade Equivalent): 8.5
  PL (Proficiency Level): 3 (Proficient)

- The last column tells us the student's response to the Informative Writing Prompt earned a 4.5 (i.e., one rater scored it a 4, and the other scored it a 5). Let's look at a second example. Look at the fifth student's scores (Birthdate 05/28/85) in math. This student is above average relative to the national percentile (NP=78) and consequently has a grade equivalent higher than average (8.9). However, the student is not proficient in math (Proficiency=Basic). This shows the difference between grade equivalents, which are set to the norm group, and proficiency levels, which are set to curricular standards for mastery. It is possible to be above average—and still not be proficient.

Finally, Page 2 of the School Record Sheet reports summary data for a fourth grade class. The top section (letter A) presents the average scores for the class, and the bottom section (letter B) presents the number and proportion of students in each proficiency level.

The first row presents the averages for fourth grade students who took the WKCE. The scores

Figure 3.9

School Record Sheet (continued)

Wisconsin Student Assessment System

MULTIPLE ASSESSMENTS

School Record Sheet

School: Any School

Grade 4

Purpose
This report provides a permanent record of test results for students in a class, or some other specified group, and summary data. The results may be used to evaluate individual and group achievement compared to the nation, determine overall performance, and identify areas of strength and need.

TerraNova

Form/Level: 0-14
Test Date: 2/15/[illegible] Scoring: PATTERN (IRT)
QM: 23 Norms Date: 1996

District: Any District

City/State: ANYTOWN, WISCONSIN

CTB McGraw-Hill Page 2

1998–99 **Knowledge and Concepts Examinations**

Students	Scores	Reading	Language	Math	Science	Social Studies	Writing Prompt: Informative
School Summary (A)	NPNCE	59	52	63	70	61	-
	MNS	5.4	5.1	5.6	6.0	5.6	-
	MNCE	54.6	51.1	57.0	60.8	55.8	-
	MSS	647.8	640.8	636.9	652.2	645.8	-
	MHOLS						3.6
Number of Students		70	70	71	71	71	71
(B)							
Proficiency Level 1 Minimal Performance	PCT	8	3	3	0	1	
	NUM	6	2	2	0	1	
Proficiency Level 2 Basic	PCT	17	39	25	11	19	
	NUM	12	28	18	8	14	
Proficiency Level 3 Proficient	PCT	57	39	51	61	58	
	NUM	41	28	37	44	42	
Proficiency Level 4 Advanced	PCT	15	17	19	26	19	
	NUM	11	12	14	19	14	
Total Enrollment: 72 (See note below)							

Group (Summary) Scores

NPNCE: NP of the Mean NCE
MNS: Mean National Stanine
MNCE: Mean Normal Curve Equivalent
MSS: Mean Scale Score
MHOLS: Mean Holistic Score (Writing)

Note: Percentages are based on total enrollment, including students for whom no scores are reported, and may not total 100%.

PCT: Percent at Level
NUM: Number at Level

Copyright © 1997 CTB McGraw-Hill. All rights reserved.

BEST COPY AVAILABLE

reported for academic subject matter areas are:

- **NPNCE** (National Percentile of the average NCE; range 1-99)
- **MNS** (Mean National Stanine; range 1.0-9.0)
- **MNCE** (Mean Normal Curve Equivalent; range 1.0-99.0)
- **MSS** (Mean Scale Score; range 450.0-899.0)
- **MHOLS** (Mean Holistic Score; range 1.0-6.0)
- The last line is the number of students who took the test.

For example, the information in the Reading column tells you that the NP for the average NCE is 59, the average stanine is 5.4, the average NCE is 54.6, the average scale score is 647.8, and 70 students took the Reading WKCE. No average holistic score is reported for reading, because only the writing sample receives a holistic score. The average holistic score for the responses to the informative writing prompt was 3.6.

Continuing our example, the bottom part of the page reports the percentage (PCT) and number (NUM) of students in each proficiency level. Looking at Language, 3 percent (or 2 students) placed in the Minimal Performance level, whereas 19 percent (14 students) placed in the Advanced level on the Social Studies test. Note that the percentages are based on the number of students enrolled (72), not the number of students who took the WKCE (70 or 71). Consequently, none of the percentages adds to 100 percent (see Playing the Percentages on page 37).

Writing Frequency Distribution

Figure 3.10 presents the Writing Frequency Distribution report for a class of 12 eighth graders. The first column (letter D) lists the scores obtained by class members. The second column shows the number (Frequency) of students who

Figure 3.10

Writing Frequency Distribution

Wisconsin Student Assessment System

MULTIPLE ASSESSMENTS

Writing Frequency Distribution

School: ANY SCHOOL

Grade 8 (A)

Purpose (B)
This report shows summary information and the distribution of holistic scores for the writing prompt administered to the local group.

TerraNova

Simulated Data

(A)

Test Date: 2/15/99
QM: 23

District: ANY DISTRICT

City/State: ANYTOWN, WI

CTB McGraw-Hill

Page 1

1998–99 Knowledge and Concepts Examinations

Prompt: Informative (C)

(E)

Frequency Distribution

Holistic Score (D)	Frequency	Percent	Cumulative Frequency	Cumulative Percent
5.0	1	8	12	100
4.5	3	25	11	92
4.0	3	25	8	67
3.5	2	17	5	42
3.0	1	8	3	25
0	2	17	2	17

(F)

Condition Codes (Scored as 0)

Illegible	1
Primarily in Another Language	1

Summary Scores (G)

Number of Students	12
Mean Holistic Score	3.4
Standard Deviation	1.6

Copyright © 1997 CTB/McGraw-Hill. All rights reserved.

obtained each score. The third column converts the number to the percent of the class receiving each score. The fourth column converts the number to a cumulative frequency (i.e., the number of students in that category plus the number below that category) and the fifth column converts the cumulative total to the cumulative percent.

In the example in Figure 3.10, the second column tells you that 2 students received a holistic score of 0; 1 received a score of 3.0; 2 received scores of 3.5; 3 received scores of 4.0, 3 received scores of 4.5, and 1 received a score of 5.0. No students received scores of 1-2.5, and no student scored above 5.0 in this classroom.

The third column shows that a quarter of the class (3/12 = 25%) earned a score of 4.0, but only 8 percent of the class (1/12 = 8%) earned scores of 3.0 and 5.0. The fourth column shows that 8 children had scores of 4.0 or less. The last column shows that 25 percent of the class (3/12 = 25%) had scores of 3.0 or less, and 67 percent (8/12) had scores of 4.0 or less.

The rows in the Condition Codes section (letter F) explain why two students earned scores of 0 (remember, the lowest Holistic score is 1). One student's response was illegible, and one was written primarily in another language. None were off-topic or insufficient (i.e., too short) to evaluate.

The Summary scores at the bottom of the report (letter G) tell you that 12 students took the exam. The mean holistic rating for these 12 students is 3.4, and the average spread (standard deviation) of holistic ratings is 1.6.

Objectives Performance Summary

This section provides information about how well the class performs within specific academic skill objectives (i.e., academic objectives). Figure

Figure 3.11

Objectives Performance Summary

1998–99 Knowledge and Concepts Examinations

Wisconsin Student Assessment System

MULTIPLE ASSESSMENTS

Objectives Performance Summary

District: ANY DISTRICT

Grade 4 (A)

Purpose (B)

This report provides an analysis of group objectives mastery using the percent of students mastering each objective and average Objectives Performance Indexes (OPI) for this group. The information may be used to analyze curriculum strengths and areas of need.

TerraNova

WISCONSIN 99 TEST DATA END

No. of Students: 61 (A)

Form/Level: B-14

Test Date: 2/15/99 Scoring: PATTERN (IRT)

QM: 23 Norms Date: 1996

District: ANY DISTRICT

State: WISCONSIN STATE DEPT OF

City/State: ANYTOWN/ANYWHERE, WI

CTB McGraw-Hill Page 1

Objectives Performance by District/School (D)

Objectives Performance Index (OPI)
OPI is the estimated number of items correct out of 100

● Mastery (Range: 75 – 100 correct)
◒ Partial Mastery (Range: 50 – 74 correct)
○ Non-Mastery (Range: 0 – 49 correct)

** National Reference Group Grade 8.2

		Percent of Students Mastering Each Objective						Average Objectives Performance Index (OPI)					
					SCHOOL						SCHOOL		
		Total District	National **	Difference	OTTENWALTER	POLK	SMITH	Total School	National **	Difference	OTTENWALTER	POLK	SMITH
	Reading												
02	Basic Understanding (C)	43	45	–2	42	36	48	◒ 71	◒ 70	+01	◒ 69	◒ 72	◒ 72
03	Analyze Text	43	47	–4	38	36	52	◒ 57	◒ 67	–01	○ 45	◒ 65	◒ 63
04	Evaluate & Extend Meaning	28	30	–2	38	21	22	◒ 61	◒ 63	–02	◒ 60	◒ 64	◒ 59
05	Identify Reading Strategies	18	16	+2	29	7	15	○ 48	○ 44	+04	◒ 54	○ 42	○ 45
	Language												
07	Sentence Structure	62	47	+15	63	57	63	● 76	◒ 67	+09	● 78	● 75	● 75
08	Writing Strategies	57	39	+18	58	57	56	◒ 73	◒ 63	+10	● 75	◒ 71	◒ 71
09	Editing Skills	45	32	+13	50	36	44	◒ 68	◒ 61	+07	◒ 71	◒ 69	◒ 68
	Mathematics												
10	Number & Number Relations	56	40	+18	46	64	67	◒ 74	◒ 67	+07	◒ 71	● 77	● 76
11	Computation and Numerical Estimation	35	19	+18	21	29	52	◒ 62	◒ 53	+09	◒ 56	◒ 58	◒ 71
12	Operation Concepts	77	55	+18	79	71	78	● 80	◒ 72	+08	● 79	● 77	● 81
13	Measurement	51	23	+18	46	57	52	◒ 67	◒ 52	+15	◒ 70	◒ 62	◒ 67
14	Geometry & Spatial Sense	98	95	+3	100	100	96	● 94	● 95	–01	● 93	● 92	● 98
15	Data Analysis, Statistics, Probability	85	65	+20	92	71	85	● 80	● 75	+05	● 83	● 79	● 79
16	Patterns, Functions, Algebra	95	87	+8	96	65	96	● 92	● 89	+03	● 92	◒ 55	● 92
17	Problem Solving & Reasoning	85	62	+23	96	48	85	● 91	● 85	+06	● 82	○ 44	● 91
	Science												
10	Number & Number Relations	62	47	+15	63	57	63	◒ 71	◒ 70	+01	◒ 69	◒ 72	◒ 72
11	Computation and Numerical Estimation	57	39	+18	58	57	56	◒ 57	◒ 67	–01	○ 45	◒ 65	◒ 63
12	Operation Concepts	45	32	+13	50	36	44	◒ 61	◒ 63	–02	◒ 60	◒ 64	◒ 59
	Social Studies												
10	Number & Number Relations	58	40	+18	46	64	67	◒ 74	◒ 67	+07	◒ 71	● 77	● 78
11	Computation and Numerical Estimation	35	19	+18	21	29	52	◒ 62	◒ 53	+09	◒ 56	◒ 58	◒ 71
12	Operation Concepts	77	55	+18	79	71	78	● 80	◒ 72	+08	● 79	● 77	● 81
	Number of Students:	65			24	14	27	65			24	14	27

Copyright © 1997 CTB/McGraw-Hill. All rights reserved.

BEST COPY AVAILABLE

3.11 reports the outcomes for a district's fourth grade of 65 students taking the WKCE. The major row divisions present specific academic objectives in five subject domains (Reading, Language, Mathematics, Science, and Social Studies). There are two groups of columns. The left-most group of six columns presents information about the percentage of the class whose Objective Performance Index (OPI) score is greater than or equal to 75 percent (i.e., the percentage of students whom you might assume have "mastered" the objective). The right-most group of six columns describes the average, or mean, OPIs for the class by academic objective.

Within each of these divisions (left and right), there are six columns. Each of the columns presents information as follows:

- Total District (information for the entire district; note the first column on the right side of the page is mislabeled "Total School"; it should read "Total District")
- National (the national average)
- Difference (the difference between the district minus the national average—negative numbers imply the district is below the national average; positive numbers imply the district is above the national average)
- OTTENWALTER (name of first school in the district)*
- POLK (name of second school in the district)*
- SMITH (name of third school in the district).*

* The number of school columns may vary from 1 (repeats data for the entire district) to as many classrooms or schools as the district wants to report.

An illustrative example will help you understand what these numbers mean. First, look at the first row. It summarizes information for the grades of 65 students regarding four Reading objectives (Basic Understanding, Analyze Text, Evaluate and Extend Meaning, and Identify Reading Strategies). (Figure 3.2 describes all of these objectives.) Look at the first line in this column; it reports information about how well students did on the academic objective of Basic Understanding.

- The first column to the right reports the outcomes for all the fourth graders in the district (Total District). So, the first number (43) means that 43 percent of the fourth graders in this district earned an OPI of at least 75 percent. Another way of saying this is that you might guess 43 percent of the children in this grade have mastered Basic Understanding skills in reading.
- The second column (National) presents the proportion of students in the national sample who earned OPIs of 75 or greater. In this example, the number is 45, meaning in a typical classroom, you might expect 45 percent of the students to have mastered Basic Understanding.
- The third column (Difference) reports the difference between the Total District and National columns. In this case, the number (-2) means the percentage of fourth graders in this school who have mastered Basic Understanding skills in reading is slightly less (by 2 percent) than the proportion of the national sample who have mastered these skills. When a district performs better than the national average, the numbers in the Difference column will be positive; when the district performs worse, the numbers will be negative.
- The fourth through sixth columns report the percentage of students at each school in the district who have mastered each objective. So, the percentage of fourth graders at Ottenwalter who have mastered Basic Understanding is 42 percent, whereas only 36 percent of fourth graders at Polk have mastered the skill.

The columns on the right side of the report present the mean, or average, OPI for the class. Means are presented in two ways: visual symbols reflecting three levels of achievement (non-mastery ○, partial mastery ◒, and mastery ●), and the actual number of the mean. Look at the top line of results to see how students did for Basic Understanding in Reading.

- The first column in this section (Total School) reports the mean OPI for all fourth graders in the district. ◒ 71 means the average OPI was 71; because 71 is between 50-74, it falls in the partial mastery range (50-74), (see upper left corner of the report for a key).
- The second column in this section reports the mean OPI for the national sample of fourth graders. ◒ 70 means the average for the national sample was 70, which falls in the partial mastery range of 50-74, and so is illustrated with a half-filled circle ◒.
- The third column (Difference) reports the difference between the mean OPI for the Total School (really, district) and the mean OPI for the national sample. In this case, 71-70 = +01. In other

words, the average OPI for this school's fourth grade was higher (by +1 percent) than the national average OPI.

- The fourth through sixth columns in this section report the mean OPI for each fourth grade at each school in the district. The average OPI for Basic Understanding at Polk was higher (72) than the average for Ottenwalter (69).

Examination of the Objectives Performance Summary is probably the most useful activity for planning instruction. You might look down through the first (left-most) column to find objectives students have mastered (i.e., those you have successfully taught), and those that students have not mastered (i.e., those you have not successfully taught). For example, the high proportion of students mastering Mathematics objectives suggests these are strong areas of instruction. However, within this instructional domain, student mastery of Computation and Numerical Estimation is relatively low. By examining the proportion of a class that has mastered objectives, or by examining objectives with relatively high and low mean OPIs, you can identify areas of strength and areas in need of improvement within your instruction.

Note: the example in Figure 3.11 lists the same three objectives (10, 11, 12) for Mathematics, Science, and Social Studies. This is an error. The Science and Social Studies objectives are incorrectly identified.

Please keep in mind two important points when using OPIs to shape your teaching. First, *OPIs are not the same as proficiency levels.* OPIs are linked to specific academic objectives, not general academic proficiency. Their specificity can help you focus your teaching by suggesting relatively weak or strong areas of instruction within academic domains. However, proficiency levels reflect an aggregate performance within a broader domain.

Second, *always validate the results of standardized tests with your own assessments.* That is, check the results of tests against student work, quizzes, exams, and other evidence of student performance you collect in your classroom. Often, teachers do not teach, or test, the academic skills on which children do poorly. For example, the results in Figure 3.11 might suggest the teacher's approach to mathematics may overlook or fail to provide sufficient practice in computation and estimation. You may want to align instructional content with assessment. However, if you find that results of standardized tests conflict with classroom tests (e.g., students do well on your exams but not on the WKCE) examine *how* you ask students to perform versus how the standardized examinations ask students to perform. You may find it useful to align your assessment methods to those of the WKCE.

Commonly Asked Questions and Answers About the WKCE

This chapter is intended to enhance your assessment literacy for understanding and interpreting WKCE results. The first part of the chapter outlined why assessment literacy is important to teachers. The second part of the chapter described the WKCE content and results, and the third part of the chapter provided opportunities for you to apply your knowledge of the WKCE to interpreting results. However, you still may have some questions about the examinations. For example, we often have been asked some of the following questions.

1. How well does the content of the WKCE align with state academic standards?

The WKCE assesses about 55 percent of the Wisconsin academic standards, but about 98 percent of the examination content is included in the state standards. In other words, the examinations are essentially free from irrelevant academic skills and content, but they are incomplete. Some entire domains, such as oral communication, music, and physical education, are not included in the WKCE, and neither are some parts of some domains.

2. How does the WKCE measure what children learn in a classroom?

It depends on the degree of alignment between classroom instructional activities and exam content. If the classroom's curriculum and instruction are closely aligned to WKCE content, the exams will provide a good measure of student learning. However, if the classroom's curriculum and instruction are poorly aligned to exam content, the exam will not reflect student learning.

3. How is curriculum alignment different from "teaching to the test?" Isn't it wrong to "teach to the test?"

Aligning curriculum, instruction, and assessment is essential to effective education but teaching

BEST COPY AVAILABLE

to the test is cheating. How should you separate CIA alignment from teaching to the test? The answer is in the specificity of the teaching. If you teach to the WKCE's instructional objectives, and assess student progress by requesting similar kinds of responses, you are aligning curriculum, instruction, and assessment. If you teach the answers to a specific set of items or questions you think might be on the WKCE, you are "teaching to the test." Alignment promotes knowledge and skills students can use regardless of specific item content; teaching to the test promotes knowledge and skills that are useful only for a specific set of items.

4. What is the reading level of the WKCE?

Level 14, the fourth grade test, has a readability range from approximately second grade to fourth grade. Level 18, or the eighth grade test, has a readability range from fourth grade to eighth grade, and level 20, the tenth grade test, has a readability range of approximately seventh grade to tenth grade.

5. What does the publisher of the WKCE say about the use of testing accommodations with the test?

The publisher has not taken a position on the use of testing accommodations with TerraNova, the current examination, and consequently does not provide users with any guidance about appropriate or inappropriate testing accommodations.

6. The examination appears to measure knowledge, skills, and the application of these within subject matter areas, but does little to assess integration of skills across subject areas like mathematics and science or language arts and social studies. Why? This is inconsistent with our efforts to provide students integrated curriculum and instruction.

The WKCE has been designed to focus on knowledge, skills, and the application of these primarily within core subject matter areas of reading/language arts, mathematics, science, and social studies because that is how the state content and performance standards conceptualize important learning objectives. This approach maximizes the ability to isolate academic skills within subject matter, but it minimizes the understanding of integrated subject matter knowledge.

7. How long will the TerraNova test be used in the WSAS?

The current contract with CTB/McGraw-Hill, the publisher of TerraNova, runs through 2002.

8. Are there practice materials or recent past tests available so teachers and students can get a clear understanding of the type of questions asked on the test and the array of item formats or types?

Yes. First, the publisher of the TerraNova test publishes Practice Activities for students in grades 1 through 12. These are booklets with five or six practice items in each of the core subject matter areas. The items and the test directions are representative of those on recent versions of the test. Second, you can review a copy of last year's examination by contacting your school assessment coordinator. Copies of the forthcoming year's examination are secure until after the test is given and the test response forms are returned for scoring.

9. Is the WKCE available in other languages for students with limited English proficiency?

The TerraNova is available in Spanish as SUPERA. However, you may not give the Spanish version and substitute scores from it for the English version because the scores have not been demonstrated to be equivalent, and because the state standards call for proficiency in knowledge and skills in English.

References

CTB/McGraw-Hill. *Teacher's Guide to TerraNova*. Monterey, CA: CTB/McGraw-Hill, 1996.

Newmann, F.M.; H.M. Marks; and A. Gamoran. *Authentic Pedagogy and Student Performance*. Madison, WI: Wisconsin Center for Education Research, 1995.

Facilitating the Participation of All Students in Assessments

4

Wisconsin's public schools serve more than 879,000 students, all of whom are expected to learn and progress toward productive lives as citizens. Included in this population are more than 116,000 students with identified disabilities. All with special needs have individualized education program (IEPs) developed with input from parents and educational specialists. The majority of these students have relatively mild disabilities and, in most cases, learn much of the same content as their nondisabled peers, but possibly using different instructional methods or different developmental timelines.

Documenting students' achievements and educational progress is a critical aspect of an appropriate education and is required by law for students with disabilities. Consequently, educators are responsible for collecting evidence that students are learning. Assessment practices, especially testing, are one of the primary methods educators use to collect evidence of students' learning. Typically, when educators think of testing students with disabilities, they think about individualized, norm-referenced tests of cognitive abilities, achievement, and social and adaptive behavior, which are used to identify students who may be disabled and have special educational needs. Such tests often are helpful in identifying students with disabilities, but are of limited use as evidence concerning educational progress because they usually do not contain specific content that is aligned with students' daily instruction. In addition, such tests do not allow for progress comparisons to other students in the same schools.

In communities across the state and nation, many educational stakeholders want educators to be more accountable and to emphasize high standards for all students. Assessment programs have been and will continue to be part of the evidence used to document what students are learning and how well they are learning it.

All Means ALL

Historically, many of the statewide or schoolwide assessment efforts have not included all students. Participation rates for students with disabilities during the past several years in statewide assessments such as the Wisconsin Reading Comprehension Test (WRCT) at third grade and the Wisconsin Knowledge and Concepts Examinations (WKCE) at grades 4, 8, and 10, have ranged from a low of 41 percent to a high of 100 percent. Many of the students who did not participate were students with disabilities or with limited English proficiency. There are several possible reasons for these varying participation rates. However, if educators and other educational stakeholders who aspire to high standards for all students are to have a meaningful picture of how well students are learning and applying valued content knowledge and skills, all students need to be assessed periodically.

Before going further, let's look beyond the numbers at the cases of two students with disabilities.

The Case of Michele

Michele is a fourth-grade student who is classified as learning disabled. Her instructional reading level is second grade, but she receives all her instruction in regular classes with some support from a consulting special education teacher. She has good listening and memory skills, and is a highly motivated student who gets along well with her classmates. She often requires extra time to complete her work because she reads slowly and due to her poor spelling skills benefits from assistance with her written assignments.

The Case of Ben

Ben is chronologically an eighth grader who was diagnosed as autistic at three years of age. Due to his pervasive communication difficulties he receives much of his education in a highly structured special education classroom with six other students with developmental disorders. He has a limited vocabulary and interacts with only his teacher and her aide. He does, however, participate in a sixth-grade math class that focuses on basic skills and is doing quite well with the support from the aide.

Historically, state and district testing programs have excluded students like Michele and Ben. The reasons typically given for excluding students like these from testing programs include

- the concern that students with disabilities will lower a school's mean score,
- the desire to "protect" students with disabilities from another frustrating testing experience,
- the perception that the tests are not relevant, especially for students with disabilities,
- the fact that some parents do not want their child spending valuable class time taking a test that doesn't count toward a grade, and
- the belief that the guidelines for administering standardized tests prohibit, or at least greatly limit, what can be changed without jeopardizing the validity of the resulting test score.

The limited participation of students with disabilities in state and district assessments results in

- unrepresentative mean scores and norm distributions,
- a belief that students with disabilities cannot do challenging work, and
- the undermining of inclusion efforts for many students.

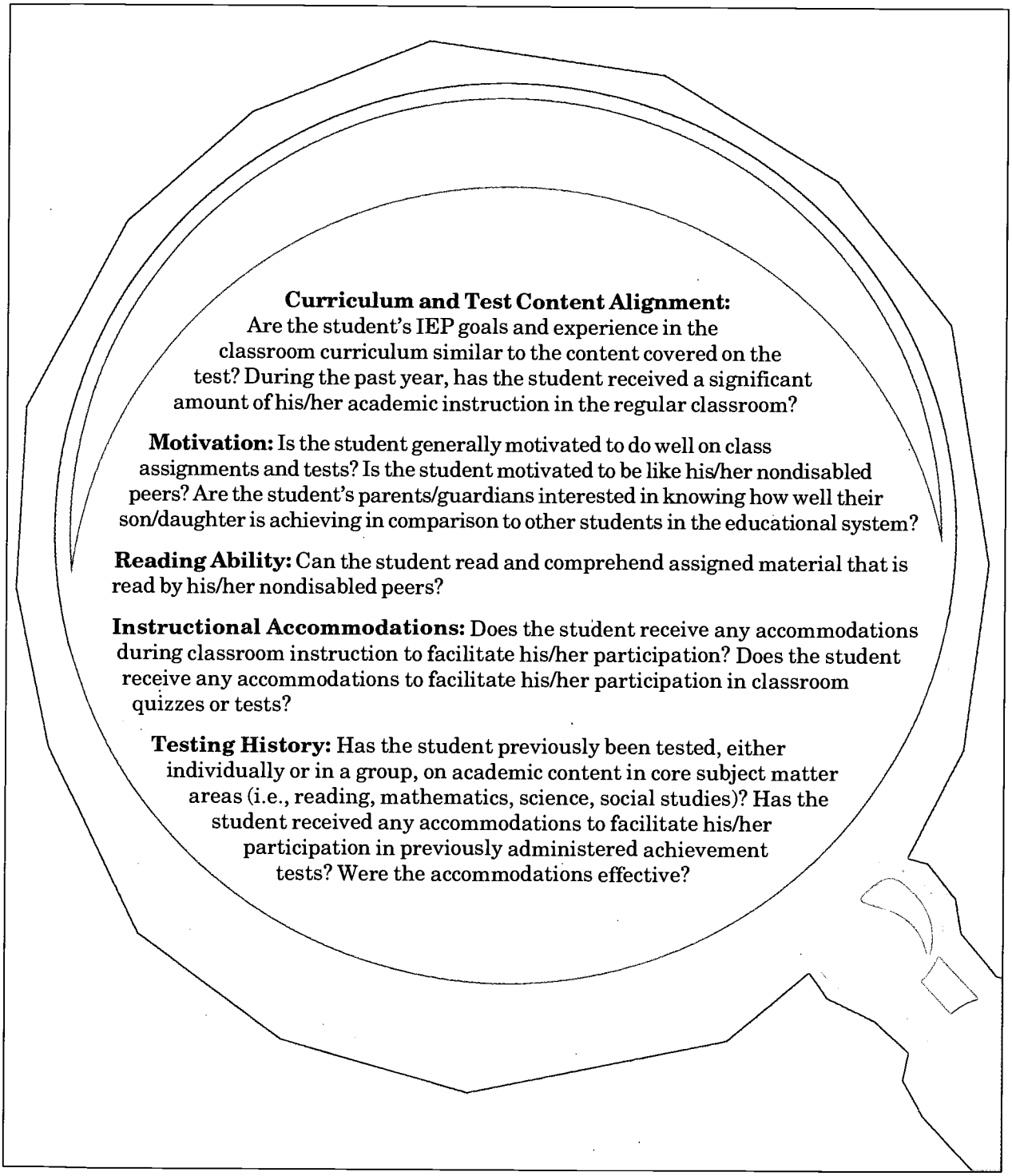
Since the passage of federal and state legislation in the 1970s, students with disabilities have been guaranteed access to a free appropriate public education. Therefore, when tests and assessment systems are designed to serve as indicators of progress in the subject matter content of a school's curriculum or the state's academic standards, and are used to make decisions about future educational services, all students must participate in the assessments as part of their free appropriate public education. Numerous court cases under the Americans with Disabilities Act of 1990 established the legal basis for this position and, most recently with regard to children with disabilities, in the amendments to the Individuals with Disabilities Education Act or IDEA (Public Law No. 105-17). The 1997 amendments to IDEA include requirements concerning

- the participation of children with disabilities in general state and district assessment programs, with appropriate accommodations when necessary,
- documenting in a student's IEP any individual modifications in the administration of state or district tests that measure achievement,
- documenting in a student's IEP a justification for exclusion from a standardized test and indication of how the student will be assessed with an alternate method, and
- reports to the public about the participation and performance of children with disabilities with the same details as reports for nondisabled children.

Decisions about including students with disabilities in assessment programs and validly implementing assessments can be challenging and require teachers' involvement on IEP teams. One of the first challenges confronting educators is to determine the "right" assessment program for students with disabilities. Practically speaking, students with disabilities could participate in: (a) the regular assessment, e.g., WRCT or WKCE, without accommodations; (b) the regular assessment with testing accommodations; (c) an alternate assessment; or (d) part of the regular assessment with testing accommodations and the remainder in an alternate assessment. In making this participation decision, educators consider an array of factors, many of which are "magnified" in Figure 4.1. As highlighted in this figure, the most critical factors include: the alignment between a student's IEP goals, classroom curriculum, and the content of the test; a student's reading ability; and the nature of instructional accommodations a student typically receives.

Figure 4.1

"Magnifying" Key Variables Discussed by IEP Teams When Making Participation and Accommodation Decisions



BEST COPY AVAILABLE

Tactics for Increasing the Meaningful Participation of ALL Students in Assessment Programs

As noted in IDEA '97 and in our state's Guidelines to Facilitate the Participation of Students with Special Needs in State Assessments (see Appendix D), testing accommodations and alternate assessment are two possible methods educators can use to facilitate the participation of all students with disabilities in assessments and accountability systems. Therefore, every teacher who works with students with disabilities should know about testing accommodations and alternate assessment if they want to facilitate their students' meaningful involvement in assessment programs.

Testing Accommodations

One of the most frequent steps for increasing the meaningful participation of students with disabilities in assessments is allowing changes to test procedures, rather than allowing changes in the test content. Such changes are commonly referred to as testing accommodations. Teachers are familiar with instructional accommodations like extra time to complete work or a quiet location to minimize distractions.

Testing accommodations are changes in the way a test is administered or responded to by a student. Testing accommodations are intended to offset distortions in test scores caused by a disability without invalidating or changing what the test measures (McDonnell, McLaughlin, and Morison, 1997). Many different testing accommodations are allowable as long as they do not reduce the validity of the test scores. In Wisconsin, the IEP team is entrusted to determine the appropriate testing accommodations for individual students with disabilities.

Educators can alter tests and assessment programs in a variety of ways to facilitate the participation of students with disabilities and still provide valid results. As increasing numbers of students with disabilities are included in assessment programs and take the same tests as their nondisabled peers, it is likely that teachers and other members of IEP teams will need to consider the use of testing accommodations. It is important to understand that accommodations are intended to maintain and facilitate the measurement goals of an assessment, not to modify the actual questions or content of the tests. Accommodations usually involve changes to the testing environment, e.g., Braille or large print materials, the amount of time a student has to respond, the quietness of the testing room, assistance in reading instructions, or the method by which a student responds to questions, e.g., orally with a scribe, pointing to correct answers. Testing accommodations should not involve changes in the content of test items. When changes are made to the test content, the test is very likely to be measuring skills or levels which are different from those skills or levels measured by the original test. If educators do make such changes to test content, the results from this "changed" test could not be compared validly with results from the "unchanged" test.

Accommodations generally result in some minor changes in the procedures for administration or response upon which a test was standardized. Consequently, because many educators have been taught to follow standardization procedures exactly, they may be reluctant to use accommodations. The keys to the selection and appropriate use of testing accommodations are threefold. First, educators must determine accommodations on a case-by-case basis for each student. Second, knowledge of student's current instructional accommodations should guide considerations of testing accommodations. Third, accommodations should make the test a more accurate measure of what a student knows or can do. That is, IEP teams must select accommodations that are likely to facilitate a student's participation in a testing program, but not likely to change or invalidate the intended meaning of a test score.

To date, there is no comprehensive research base to guide educators' decisions about which accommodations invalidate test results and which accommodations improve test performance without invalidating test results. Studies of the effects of testing accommodations on test scores of students with disabilities have been published and numerous investigations are underway in research centers in Wisconsin and across the country. However, research on testing accommodations is unlikely to be prescriptive because decisions about accommodations and their effect on a student's test performance are highly individualized events. Given that most researchers have used group research designs and compared the effects of accommodations across groups—e.g., one

group of students receives an accommodation and a second group doesn't—it is difficult, if not unreliable, to apply many generalizations from the existing published research to your current and future students and testing situations. However, if you have a clear understanding of what a test or subtest measures, then many of the decisions about appropriate or valid accommodations become rather straightforward. For example, reading aloud questions and answers on a reading test designed to measure sight vocabulary and comprehension certainly would invalidate the resulting score because these accommodations are changing the skills or competencies the test is designed to measure. Conversely, reading aloud a complex story problem on a test designed to measure mathematics reasoning and calculation could be appropriate for some students with disabilities. In this latter case, assistance with reading is designed to increase the likelihood that the test score is a better indicator of what the student has learned in mathematics. If the accommodation does this, then the test score is said to be valid.

Researchers group commonly-used accommodations into four categories:

- accommodations in timing;
- accommodations to the assessment environment;
- accommodations in the presentation format; and
- accommodations in the recording or response format.

Figure 4.2

Examples of Accommodations Frequently Considered Appropriate for Students with Disabilities

Time Accommodations

Administer a test in shorter sessions with more breaks or rest periods
Space testing sessions over several days
Administer a test at a time most beneficial to a student
Allow a student more time to complete the test

Setting Accommodations

Administer the test in a small group or individual session
Allow a student to work in a study carrel
Place student in a room or part of a room where he or she is most comfortable
Allow a special education teacher or aide to administer the test

Format Accommodations

Use an enlarger to facilitate vision of material
Use a Braille transcription of a test
Give practice tests or examples before actual test is administered
Assist a student in tracking test items by pointing or placing the student's finger on items
Allow use of equipment or technology that a student uses for other school work

Recording Accommodations

Use an adult to record a student's response
Use a computer board, communication board, or tape recorder to record responses

Figure 4.2 provides some specific examples of each of these categories of accommodations.

It is important to note that not all students with disabilities will need testing accommodations to participate and provide a valid or accurate account of their abilities. On the other hand, for a small number of students with more severe disabilities, testing accommodations will not be appropriate or reasonable. These students' educational goals and daily learning experiences concern content which may differ significantly from that contained in state or district content standards. Although many of the IEP goals of these students should be aligned with the state's academic content standards, a student's current performance may differ significantly from the performance standards expected for a given student's grade level. Consequently, students in this situation will need to participate in an alternate assessment to meaningfully measure their abilities and provide valid results.

Many educators find it difficult to make decisions concerning the selection and use of testing accommodations with students. They also find it difficult to explain the use of testing accommodations to other educational stakeholders. As a result of numerous discussions about testing accommodations with teachers, parents, and testing experts, let us suggest two useful metaphors for thinking about the role and function of testing accommodations.

The first metaphor concerns eyeglasses. Look around any room with other adults present and you will see at least one-third and maybe one-half of them wear eyeglasses to correct for vision impairments. Eyeglasses are an accommodation for imperfect or poor vision. If you wanted to test the natural vision ability of a person who wears glasses for driving and outdoor activities, then wearing glasses during a test of distant vision would invalidate the test score, assuming your purpose is to make an inference about the person's natural or uncorrected vision. On the other hand, if your purpose was to determine the same person's driving ability, then allowing glasses during the driving test would be a valid accommodation because it would facilitate a more accurate assessment of the person's driving skills by minimizing or eliminating problems due to vision impairments. Remember, even in the absence of disabilities or other complicating factors, tests are imperfect measures of the constructs they are intended to assess.

Using the same metaphor of a corrective lens, envision a student's "true" competence in reading, for example, as a point on a vertical scale. Next to it, imagine an identical scale of that student's "observed" competence, as reflected by performance on an assessment. Between the two scales is a lens causing some diffraction of light, so that true competence is represented (over repeated measurements) by an array of points on the observed-competence scale that forms a blurry image of the true, unmeasured competence. If the test is well designed, this image will be centered on the "true" value, i.e., it will be unbiased, and it will not be too blurry. In other words, it will be reliable. In summary, testing accommodations are intended to function like a corrective lens that will focus the distorted array of observed scores on a more valid image of the performance of individuals with disabilities.

The second metaphor about testing accommodations is an access ramp. One may conceptualize an access ramp as part of a package of testing accommodations for individuals with physical impairments which influence mobility. If individuals can't get to the testing room, then they certainly can't demonstrate what they know or can do! The conceptual value of an access ramp has additional meaning, however, when addressing issues of construct validity. Testing accommodations facilitate access to a test for students with a wide range of disabilities just like a ramp facilitates access to a building for individuals with disabilities related to mobility. The tests that students are required to take are designed to measure some specific target cognitive skills or abilities, such as mathematical reasoning and computations, but almost always assume that students have the skills to access the test, such as attending to instructions, reading story problems, and writing responses. Thus, knowledge and concepts tests like those included in the WKCE target broad constructs like mathematics, science, social studies, and language arts and are used to determine how students are doing in these subjects. Some students, in particular those with disabilities, have difficulty with the access skills needed to get "into" the test (see Box 4.1 for Target Skills vs. Access Skills Activity). Thus, valid testing accommodations, just like an access ramp, should be designed to reduce problems of access to a test and enable students to demonstrate what they know and can do with regard to the skills or abilities the test targets.

By now, you should have a good understanding of what testing accommodations are and how

Target Skills vs. Access Skills Activity

Test items are designed to measure specific or general skills or abilities. For example, many mathematics items are intended to measure a student's ability to reason, compute, and communicate a solution or result. The skills or abilities that test developers intend the items to measure can be called target skills or abilities. The same mathematics items require a student to attend, read, remember some information, and ultimately respond by bubbling in an answer choice or writing an extended response. These latter skills are generally not what the test developers designed the mathematics items to measure, but without these skills or abilities students cannot access or interact with the test items to demonstrate whether or not they possess the target skills measured by the items. Thus, skills or abilities such as attending, seeing, writing, etc. are considered access skills or abilities. A list of common access skills is provided below. Can you think of additional access skills?

1. Attending
2. Listening
3. Reading*
4. Remembering
5. Writing*
6. Following directions
7. Working by oneself
8. Sitting quietly
9. Turning pages of test booklet
10. Locating test items
11. Locating answer spaces
12. Erasing completely
13. Seeing
14. Processing information in a timely manner
15. Working for a sustained period of time
16. Spelling*

* Some skills such as reading, writing, and spelling are access skills for tests designed to measure mathematics, science, and social studies, but are target skills on most tests designed to measure reading/language arts skills.

Key Premise

Testing accommodations should be designed to only effect deficits in access skills, not target skills. If an accommodation involves one or more of the target skills or abilities a test is designed to measure, it will invalidate the test score.

Box 4.1

Target Skills vs. Access Skills Activity (continued)

Background Information about Subskills Measured on TerraNova

CTB/McGraw-Hill in developing language arts and mathematics for tests like TerraNova uses the following describers to characterize the many subskills their items are designed to measure.

Reading/Language Arts Objectives and Subskills

01 Oral Comprehension
Subskills: literal; interpretive

02 Basic Understanding
Subskills: sentence meaning; vocabulary; stated information; sequence, initial understanding; stated information graphics

03 Analyze Text
Subskills: main idea/theme; supporting evidence; conclusions, cause/effect; compare/contrast; story elements—plot/climax/character/setting, literary techniques; persuasive techniques; nonfiction elements

04 Evaluate and Extend Meaning
Subskills: generalize; fact/opinion; author-purpose/point of view/tone/bias; predict/hypothesize; extend/apply meaning; critical assessment

05 Identify Reading Strategies
Subskills: make connections; apply genre criteria; utilize structure, vocabulary strategies; self-monitor; summarize; synthesize across texts; graphic strategies; formulate questions

06 Introduction to Print
Subskills: environmental print; word analysis; sound/visual recognition

07 Sentence Structure
Subskills: subject/predicate; statement to question; complete/fragment/run-on; sentence combining; nonparallel structure; misplaced modifier; mixed structure problems; sentence structure

08 Writing Strategies
Subskills: topic sentence; sequence; relevance; supporting sentences; connective/transitional words; topic selection; information sources; organize information; writing strategies

09 Editing Skills
Subskills: usage; punctuation; capitalization; proofreading

Mathematics Objectives and Subskills

10 Number and Number Relations
Subskills: counting; read, recognize numbers; compare, order; ordinal numbers; money; fractional part; place value; equivalent forms; ratio, proportion; percent; roots, radicals; absolute value; expanded notation; exponents, scientific notation; number line; identify use in real world; rounding, estimation; number sense; number systems; number properties; factors, multiples, divisibility; odd, even numbers; prime, composite numbers

11 Computation and Numerical Estimation
Subskills: computation; computation in context; estimation; computation with money, recognize when to estimate; determine reasonableness; estimation with money

12 Operation Concepts
Subskills: model problem situation; operation sense; order of operations; permutations, combinations; operation properties

13 Measurement
Subskills: appropriate tool; appropriate unit; nonstandard units; estimate; accuracy, precision; time; calendar; temperature; length, distance; perimeter; area; mass, weight; volume, capacity; circumference; angle measure; rate; scale drawing, map, model; convert measurement units; indirect measurement; use ruler

14 Geometry and Spatial Sense
Subskills: plane figure; solid figure; angles; triangles; parts of circle; point, ray, line, plane; coordinate geometry; parallel, perpendicular; congruence, similarity; Pythagorean theorem; symmetry; transformations; visualization, spatial reasoning; combine/subdivide shapes; use geometric models to solve problems; apply geometric properties; geometric formulas; geometric proofs; use manipulatives; geometric constructions

15 Data Analysis, Statistics and Probability
Subskills: read pictograph, read bar graph; read line graph; read circle graph, read table, chart, diagram; interpret data display; restructure data display; complete/construct data display; select data display; make inferences from data; draw conclusions from data; evaluate conclusions drawn from data; sampling; statistics; probability; use data to solve problems; compare data; describe, evaluate data

16 Patterns, Functions, Algebra
Subskills: missing element; number pattern; geometric pattern; function; variable; expression; equation; inequality; solve linear equation; graph linear equation; solve quadratic equation; graph quadratic equation; model problem situation; system of equations; use algebra to solve problems

17 Problem Solving and Reasoning
Subskills: identify missing/extra information; model problem situation, solution; formulate problem; develop, explain strategy; solve nonroutine problem; evaluate solution; generalize solution; deductive/inductive reasoning; spatial reasoning; proportional reasoning; evaluate conjectures

18 Communication
Subskills: model math situations; relate models to ideas; make conjectures; evaluate ideas; math notation; explain thinking, explain solution process

Target Skills vs. Access Skills Activity (continued)

Application Activity

Below are several items like those used on tests such as *TerraNova*. Read through each item with the purpose of identifying the *target skills* (the skills the test developers intended to measure) and key *access skills* (skills needed to "get into" the item and to document a response)

Item #1

Target Skills: _____

_____

_____

Access Skills: _____

_____

_____

Choose the sentence that best combines the underlined sentences into one.

The train sped through the tunnel.

The train sped across the bridge.

✓ **A** The train sped through the tunnel and across the bridge.
B The train sped through and across the tunnel and the bridge.
C The train that sped through the tunnel sped across the bridge.
D The train sped through the tunnel and it sped across the bridge.

Item #2

Target Skills: _____

_____

_____

Access Skills: _____

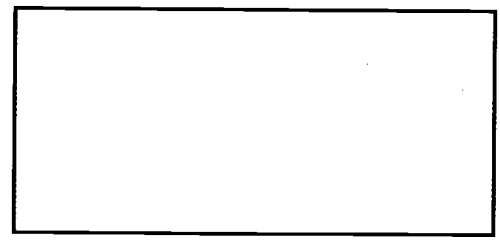
_____

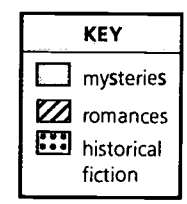
_____

This chart shows the number of different types of fiction books on a bookstore shelf.

mysteries	10
romances	30
historical fiction	30

The bookstore owner put 10 more mysteries on the shelf. Draw a circle graph that shows the fraction of the total number of books for each type of fiction that are now on the shelf. Use the key to label your graph.





Please note that all the test items used in this box are examples from *Teacher's Guide to TerraNova* (McGraw-Hill, 1997) and copied with permission.

BEST COPY AVAILABLE

Box 4.1

Target Skills vs. Access Skills Activity (continued)

Item #3

Target Skills: _____

_____

_____

Access Skills: _____

_____

_____

Choose the topic sentence that best fits the paragraph.

_____. *Some of the rain runs off into brooks and streams. Some of it goes into the roots of plants and trees. Some of it even goes back up into the air!*

- ○ All living things need water.
- ○ Rain is often collected in tanks.
- ✓ ○ The rain that falls from the sky is not lost or wasted.
- ○ Plants that live in the desert have special ways of storing water.

Item #4

Target Skills: _____

_____

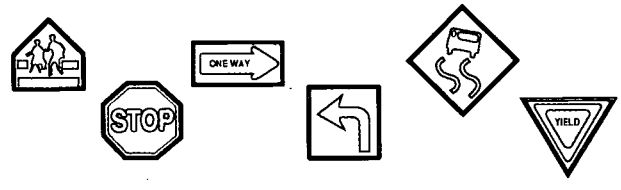
_____

Access Skills: _____

_____

_____

Look at the shapes of the road signs.



Sort the shapes into two groups by drawing them on the notepads below. Then explain why the shapes in each group go together.

Group 1

Why do these shapes make a group?

_____

_____

_____

Group 2

Why do these shapes make a group?

_____

_____

_____

Please note that all the test items used in this box are examples from *Teacher's Guide to TerraNova* (McGraw-Hill, 1997) and copied with permission.

they should function to improve the validity of a student's test score. In addition, you should be aware that testing accommodations are sanctioned by federal and state policies, and that IEP team members are responsible for selecting and implementing them for qualified students. But you may legitimately ask, "How do you go about selecting specific testing accommodations for specific students with specific disabilities and well-defined instructional plans?" The key to selecting and implementing testing accommodations for an individual student lies in the classroom(s) where that student is taught each day. That is, the instructional accommodations teachers frequently use to facilitate a student's teaching-learning interactions are prime candidates as accommodations when that same student is participating in a statewide or districtwide test. This premise is reasonable, particularly when there is good alignment between what is taught in the classroom and what is on the test. This does not mean, however, that all accommodations used to support a student during instruction will result in valid testing accommodations. We will say more about selecting and implementing testing accommodations later in this chapter via two student case illustrations.

In summary, think about the list of do's and don't's in testing accommodations offered by Thurlow, Elliott, and Ysseldyke (1998, pp. 61-62):

- Don't introduce a new accommodation for the first time for an assessment.
- Don't base the decision about accommodations for a student on the student's disability category.
- Don't start from the district or state list of approved accommodations when considering a student's accommodations in an upcoming test.
- Do systematically use accommodations during instruction and carry these into the assessment process.
- Do base the decision about accommodations, both for instruction and for assessment, on the needs of the student.
- Do consult the district or state list of approved accommodations after determining what accommodations the student needs. Then, reevaluate the importance of the accommodations that are not allowed. If they are important for the student, request their approval from the district or state.

As you work with students to provide testing accommodations, revisit this list and try to add to it. Now it is time to examine another assessment tactic, alternate assessments, designed to facilitate the participation of students with some of the most severe disabilities.

Alternate Assessments: The Ultimate "Accommodation"

For many students with severe disabilities, educators need to make changes beyond test administration procedures or format to ensure meaningful assessment results. Thus, the content of the assessment also must change to provide for a valid measure of what these students are learning. This approach has lead to the development of alternate assessments for approximately 15-20 percent of students with disabilities who are functioning at developmental and instructional levels significantly below those assessed by tests such as the WKCE or WRCT.

By definition, an alternate assessment is an assessment used in place of the regular test (Ysseldyke and Olsen, 1999). Procedures for conducting an alternate assessment are still evolving in Wisconsin and in most other states. Two states, Kentucky and Maryland, have been operating alternate assessments with some success for several years as part of a high-stakes state assessment system. These state assessments, unlike in Wisconsin, both emphasize performance assessments of academic and functional skills and require the use of portfolios that are scored by teams of raters using proficiency rubrics (Ysseldyke, Thurlow, Erickson, Babrys, Haigh, Trimble, and Gong, 1996). The Department of Public Instruction in Wisconsin published an Information Update Bulletin in 1998 that describes the state's vision of alternate assessment (see Appendix E for guidelines for complying with the assessment provisions of the Individuals with Disabilities Education Act, Bulletin 98.14). According to this document, when a student cannot take the regular assessment even with accommodations, starting in 1999 data were to be collected and thoroughly reviewed by IEP teams using a wide range of assessment methods, e.g., observations, interviews, record reviews, rating scales, and other tests. Alternate assessments should be curriculum-relevant and standards-based and should reflect the IEP objectives for an individual student. One of the possible tools available to assist educators in achieving this goal is alternate performance indicators. Alternate performance indicators (commonly referred to as APIs) are descriptions of specific knowledge and skills that

follow from the state's content and performance standards, and when demonstrated by a student, serve as meaningful "predictors" or "indicators" of some of the fundamental competencies represented in our state's content and performance standards. Educators in Wisconsin have developed sample APIs in each of the four content areas (English/ language arts, social studies, mathematics, and science) for use with students with severe disabilities and limited English proficiency. Educators can assess a student's knowledge and skills in each domain using a variety of methods, including observations, tests, interviews, records reviews, and rating scales. This array of assessment options and standards-based terminology is designed to offer IEP teams flexibility in assessing students with significant disabilities. The IEP teams are encouraged to thoroughly review the current educational performance of students who are eligible for any state or district assessment using recent, representative, and reliable data. The IEP's review should occur during a time period three to four months prior to the state or district assessment the alternate assessment is replacing.

As you can see, an alternate assessment in Wisconsin requires educators to understand the state's content standards and the use of students' IEP objectives as assessment guideposts for structuring a thorough review of the educational achievement and progress of individual students. For many IEP teams, this thorough review will result in use of an array of methods for collecting information that is recent, representative, and hopefully reliable. We say "hopefully" because if care is not taken in the collection and evaluation of information, the results may not be reliable. Let's look at the issue of reliability and the related concept of validity, given that alternate assessments, like any other assessment, need to be psychometrically sound.

Remember, as we discussed in Chapter 2, central to the notion of reliability is consistency. In the case of an alternate assessment where results are based on the judgments of educators who review an array of evidence about a particular student's learning, reliability concerns the consistency among the judgments of IEP team members, the consistency of judgments over time (say three or four weeks), and the agreement between educators' or stakeholders' judgments of performance and actual test scores of students.

To date, very little research has been done under the name of alternate assessment. A review of the literature will turn up a few technical reports from research centers like the National Center on Educational Outcomes that describe alternate assessment practices in Maryland and Kentucky (Ysseldyke, et al., 1996), but no one has published empirical reports of Wisconsin's approach to alternate assessment. Do not, however, conclude that no research base exists for alternate assessments. In fact, the conceptual and measurement foundations for alternate assessment are well developed and based on years of research in education and psychology related to performance assessment, behavioral assessment, developmental assessment, structured observations, and clinical assessment. Although these assessment methods differ somewhat, they all are (a) based on direct or indirect observation of students, (b) are criterion- or domain-referenced in nature, and (c) require summary judgments about the synthesis of data and the meaning of the scores or results. This latter quality, the use of judgments by knowledgeable assessors, is the empirical foundation for alternate assessment in Wisconsin. Therefore, a brief review of the research literature follows on the accuracy of teachers' judgments of students' academic functioning.

Hoge and Coladarci (1989), as mentioned in Chapter 1, reviewed research on teacher-based judgments of academic achievement, consisting of 16 studies examining the relationships between teachers' judgments of student achievement and students' actual performances on an independent criterion of achievement. The 16 studies they reviewed, along with one additional study, are listed in Table 4.1. in this chapter. Hoge and Coladarci concluded that "the results revealed high levels of validity for the teacher-judgment measures" (p. 297). Studies differed according to how the accuracy of teachers' judgments was assessed. The majority of the studies reported judgment/criterion correlations, and a few reported judgment/performance agreement data. The judgment/criterion correlations of the studies reviewed by Hoge and Coladarci ranged from .28 to .92. "The median correlation, .66, suggests a moderate to strong correspondence between teacher judgments and student achievement" (Hoge and Coladarci, p. 303). Hoge and Coladarci also compared the judgment/criterion correlations among the different methodological dimensions used. Indirect measures had a median correlation of .62 and direct measures had a median correlation of .69.

On the dimension of judgment specificity, studies using rating scales had a median judgment/criterion correlation of .61. This was somewhat lower, although generally consistent with the correlations in studies using ranks (.76), grade equivalents (.70), number correct (.67), and item judgments (.70). Peer-referenced versus norm-referenced judgments did not seem to affect the judgment/criterion correlations. The peer-referenced median judgment/criterion correlation was .68 and the norm-referenced judgment/criterion correlation was .64.

A study by Gresham, Reschly, and Carey (1987) examined the accuracy of teachers in judging academic performance, and in classifying students as learning disabled or nonhandicapped. Although alternate assessments in Wisconsin are not part of the process for identifying children with disabilities, this study is relevant because of its examination of the accuracy of teachers' judgments. The teachers' judgments in the Gresham study classifications were compared to the students' standardized test results. This study consisted of 100 children with learning disabilities and 100 children without learning disabilities. All of the students were given the Wechsler Intelligence Scale for Children-Revised (WISC-R; Wechsler, 1974) and the Peabody Individual Achievement Test (PIAT; Dunn and Markwardt, 1970). Teachers were asked to fill out the Teacher Rating of Academic Performance (TRAP; Reschly, Gresham, and Gresham-Clay, 1987), a five-item scale focusing on reading and math performance. The researchers reported that teachers' judgments of academic achievement were accurate in identifying students as learning disabled or nonhandicapped. Furthermore, teachers' ratings on the TRAP identified children with learning disabilities somewhat more accurately than the WISC-R and the PIAT combined: 96 percent versus 91 percent. The opposite was true for the identification of students without handicaps; the WISC-R and the PIAT were slightly more accurate: 88 percent versus 86 percent. The researchers concluded that regular classroom teachers are accurate "tests" of student academic achievement and could be used as one of the criteria by which psychoeducational tests are validated (Gresham et al., 1987).

In summary, information collected through alternate assessments will be different from that collected for most students in either the WRCT or WKCE, but it still can serve as an index of student progress toward meeting skills related to the academic standards for all students in our state. For reporting and accountability purposes in state and school district reports starting in 2000, students with disabilities who take an alternate assessment in one or more content areas will be described as functioning at the Prerequisite Skills Level. This level of proficiency is so named because it is assumed that a student would have great difficulty answering the vast majority of items on the regular assessment and therefore currently is working on subject matter content that is prerequisite to the skills or competencies measured by the regular test, e.g., WRCT or WKCE.

Reporting Test Results of Students with Disabilities

Score reporting is another aspect of testing programs that influences the participation of students with disabilities. It seems that some educators are concerned that students with disabilities will score lower on tests than many other students, and consequently will lower the overall average score earned by a school and district. A close examination of recent WKCE Performance Reports indicates that some students with disabilities already are functioning at the Proficient and Advanced levels of performance. In addition, the preliminary findings from research with fourth graders indicates that appropriate use of testing accommodations can result in significant increases in students' scores; in 30 percent of the cases examined, students with disabilities who received testing accommodations scored equal to or better than students without disabilities when they were given complex mathematics and science performance tasks (Elliott and Kratochwill, 1999). However, to address the possible concern that students with disabilities will lower a school's or district's scores, test scores for students with disabilities will be reported both together with scores for their nondisabled peers and also will be disaggregated, or reported separately, from those of other students. In addition, the test results from the WKCE and many other assessments are reported via graphic methods which highlight a student's relative strengths and weaknesses within subject matter areas. Such an account of a student's performance provides feedback which many teachers can use to influence their instructional plans for a student. At this time, it

Table 4.1

Studies of Teachers' Judgments of Students' Achievement*

Author	Direct vs. Indirect	Judgment Measure	Reference Group
Airasian, Kellaghan, Madaus and Pedulla (1977)	I	Ratings	NR
Coladarci (1986)	D	Item Response	PI
Doherty and Conolly (1985)	D	Grade Equivalents	NR
Farr and Roelke (1971)	D	Ratings	NR
Gresham, Reschly, and Carey (1987)	I	Ratings	NR
Helmke and Schrader (1987)	D	Number Correct	PI
Hoge and Butcher (1984)	D	Grade Equivalents	NR
Hopkins, Dobson, and Oldridge (1962)	I	Rankings	NR
Hopkins, George, and Williams (1985)	I	Ratings	NR
Leinhardt (1983)	D	Item Response	PI
Luce and Hoge (1978)	I	Rankings	NR
Oliver and Arnold (1978)	I	Grade Equivalents	NR
Pedulla, Airasian, and Madaus (1980)	I	Ratings	NR
Sharpley and Edgar (1986)	I	Ratings	NR
Silverstein, Brownlee, Legutki, and Macmillan (1983)	I	Ratings	NR
Wright and Wiese (1988)	I and D	Ratings and Grade Equivalents	NR
DuPaul, Rapport, and Perriello (1991)	I	Ratings	NR and PI
Demaray and Elliott (1998)	I and D	Ratings and Item Response	NR and PI

Note: D = Direct judgments, I = Indirect judgments, NR = Norm-referenced, and PI = Peer-independent

*Source adapted from "Teacher-based judgments of academic achievement: A review of the literature" by R.D. Hoge and T. Coladarci, 1989, Review of Educational Research, 59, p. 301.

is not considered appropriate to report which students received accommodations and which did not (McDonnell, McLaughlin, and Morison, 1997; Phillips, 1984) because of the possibility of flagging a student as having a disability.

Guidelines for Testing Students with Disabilities: Putting Testing Accommodations and Alternate Assessments into Practice

Up to this point, we have attempted to provide you with a legal and conceptual foundation—with a few do's and don'ts sprinkled in—for understanding testing accommodations and alternate assessments. It is now time to look into some of the details of putting this new knowledge into practice.

Many of the details for guiding the use of testing accommodations are in a document entitled Guidelines to Facilitate the Participation of Students with Special Needs in State Assessments (Wisconsin DPI, 1998; see Appendix D). This document provides specific criteria to facilitate decisions about participation in the WRCT and WKCE assessments for students with disabilities, students receiving services under Section 504 of the Vocational Rehabilitation Act, and students with limited English proficiency. As a starting point for this examination of practical steps for including all students with disabilities in assessment programs, here are the key recommendations highlighted by DPI's guidelines:

- A student's IEP team, which includes the parent(s) as an equal participant, addresses all questions regarding the participation of a student in statewide and districtwide tests.
- State and federal special education laws require that a student's IEP include statements of
  - whether or not the child will participate in the standardized test,
  - accommodations necessary to allow the child to participate in the test, and,
  - if the child is not participating in the test, a statement of why the test is not appropriate and how the child will be assessed.
- To make these statements, the IEP team must know about the child's present level of educational performance and measurable annual goals, the general curriculum, the format and content of the state or district test, and the alignment between the curriculum and the academic content standards assessed by the statewide or districtwide assessment system.
- Participation in the state (or district) test for students with disabilities is not an "all or nothing" decision. Instead, there are multiple options for enabling a student with a disability to participate. These options include
  - participation in the test without accommodations,
  - participation in the test with accommodations, or
  - participation in alternate assessments.
- For the WRCT, the IEP team must choose only one of these options because the test assesses only one content domain (reading comprehension). For the WKCE, however, educators may use these options together depending on the individual needs of the student. That is, you must make separate decisions regarding the need for accommodations or alternate assessment for each content domain (math, social studies, reading, language arts, and science) included in the WKCE. For example, some students with disabilities may not require any accommodations to participate in the WKCE. Other students with disabilities, however, may need accommodations for some of the content domains but not for others. Still other students may need accommodations for some areas within the WKCE and alternate assessment for one or more content domains. Finally, for a limited number of students with disabilities the WKCE will not be appropriate, and the performance of these students will be assessed only through an alternate assessment.
- The IEP team decision regarding student participation in state assessment must be made on an individual basis. As a result, this decision is based upon a thorough review of child-specific data to assess the student's current educational performance relative to the academic performance standards for ALL students.
- This thorough review includes consideration of existing student records, including the most recent evaluation data, formal and informal evaluations conducted by team members, reports by parents and teachers, classroom work samples, independent educational evaluations, and any other information available to the IEP team.
- To make appropriate decisions regarding the student's need for accommodation and/or alternate

assessment, the IEP team should consider the following:

— Begin with the assumption that the student will participate in the test.

— Assess need for accommodation and/or alternate assessment based on the student's present level of educational performance, IEP goals, and the content and format of the test. For the WKCE, independently assess the need for accommodation for each content domain.

— Consider the accommodations the student receives in classroom assessments as possible accommodations for the test.

— Select accommodations that do NOT change the skills or content tested. If the necessary accommodations would change the skills or content tested, assess the student's knowledge and skills through alternate assessment. For example, an accommodation that included reading aloud passages and/or items to students would not be an acceptable accommodation if the purpose of the assessment is to measure reading skills. Thus, a student who would require this accommodation should participate in an alternate assessment for the WRCT or the reading test of the WKCE.

— Use an alternate assessment only if a student would not be able to demonstrate at least some of the knowledge and skills on the WRCT or WKCE assessment with appropriate accommodations.

• Based on the thorough review of the student's current educational performance relative to the academic standards, the IEP team determines how a child with a disability will participate in the assessment system. For those students who are identified as needing accommodations on the standardized test, the IEP team must specify which accommodations are necessary for the child to participate in the assessment.

• The IEP team may determine that, even with accommodations, a child with a disability would be unable to demonstrate at least some of the knowledge and skills on the test. As a result of this decision, alternate assessment will gauge the student's performance. The thorough review used to reach this decision can serve as an alternate assessment if it is documented as part of the IEP process. The review of child-specific data must be recent, reliable, and representative of the student's present level of educational performance relative to the academic standards. In addition, to qualify as an alternate assessment, the IEP team must conduct its review within three or four months prior to the administration of the test. Additional information regarding the Department's position on alternate assessment for children with disabilities under the IDEA can be found in DPI Bulletin 98.14 (see Appendix E or web site http:// www.dpi.state.wi.us/dpi/dlsea/een/bul98-14.html).

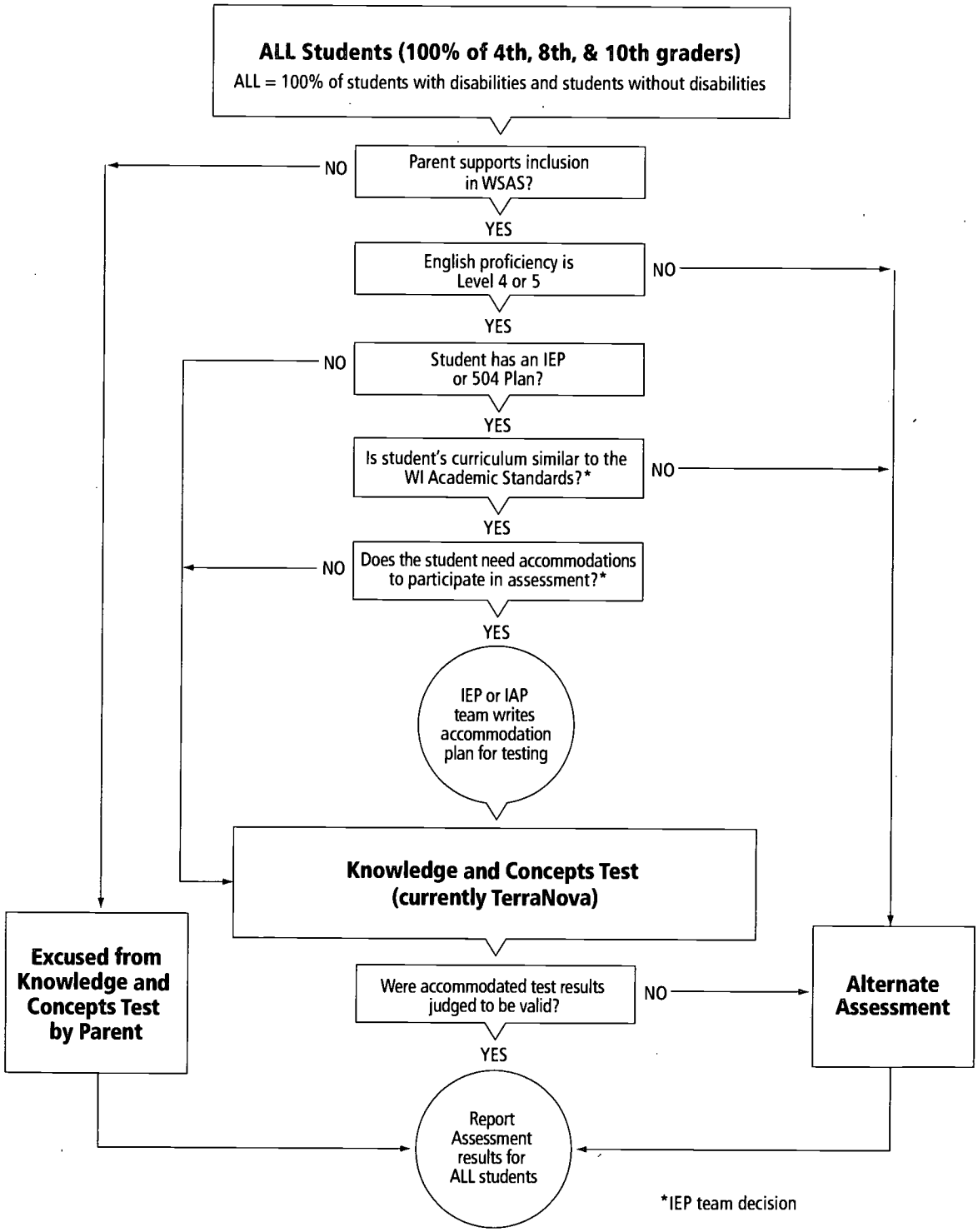
• Test results are not the sole method for making educational decisions involving students with disabilities. Test results are only part of the information used to understand a student and to monitor his or her educational progress.

The flowchart illustrated in Figure 4.3 summarizes the questions educators need to address and the decisions they need to make according to the state's testing guidelines for students with disabilities or limited English proficiency. Take a close look at this flowchart and try to use it to explain the assessment options available to students with disabilities for participating in the WKCE or the WRCT.

Case Applications

In the early pages of this chapter, we briefly introduced you to two students with disabilities: Michele and Ben. If you don't recall them well, take a minute and refresh your memory. To illustrate the application of testing accommodations and alternate assessment procedures, you will receive more information in this section about assessment efforts with Michele and Ben. Before examining the details of Michele's or Ben's cases, let's consider in which type of assessment these two students should participate given their respective grade levels, educational programs, and general competencies. For Michele and Ben, the options include completing a test like the WKCE with or without accommodations, an alternate assessment, or some combination of the WKCE and an alternate assessment. Given that the WKCE format is essentially five tests, i.e., reading, language arts, mathematics, science, and social studies, you are encouraged to think about a student's participation in each of these five tests based on his/her current instructional program. Educators who serve students with severe disabilities have reported that when they are making participation decisions it is helpful to address the following seven questions with regard to each subject matter area:

Decision Flow Chart for Assessing All Students in the Wisconsin Student Assessment System (WSAS)



BEST COPY AVAILABLE

- Is the student's curriculum very different from the district or state grade level content standards? Yes or No?

- Does the student demonstrate cognitive ability and adaptive behavior that prevents completion of the general education curriculum, even with program modifications and adaptations? Yes or No?

- Are the student's management needs intensive and do they require a high degree of individualized attention and intervention from educators? Yes or No?

- Does the student's current adaptive behavior require extensive direct instruction in multiple settings to accomplish the application and transfer of skills? Yes or No?

- Is the student's inability to complete a course of study primarily due to his or her disability, rather than excessive or extended absences, language differences, or social, cultural or environmental factors? Yes or No?

- Is the student unable to apply or use academic skills at a minimal competency level in natural settings such as home, community, or work site? Yes or No?

- Does the student require intensive, frequent, and individualized community-based instruction to acquire, maintain or generalize skills and to demonstrate performance in settings such as prevocational/vocational settings? Yes or No?

The seven questions or issues listed above serve as a participation decision checklist, and when completed can serve as the basis for a justification to include or exclude a student from one or more of the WKCE tested areas. If four or more of the seven questions are answered "Yes," it seems unlikely that the results of the WKCE will be meaningful even with appropriate testing accommodations.

As you read about Michele and Ben, come back to this Participation Decision Checklist (Elliott and Kratochwill, 1998) and see if you agree with the participation decisions made by these students' IEP teams. Remember, answering "Yes" to four or more of the seven points serves only as a guideline for making participation decisions. In most cases, answering "Yes" to four or more of the points in the checklist would suggest that an IEP team believes a student's cognitive capabilities are well below that of agemates, that his or her curriculum is very different in content from what would be expected if it were reasonably well aligned with the state's content standards, and that the student needs extensive assistance to function at school and other community settings. Thus, in effect the content covered in each of the four subject matter areas of the WKCE is highly likely to be very different from the subject matter in the student's daily curriculum. Consequently, to achieve a meaningful assessment of a student with a severe disability, the IEP team will have to utilize an assessment method other than the WKCE. If most of the responses to the checklist items are "No," it is highly likely that those students can participate meaningfully in tests like the WKCE with or without testing accommodations.

Now let's examine the cases of Michele and Ben, and apply what we know about making participation decisions, selecting valid testing accommodations, and conducting an alternate assessment.

The Case of Michele

Michele is a fourth grade, female student with a moderate learning disability primarily with difficulties in reading. She currently receives all her instruction in the regular classroom; however, the regular classroom teacher receives support from a consulting teacher, Mr. Bartlett, who frequently helps to individualize some aspects of instructional tasks for Michele. Ms. Ware, Michele's regular teacher, stresses the use of authentic performance tasks throughout instruction and assessment, particularly in mathematics and science. Ms. Ware also is quite knowledgeable about the state's content and performance standards in the areas of mathematics and science.

Michele is cooperative and motivated to do well. She works more slowly than most of her classmates because she reads slowly and has difficulty composing written responses. Her IEP listed the following instructional accommodations: use of spelling aids to facilitate accuracy in spelling of basic words, additional time to read and comprehend materials, read-along method to facilitate pace and comprehension of difficult text, and use of simple writing webs or diagrams to facilitate planning of written responses.

Classroom teaching and testing experience with Michele

In preparation for the forthcoming IEP team meeting concerning Michele's participation in WKCE, Ms. Ware decided to try and figure out what testing accommodations Michele would ben-

efit from by administering several mathematics performance tasks which she had used in previous years to evaluate all students. She knew these tasks were challenging, requiring quite a bit of reading and spelling, but based on her previous experience administering the WKCE she believed the tasks were a lot like many of the constructed response items on the mathematics and science tests. So she decided to administer the tasks to Michele with as many of her instructional accommodations in place as possible, and then compare her results to the mean scores of students without disabilities in her class. Therefore, Michele was allowed extra time to read and respond to all the tasks, given assistance with reading when she requested, and allowed to use a dictionary and spelling "cheat sheet" with many of her problem words written correctly. Ms. Ware scored all the tasks using a rubric that had been posted in the room and which all her students understood. Specifically, the mathematics and science scoring rubrics ranged from a low of 0 = Not Scorable to a high of 5 = Advanced Response. A score of 1 = Attempted Response, 2 = Minimal Response, 3 = Nearly Proficient Response, and 4 = Proficient Response. Ms. Ware also felt it would be helpful to ask Michele what she thought about the accommodations after she completed the tasks.

Task 1: The Race. "The Race" asks students to analyze various plans presented for a fair running race involving five students. Students need geometric knowledge and direct measurement skills to complete the task. Michele's responses were brief and lacked detail. She set forth her opinion without sufficient rationales. The rules of the race she designed herself were unclear and her rationale was incomplete. Michele's score on the task was a 2.

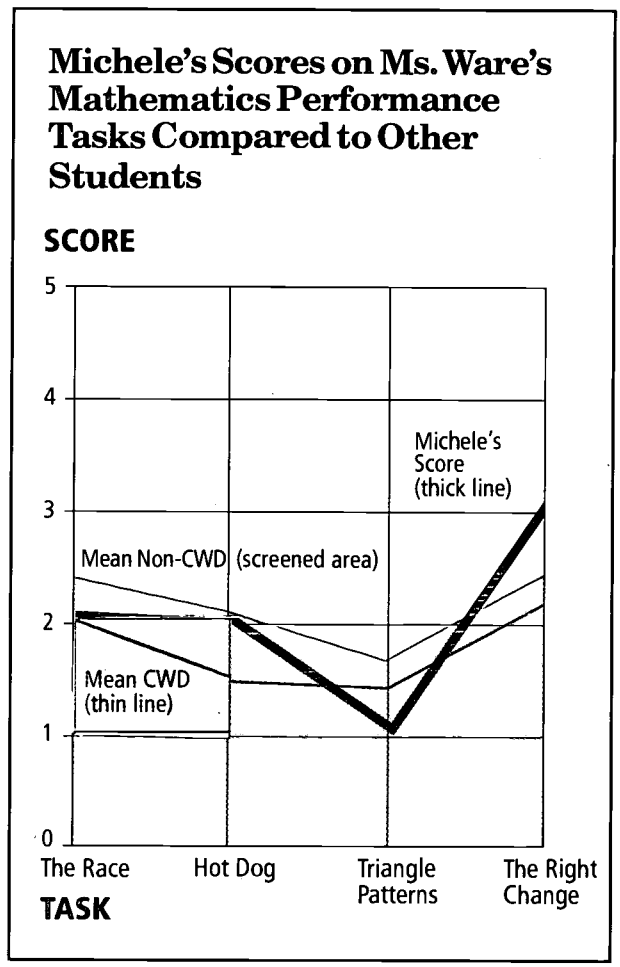
Task 2: Hot Dog. "Hot Dog" requires students to decide how many hot dogs and buns to buy for a picnic. The students need to read and interpret a table, use remainders in division, compute whole numbers, and estimate more than half of an odd number. Michele's calculations were confusing. Her responses were not supported with meaningful rationales or calculations. Michele's score on the task was a 2.

Task 3: Triangle Patterns. "Triangle Patterns" asks students to perform various manipulations with a triangle figure. They need to have geometric knowledge and an understanding of patterns and relationships. Some of Michele's answers about area were incorrect and her figure manipulations were poor. She erased her answers for the items in which she had to state area without an accompanying figure and did not replace them. She did not describe or illustrate the pattern she found. Her score on the task was a 1.

Task 4: The Right Change. "The Right Change" describes a student who purchased various items at the store and asks about the change the student could receive. Students need to calculate with money and know the denominations of coins. Michele provided two combinations of coins that could be given as change. Her explanation of why the clerk would choose a certain combination of coins was imaginative and original. She addressed all aspects of the task and earned a score of 3.

The graphic below (see Figure 4.4) provides a summary of Michele's scores on the four mathematics performance tasks. In addition, Ms. Ware also has included data from a previous class of

Figure 4.4



students who completed the same four performance tasks. This figure shows that Michele, with the use of accommodations to which she was accustomed during instruction, performed similarly to the average of her nondisabled peers and above the average of other students with disabilities that Ms. Ware has taught over the past two years.

Once the tasks were complete, Ms. Ware asked Michele what she liked and didn't like about the tasks. Ms. Ware also wanted to find out what Michele thought about the testing accommodations she had used. Michele stated that what she liked most about the math tasks were that they asked interesting questions and that they were challenging. Regarding the aspects of the tasks she liked the least, Michele said some of them needed more explaining and that the "triangle problem" was too complicated. She mentioned that she had never studied parts of a few of the math tasks. She suggested that the tasks might have been easier for her if they provided more explanation of what students were expected to do and if she had more time to complete them. Armed with the data and knowledge from this practice testing experience with Michele, Ms. Ware listed the following possible testing accommodations for Michele:

- extra testing time;
- more frequent or extended rest breaks;
- distraction-free space or alternative location for an individual;
- directions read and reread as needed;
- clarify student's questions about what to do by asking the student about what is written in the test booklet;
- have student reread directions to teacher and restate in his or her own words;
- allow the special education teacher to administer the test;
- read questions and content to student;
- spelling assistance (use Spellmaster); and
- use of calculator, manipulatives, and ruler.

With this information and the classroom testing experience with Michele, Ms. Ware felt ready for the forthcoming IEP meeting, which she knew would address testing accommodations.

Michele's IEP team meeting

Michele's IEP team needed to meet to update her IEP with regard to participation in WKCE and the possible need for testing accommodations. Michele's teacher in third grade was new to the school district and state, and subsequently had not felt comfortable at the end of the year making decisions about testing accommodations for Michele. Ms. Ware, Michele's mother, the school principal, Mr. Bartlett, and the school psychologist all met before the holiday break to discuss Michele's current educational functioning and her IEP goals, and specifically to make a decision about participation in WKCE and the need for any testing accommodations.

To facilitate and focus participation at the meeting, the school psychologist, Dr. Corey, gave a brief overview of recent changes in federal and state law regarding the participation of all students in assessment programs and provided Michele's mother, Mrs. Moore, a copy of a handout on testing accommodations. Mrs. Moore asked several questions about the WKCE and why it was necessary for students with disabilities to be involved, given that they already had been tested quite a bit in the process of being identified with a disability. Mrs. Moore also indicated that she was unaware of any state academic standards and requested a copy to review. After a rather lengthy discussion about the state's standards and the reasons for all students to participate in assessment programs like the WKCE, the team addressed the issue of Michele's participation in the WKCE. The IEP team answered "No" to each of the seven participation questions. That is, the IEP team believed it was possible for Michele to meaningfully participate in the WKCE.

The team then focused on identifying necessary accommodations to facilitate Michele's meaningful participation in the forthcoming WKCE. At this point, Ms. Ware shared the results of her work with Michele. Each member of the team expressed interest in her findings, but wondered if her findings were applicable to the WKCE. Given her testing expertise, Dr. Corey knew the WKCE well, and consequently was able to address questions about it. She assured the team that although the items might cover different content, many of the skills needed to access Ms. Ware's performance tasks were similar to those needed to access the constructed response items on the mathematics and science portions of WKCE. At this point in the meeting, Dr. Corey reaffirmed that there was a consensus among the team that Michele should participate in the forthcoming statewide test and that she would need some accommodations to minimize the effect of her disability on the validity of the test results. Each team member voiced

Figure 4.5

Accommodations Selected from the Assessment Accommodations Checklist by Michele's IEP Team Members

Assessment Accommodations Checklist™

Assistance Prior to Administering the Test

1 Teach test-taking skills
2 Administer practice activities
3 Other _____

Motivational Accommodations

4 Provide treats, snacks, or prizes, as appropriate
(5) Provide verbal encouragement of student's efforts
(6) Encourage student who may be slow at starting to begin
(7) Encourage student who may want to quit to sustain effort longer
(8) Encourage student to remain on task
9 Other _____

Scheduling Accommodations

(10) Provide extra testing time (indicate how much on student form) ×1½
(11) Allow frequent or extended rest breaks
12 Schedule testing over extra days
13 Administer the test at a time most beneficial to the student
14 Other _____

Setting Accommodations

15 Provide distraction-free space or an alternative location for the student (e.g., study carrel, front of classroom)
16 Place the student in the room or part of the room where he/she is most comfortable
17 Conduct the testing in a special education classroom
18 Conduct the testing at home or at a hospital location
19 Provide for an individual test administration
20 Provide special lighting
21 Provide adaptive or special furniture
22 Provide special acoustics
23 Play soft, calming music to minimize distractions
24 Allow the student freedom to move, stand, or pace during an individualized administration of the test
25 Other _____

Assistance with Test Directions

(26) Read directions to student
(27) Reread directions for each subtask as needed
28 Simplify language in directions (paraphrase)
(29) Clarify student questions regarding what to do by asking the student about what is written in the test booklet.
30 Underline verbs in the test instructions
31 Circle or highlight the task in the directions
(32) Have student reread and restate directions in his/her own words
33 Provide additional practice activities before administering the test.
34 Use sign language or oral interpreters for directions and sample items
35 Color-code instructions to emphasize steps
36 Other _____

Assistance During the Assessment

37 Arrange for a special education teacher or other qualified person to administer test
38 Read questions and content to student
39 Sign questions and content to student
40 Restate the question with more appropriate vocabulary or define unknown vocabulary in the question
41 Turn pages for the student
42 Record student's response (in writing or by audio taping)
(43) Assist the student in tracking the test items by pointing or by placing student's finger on the items
44 Provide spelling assistance, where appropriate
45 Have teacher sit near student
46 Other _____

Equipment or Assistive Technology

47 Text-talk converter
48 Speech synthesizer or electronic reader
49 Visual magnification devices
50 Auditory amplification devices
51 Masks or markers to maintain place
52 Tape recorder
53 Computer or word processor for recording responses
54 Braille writer for recording responses
55 Communications device to indicate responses
(56) Calculator
(57) Manipulatives
(58) Ruler
59 Pencils adapted in size or grip
60 Device that transforms print into a tactile form
61 Arithmetic tables
62 Written list of necessary formulas
63 Noise buffers
64 Other _____

Test Format Accommodations

65 Use lined or grid paper for recording answers when only blank space was provided
66 Provide Braille or large-print editions of the test
67 Audio tape test questions
68 Change presentation format of written material (e.g., increase spacing between lines, reduce number of items per page, print one complete sentence per line)
69 Provide a copy or overhead transparency of diagrams/tables needed for tasks so student does not have to flip back and forth in test booklet
70 Use large-print answer document
71 Use test form with vertically arranged multiple-choice items that have an answer circle to the left of each choice
72 Provide cues such as stop signs or arrows on the test form
73 Mark responses in test book rather than on separate answer document
74 Use a computer for task presentation
75 Other _____

Dr. Elliott and Dr. Kratochwill are faculty members in the Department of Educational Psychology at the University of Wisconsin-Madison. Aleta Gilbertson Schulte is a doctoral student in that department.

Published by CTB/McGraw-Hill, a division of the Educational and Professional Publishing Group of The McGraw-Hill Companies, Inc., 20 Ryan Ranch Road, Monterey, California 93940-5703.
Copyright © 1999 by Stephen N. Elliott, Ph.D., Thomas R. Kratochwill, Ph.D., and Aleta Gilbertson Schulte, M.S. All rights reserved. No part of this publication may be reproduced or distributed in any form or by any means, or stored in a database or retrieval system, without the prior written permission of the publisher. Assessment Accommodations Checklist is a trademark of the McGraw-Hill Companies, Inc.

To order additional copies, call 1-800-538-9547.

BEST COPY AVAILABLE

Figure 4.6

Summary of Michele's Accommodation Plan as Written on the Assessment Accommodations Checklist Form

Student Name Michele Moore Student Identification Number 11642

Grade 4 Test Date 2/24/99

Step 4 Implement the Testing Accommodation Plan

In the space provided, list the recommended testing accommodations. Then detach this page and give it to the person who will administer the test. It should be returned to the student's IEP file when testing is completed.

Accommodation Category	Detailed Description of the Accommodation to Be Used	Subject Areas
Motivational	Verbally encourage student's effort; Encourage her to get started, remain on task, & persist.	All
Scheduling	Provide for 1½ amount of time to complete each task & allow for several breaks.	All
Test Directions	Read test directions, clarify questions about directions, & have her reread or restate directions.	All
Assistance During the Assessment	Help student keep her place and locate correct place for answers, also help with some words but be careful not to invalidate test.	All areas Only Math, Sci, SS
Equipment	Calculator, manipulatives, ruler	Math

Step 5 Report and Evaluate the Use of the Testing Accommodations

After the actual testing session, use the space below to note any changes you made to the testing accommodation plan. If no changes were made, check the box to the right: † No Changes

Accommodation Category	Changes Made to the Accommodation During Testing	Subject Areas

- List any accommodations that may have interfered with the student's performance or invalidated the test score.

  Possible interfering or invalidating accommodations _____

- List additional accommodations that you would recommend on future tests.

  Possible future accommodations _____

† *Page 4 should be detached and given to the person administering the test. It should be returned to the student's IEP file when testing is completed.* 4

agreement with Dr. Corey, although it was clear that Mr. Bartlett, the special education teacher, had some reservations. Dr. Corey then introduced a copy of the Assessment Accommodations Checklist (see Figure 4.5) and noted that the team could use it to help develop a testing accommodations plan and to communicate the plan with others who would be responsible for administering tests. The team members agreed to try the Assessment Accommodations Checklist and came up with the following list of accommodations, which they thought would be reasonable and would increase the validity of Michele's test scores on the mathematics, science, and social studies portions of WKCE:

- verbal encouragement of student's effort;
- extra testing time (1 1/2 time);
- more frequent or extended rest breaks;
- distraction-free space/alternative location for a small group;
- directions read and reread as needed;
- clarify student questions about what to do by asking the student about what is written in the test booklet;
- have student reread directions to teacher and restate in his/her own words;
- allow the special education teacher to administer the test;
- read questions and content to student;
- assist the student in tracking test items by pointing or placing the student's finger on the items;
- encourage student to begin, remain on task, and sustain effort longer before quitting; and
- use of calculator, manipulatives, and ruler.

There was more disagreement about the accommodations Michele needed to meaningfully participate in the reading and language arts test. They knew, of course, that the content of the items could not be read to Michele, but both teachers felt it might be reasonable to read the possible answer choices on the multiple-choice items. In addition, there was some debate about the amount of time Michele would need to complete the test. Ultimately, the team endorsed the same list of accommodations for the reading and language arts tests excepting only the accommodation of "reading questions and content to student."

As a result of the meeting, the IEP team developed a feasible testing accommodation plan that should facilitate Michele's meaningful participation in the forthcoming statewide test. Implementation of the plan will require the attention of a test administrator who is responsible for only a few students and a testing setting where communication between Michele and the test administrator can occur without disrupting other test takers. A copy of the testing accommodation plan summarized from the Assessment Accommodations Checklist for a test administrator is displayed as Figure 4.6. If these accommodations are carried out, it is the professional judgment of the IEP team that the resulting scores will be better indicators of Michele's abilities in mathematics, science, social studies, and reading and language arts. Thus, the accommodation plan is designed to increase the likelihood that Michele actually takes the test and that her scores provide a valid indication of her abilities.

The Case of Ben

Ben is an eighth-grade student with autism. He receives the majority of his instruction in a highly structured classroom with six other students, his teacher, Ms. Zwick, and her teaching aide. For mathematics, however, Ben participates in a consumer mathematics class with sixth grade students. Like many students with autism, Ben's oral communication and interpersonal skills are limited. Consequently, he requires extensive instructional support and spends the majority of his school day working on functional communication and daily living skills.

Ben's IEP team meeting

Ben's IEP team concluded that he should receive an alternate assessment due to the pervasive nature of his disability and the fact that his current educational curriculum was very different from the curriculum of a majority of his agemates. Specifically, Ben's IEP team members answered six of the seven participation decision checklist questions "Yes" when reflecting on his work in reading, science, and social studies. Therefore, they recommended that Ben not participate in WKCE in these subject matter areas. The team disagreed about Ben's mathematics skills and curriculum alignment with the WKCE, but ultimately decided that even in mathematics the WKCE was unlikely to provide meaningful information about his skills. In place of the WKCE,

BEST COPY AVAILABLE

Ben's IEP team needed to conduct an alternate assessment.

Ms. Zwick volunteered to provide leadership in conducting the alternate assessment. Each of the other team members—Mr. and Mrs. Horner (Ben's parents), Dr. Carroll (school psychologist), and Ms. Wayley (principal)—agreed to help her. Nevertheless, they felt somewhat at a loss as to what test to use to assess Ben's skills in mathematics, reading, science, and social studies and how to accommodate him during the test. Ms. Zwick explained to the team members that an alternate assessment could involve a wide range of assessment methods in addition to a test. In fact, she indicated that it was unlikely that even a developmentally appropriate test would provide a valid indication of Ben's knowledge and skills, given the focus of his IEP. Instead, she suggested the team examine the rather substantial collection of classroom work samples that Ben had produced in mathematics and reading, and also review the weekly notes that she and her aide had written over the past four months. Most of these progress notes concerned learning objectives on Ben's IEP and focused on communication skills, social skills, and self-care skills. In addition to the notes, Ms. Willis, the classroom aide, had videotaped Ben during three instructional sessions when he was working on eye contact, listening, and expressing his approval or disapproval by using the words "Yes" or "No" without his routine hand-flapping.

With Ms. Zwick's leadership, the IEP members agreed to review the collected materials which served as evidence of Ben's current knowledge and skills. Dr. Carroll, the school psychologist, however, questioned whether the evidence was enough. Though it seemed the evidence was recent and representative of what Ben had been doing in mathematics and reading or language arts, he said he didn't see any evidence of work in science or social studies. Ben's parents disagreed mildly; they felt that the objectives on his IEP concerning social skills and self-care skills were basic social studies skills. This point provoked quite a bit of discussion among the team members and generated a number of questions that nobody could answer with confidence. For example, if Ben's IEP didn't have any learning objectives concerning science, did the alternate assessment still have to document his achievements in science? How far downward can one work developmentally from the state's content and performance standards and still be assessing skills in mathematics or reading? How does one reliably score Ben's performances and how are scores reported?

Ms. Wayley was the first to admit that she was getting confused and a little uncomfortable doing an alternate assessment. She commented, "I know there is error in any measurement, that is, all assessments have some error. But it seems like an alternate assessment can be full of error and the resulting scores might be meaningless given that every student could have a different assessment."

Ms. Zwick responded politely but firmly to Ms. Wayley's comments. "I have been a teacher for twelve years and have been responsible for evaluating the performances of hundreds of students," she said. "There is strong evidence that teachers like me can be excellent judges of students' work. That means teachers' judgments can be reliable and valid."

At this point, Dr. Carroll interrupted and asked, "What data do you have to support your belief that teachers are reliable and valid judges of students' work?"

Ms. Zwick was a little surprised by this challenge, but she welcomed it even if she couldn't quote a reference to a research article as Dr. Carroll occasionally did.

"Well," she said, "the best evidence I am aware of concerns comparing teachers' predictions of students' test performances to the students' actual performances. Several researchers have published work on this in major education and psychology journals. In addition, I have recently read work about the use of scoring rubrics as tools to enhance the reliability of teachers' evaluation of students' work in language arts and mathematics. So I think there is good research to support my belief that teachers can be highly accurate judges of students' work!"

Ms. Wayley spoke up. "Okay, I like the knowledge and confidence you have, Ms. Zwick, and I know we must do an alternate assessment, so will you take the lead and guide us through an alternate assessment of Ben?"

Ms. Zwick agreed to be the team leader for Ben's alternate assessment. She suggested that over the course of the next two weeks she and her aide would organize their evidence about Ben's learning and academic progress in mathematics, reading, language arts, and social studies. They also would review what, if anything, Ben had done in the area of science. Another meeting was scheduled about a month before the WKCE when the entire team could meet for a thorough review of Ben's work and provide a summary report of their results for purposes of accountability in the WKCE.

Figure 4.7

Ms. Zwick's Alternate Assessment Summary Form for Reading

ALTERNATE ASSESSMENT SUMMARY
READING

NAME: Ben Horner
AGE/GRADE: 13/8th

SCHOOL: Bailey
DISABILITY CATEGORY: Autism

ACADEMIC STANDARD: ENGLISH /LANGUAGE ARTS	RELEVANT TO CURRENT INSTRUCTIONAL PROGRAM	PERFORMANCE STANDARD EVIDENCE				PROFICIENCY STANDARD				
		WSAS Test TerraNova	Work Sample	Direct Observation	Other Tests	Prerequisite Skills	Minimal Performance	Basic	Proficient	Advanced
A. Reading/Literature										
4.1 Uses strategies	Yes / (No)	+								
4.2 Reads, interprets & analyzes	Yes / (No)	+								
4.3 Discuss texts	Yes / (No)	+				✓				
4.4 Acquire information	Yes / (No)	+								
Other: See IEP	(Yes) / No	-	+	+						
B. Writing										
4.1 Communicate with others	Yes / (No)	+								
4.2 Plan, revise, edit, publish	Yes / (No)	+				✓				
Other: See IEP	(Yes) / No	-	+	+						
C. Oral Language										
4.1 Communicate with others	Yes / (No)	-								
4.2 Comprehend oral communications	(Yes) / No	-								
4.3 Participate in discussions	Yes / (No)	-				✓				
Other: See IEP	(Yes) / No	-	+	+	+					
D. Language										
4.1 Understands forms & punctuation	Yes / (No)	+								
4.2 Vocabulary	Yes / (No)	+								
4.3 Interprets use & adaptations	Yes / (No)	+				✓				
Other: See IEP	(Yes) / No	-	+	+	+					
E. Media and Technology										
4.1 Uses computer to communicate	Yes / (No)	-								
4.2 Makes judgments about media	Yes / (No)	-								
4.3 Creates products	Yes / (No)	-								
4.4 Knowledge of media production	Yes / (No)	-				✓				
4.5 Analyze & edit media work	Yes / (No)	-								
Other: —	Yes / (No)	-	—	—	—					
F. Research and Inquiry										
4.1 Conduct research & communicate findings	Yes / (No)	-				✓				
Other: —	Yes / (No)	-	—	—	—					
English Language Arts Total Domain						✓				

BEST COPY AVAILABLE

Ben's alternate assessment

At the meeting, which the IEP team designated as the alternate assessment, Ms. Zwick started with a review of the state's policy on alternate assessment and an overview of the state's academic content standards. With this information as background, the team agreed on three main points at the outset of the meeting: (1) for purposes of statewide accountability reporting, Ben was functioning at the Prerequisite Skills levels in mathematics, reading, language arts, social studies, and science, even though he was not really doing any class work on any of the science standards; (2) they had collected substantial evidence concerning Ben's academic functioning such as classroom work samples, teacher's progress notes, videotapes of communication skills, and parents' observations, all of which they deemed recent and representative of his work, and the evidence was very similar to some of the alternate performance indicators (APIs) a statewide committee had developed in 1998; and (3) the main challenge the team faced was documenting and interpreting the evidence for instructional use.

Ms. Zwick presented most of the evidence about Ben's functioning in a portfolio notebook which made it easy to review and to see the date when the work was completed. She also had organized the information by subject matter and provided her own documentation system (see Figure 4.7 for a sample page) that indicated how Ben's work aligned or didn't align with state academic standards for fourth graders. She used the fourth-grade standards because they represented the most reasonable developmental level of comparison and had influenced the development of the state APIs for use with students with disabilities. As a means of interpreting the evidence and communicating with each other about the quality of Ben's work, Ms. Zwick suggested the use of a scoring rubric that emphasized three dimensions: the frequency or quality with which a skill is exhibited, the range of settings in which a skill is exhibited, and the amount of support a student needs to exhibit the skill (see Figure 4.8). She had seen similar scoring rubrics used to characterize the development of skills in novice learners regardless of whether they had a disability, Ms. Zwick noted, and she recently had attended a workshop on alternate assessment which featured a similar scoring rubric. Impressed with the rubric, Dr. Carroll remarked that he now was able to more fully appreciate Ms. Zwick's points about reliable and valid scores. Ms. Wayley also shared her positive opinion about the use of the scoring rubric, noting that she felt it could be a meaningful scoring system for all students doing an alternate assessment. Relieved to hear that her colleagues liked what she had developed, Ms. Zwick turned to Mr. and Mrs. Horner to see how they were reacting.

Mr. Horner said, "I think I can use this rubric or scoring method you have developed to talk about

Figure 4.8

Ms. Zwick's Rubric for Scoring the Evidence Collected for Ben's Alternative Assessment

Proficiency Level Scoring Criteria

0 = The student's knowledge or skill is **Nonexistent** if there is no observable evidence during the past 6 months.

1 = The student's knowledge or skill is **EMERGING** if the student is beginning to show understanding or use of the skill in one setting with extensive instructional support during the past 6 months.

2 = The student's knowledge or skill is **DEVELOPING** if the student occasionally shows understanding or use of the skill in one or more settings with moderate instructional support during the past 6 months.

3 = The student's knowledge or skill is **DEVELOPED** if the student frequently shows the ability to apply the knowledge or skill in more than one setting with little instructional support during the past 6 months.

4 = The student's knowledge or skill is **ADVANCED** if the student almost always shows the ability to apply the skill appropriately in most settings independently during the past 6 months.

NA = This knowledge or skill is currently **Not Relevant** to the student's educational program.

Ben's work, but how is it going to help me and others compare his work to other students taking the regular test?"

There was silence for a moment. Dr. Carroll started to answer the question, but recognized that Ms. Zwick was eager to respond.

"Well, we have already made a comparison of Ben to other students when we decided that he should not participate in the WKCE," Ms. Zwick said. "Our decision communicated that Ben was working on curriculum materials that were significantly below the eighth-grade level and that his overall skills in the subject matter areas were not developed to the point where participation in the test would provide a meaningful measurement of what he was learning.

"Remember," she continued, "we agreed Ben was functioning at the Prerequisite Skills level in mathematics, reading, language arts, social studies, and science. This tells educators and others who understand our state's proficiency standards that Ben is functioning at the lowest proficiency level in the targeted subject areas. However, this does not mean that Ben is not learning. We know he is because we have a lot of evidence from my classroom to prove he is making progress. Each of you has seen this evidence!"

"I didn't mean to insult you," Mr. Horner calmly responded. "I know Ben is making progress on some things, but I thought when you did an assessment or test you needed to compare Ben to other students."

Dr. Carroll now spoke up. "Mr. Horner, most people think about assessment results like you do," he said. "That is, they compare one student's scores to the other students'. That is called a norm-referenced comparison. We can do this with the results from the WKCE because each student who takes it is given the same items and they all earn a score that is comparable. The same scores, however, can be compared to a descriptive standard of proficiency. In Wisconsin, we have four levels of proficiency that are based on scores from the WKCE. The lowest level is called Minimal Performance, the next highest is called Basic, then there is Proficient, and finally the highest level is called Advanced. If a student doesn't take the WKCE, as we decided with Ben, then his or her level of proficiency is considered to be lower than Minimal Performance and consequently this level of proficiency is called the Prerequisite Skills level. The use of proficiency levels is an example of a criterion-referenced comparison and it allows us to compare a student's performance to a common standard as well as to other students who also participated in the same assessment. Is this making sense to you?"

Mr. and Mrs. Horner acknowledged to Dr. Carroll that it did make sense to them and that it was comforting to know. Though they knew Ben was not functioning at a high level academically, they still found the comprehensive review of his work to be meaningful. Mrs. Horner concluded by saying, "It seems like more parents would want their child to have an alternate assessment because it actually allows you to see what your child can and cannot do. I know it doesn't result in a fancy or high score, but it still tells me a lot about my son. It communicates well!"

Ms. Zwick, who had been monitoring the discussion, announced that the team still should spend a little more time on scoring Ben's work using the rubric she had presented. She encouraged each member of the team to look independently at the evidence and to select a number or level within the scoring rubric that best characterized the work. Once each member had done this for Ben's mathematics work they shared their perceptions and discussed any disagreements. The consensus rating for Ben's mathematics work was characterized as "Developed," which resulted in a score of 3. They used a similar process to summarize his work in reading, language arts, and social studies. In these subject matter areas, the team members came to a consensus characterization of "Developing," or a score of 2. With regard to science, there was no evidence to evaluate. Ben's IEP did not contain any skills concerning science. The IEP team agreed that, according to the rubric, Ben's work was best characterized as "Nonexistent," or quantitatively a score of 0. As a result of this assessment, however, they decided that when the IEP was reviewed at the end of the year, they should consider some basic skills in science.

After they scored Ben's evidence, Ms. Wayley encouraged the team to summarize its alternate assessment efforts in a brief report that could be placed in Ben's IEP file. Ms. Zwick echoed this recommendation and suggested using a simple report card style, similar to what students receive when their WKCE tests have been scored. Specifically, Ms. Zwick suggested the following reporting format for Ben (see Figure 4.9).

With the report completed, the team concluded their alternate assessment of Ben. The assessment had given them an opportunity to communicate

Figure 4.9

Ms. Zwick's Alternate Assessment Summary Report Form for Documenting Ben's Performance in the WSAS

ALTERNATE ASSESSMENT SUMMARY REPORT

Student: Ben Horner **School:** Bailey **Grade:** 8

Teacher: J. Zwick **Date Assessment Completed:** 1-18-99

	Reading/ Language Arts	Mathematics	Science	Social Studies
Advanced				
Proficient				
Basic				
Minimal Performance				
Prerequisite Skills	✓	✓	✓	✓

Notes: **Detailed information about this student's performance in each of the subject matters assessed is documented in an individual report written by the student's IEP team.**

about Ben's progress and to put it in the context of the state's academic standards. The team members felt that Ben's assessment was meaningful and provided valuable feedback to his parents and teachers. In addition, Ben and his assessment results also were included in the state's accountability system, helping to provide a more complete story about the achievement of all students.

Preparing All Students to Take Tests

The goal of teaching is to increase learning rather than to increase test scores. The Wisconsin DPI's Guidelines for Appropriate Testing Procedures makes this assertion in capital letters. Within these same guidelines, teachers are reminded that "students' attention and effort should be directed to learning the entire scope of the curriculum, not just the limited knowledge and skills measured by the WKCE" (1998, p. 6). It is clear from this statement and others that DPI does not encourage "school staff to buy, develop, or promote the use of extensive test practice materials that closely parallel the WKCE's items or tasks" (1998, p. 6). Test preparation, these guidelines withstanding, is a frequent concern of many educators and parents. Given changes in requirements concerning the participation of all students in assessment programs and the emphasis on testing as a major aspect of promotion and graduation decisions, it is anticipated that test preparation efforts will increase. Therefore, we believe it is worthwhile to understand the role and ethics of test preparation for all students.

Many sound test preparation practices may appear to be common sense activities; however, our experience with many educators suggests otherwise. Consider the test preparation strategies listed below that some teachers reportedly use when administering tests like those in WKCE. As you read through the list, critically evaluate the strategies to determine which ones you believe are appropriate and which are not appropriate.

- Limit instruction during the month prior to the test to only those objectives that are thought to be on the test.
- During instruction, use examples from last year's test.
- Give students an opportunity to practice taking the actual test items before they formally start the test.
- Teach students general test-taking skills, e.g., listening carefully to directions, reading the entire question before answering, to improve their test performance.

To determine which of the four strategies are educationally and ethically sound, use two guiding principles: (1) the educational objectives, the content of instruction, and the content of the achievement test should be aligned or strongly related to each other and (2) the general purpose of an achievement test is to inform educators and students how well the students have learned what has been taught. Thus according to Airasian (1994), the important issue becomes just how strong the relationship should be among learning objectives (performance standards), instruction, and the test. The National Council on Measurement in Education (NCME) Task Force (Canner, et al., 1991) actually provided a set of guidelines for determining appropriate and inappropriate test preparation. *Its basic guideline states that all test preparation activities that lower the validity of interpretations made from test scores are inappropriate and should be avoided.* The guidelines of the NCME Task Force indicate that the following test preparation activities are inappropriate or unethical:

- focusing instruction only on task or item formats used on the test;
- using examples during instruction that are identical to test items or tasks; and
- giving students practice taking the actual items on which they will be tested in the near future.

Ultimately, the issue of proper test preparation is one of validity. That is, the assessment of student achievement should provide a fair and representative indication of how well students have learned what they have been taught. In order to do this, test questions must focus on knowledge and skills similar to those students learned during instruction. Perhaps the most important word in the previous sentence is similar. There is an important ethical difference between teaching to the test (and content standards) and teaching the test itself! Teaching to the test is a desirable practice and involves teaching students the general knowledge and skills they need to answer questions on the test and to succeed in future education and work settings. Teaching the test itself involves teaching students the answers to specific

BEST COPY AVAILABLE

questions that will appear on the test. This is neither pedagogically appropriate nor ethical because it can result in a distorted or invalid picture of what students have achieved.

Good test preparation should enable students to show what they have learned in classes over the past several years. Therefore, it is helpful for all students to understand that when taking a test they should

- be well rested and comfortable at the time of testing;
- attend carefully to test directions and follow directions exactly;
- ask questions when they are unsure of what to do;
- find out how questions will be scored;
- pace themselves so they do not spend so much time on some questions that they cannot get to other questions;
- plan and organize essay questions before responding;
- act in their own interest by attempting to answer all questions; and
- when using a separate answer sheet, check often to make certain they are marking their responses accurately and in the correct place.

Besides these general test-taking guidelines, experts who study test-taking (what has become known as test-wiseness) suggest some additional skills that provide students some strategies for answering test questions (Linn and Gronlund, 1995; Sarnacki, 1979). Most of these test-wise skills relate to errors on the part of question writers who provide clues to correct answers. For example, when responding to multiple-choice questions, a test-wise student knows that

- the answer option that is longest or most precisely stated is likely to be the correct one;
- answer choices that do not attach smoothly to the item stem are not likely to be correct; and
- the use of words such as some, often, or similar vague words in one of the answer choices is likely to indicate the correct option.

In summary, virtually all students can master many good test-taking skills, but students need some practice to develop these skills and confidence in using them. Consider spending instructional time a couple weeks prior to an important test discussing and modeling good test-taking skills for all your students. Remember, however, test preparation should not raise test scores without also raising students' mastery of the general content being tested. Thus, test preparation and test-taking skills are designed to increase the validity of students' test scores, not necessarily to increase their scores.

A Concluding Thought

Educators are now empowered and entrusted to include all students in the various assessment systems that have been implemented in their districts and across our state. New policies and practices that involve testing accommodations and alternate assessments should help to achieve this expectation. In this chapter, we have examined in some detail the use of testing accommodations and alternate assessments as the two primary tactics available to facilitate the meaningful participation of students with special needs in assessments. Wise use of these assessment tactics rests upon an understanding of the concept of test score validity and an appreciation of good assessment as part of good instruction.

References

Airasian, P. W. *Classroom Assessment, 2nd Edition.* New York: McGraw-Hill, 1994.

Canner, J., et al. *Regaining Trust: Enhancing the Credibility of School Testing Programs.* National Council on Measurement in Education Task Force. Mimeo, NCME (1991).

Elliott, S.N., and T.R. Kratochwill. *Experimental Analysis of the Effects of Testing Accommodations on the Scores of Students With and Without Disabilities: Mid-project Results.* Presented at CCSSO's Large-Scale Assessment Conference, Snowbird, UT, 1999.

Linn, R.L., and N.E. Gronlund. *Measurement and Assessment in Teaching.* Englewood Cliffs, NJ: Merrill, 1995.

McDonnell, L.M.; M.J. McLaughlin; and P. Morison, eds. *Educating One and All: Students with Disabilities and Standards-Based Reform.* Washington, DC: National Academy Press, 1997.

Sarnacki, R.E. "An examination of test-wiseness in the cognitive domain." *Review of Educational Research* 49 (1979), pp. 60-79.

Thurlow, M.L.; J.L. Elliott; and J.E. Ysseldyke. *Testing Students with Disabilities: Practical Strategies for Complying with District and State Requirements.* Thousand Oaks, CA: Corwin Press, 1998.

Best Practices in Assessment Programs for Educational Accountability

5

Today, perhaps more than ever, there is a strong interest in getting a clear and complete picture of how well students are learning and how well schools are functioning. Consequently, assessing all students is not only an important part of educational accountability, it also is the law. As you know, however, our schools educate a diverse group of students, including students with disabilities. To meaningfully assess the learning of all students with disabilities is a challenging task. Fortunately, the laws and regulatory procedures that guide the delivery of services for students with disabilities allow for the use of two assessment tactics, testing accommodations and alternate assessments, to facilitate the participation and meaningful assessment of students with disabilities. This book primarily has focused on an understanding and intelligent use of these two assessment tactics.

A Summary of Inclusive Assessment Practices

Because there is little published research to date on these tactics, decisions about the use of testing accommodations and alternate assessment need to be guided by common sense, state testing guidelines, and a sound understanding of test validity. To guide your use of testing accommodations, we have stressed the following points:

- An IEP team must make decisions about testing accommodations and base them on the *individual needs* of a student, not on the student's disability category.
- Document the testing accommodations to be used and those actually used on the student's IEP, and be sure to communicate the IEP team's plan to accommodate a student to his or her parents and the individual responsible for administering the test.
- Accommodations a student currently receives during classroom instruction provide the starting point for selecting possible accommodations that will facilitate test-taking. Using accommodations a student has not experienced previously can actually create more problems.
- The purpose of a testing accommodation is to enhance the validity of the inference made from a student's test score; therefore, appropriate testing accommodations should affect access or enabling skills, not the skills or abilities targeted by the test.
- The list of known invalidating accommodations is actually quite short, and includes reading aloud a reading test, using a calculator on a mathematics test designed to measure mental mathematics, using spelling aids on a test where points are allocated for correct spelling, and using excessive paraphrasing of content that results in changing the meaning or level of difficulty of the material.
- IEP teams will need to meet to make testing accommodations plans several weeks prior to the actual test to ensure that testing personnel have time to coordinate accommodation plans for the entire group of students who need them.
- Testing accommodations must be reasonable and feasible: reasonable with respect to the number and type of accommodations the student receives on a regular basis in his or her classroom, and feasible in that the individual administering or managing the accommodation has the resources and skills to accurately implement the accommodation.
- If after completing a test, you believe the accommodation(s) used invalidated the results, report it to the test coordinator and arrange for another administration without using the specific invalidating accommodation(s) or consider conducting an alternate assessment of the student.

When a student cannot meaningfully participate in a test, such as the reading portion or math portion of the WKCE or the WRCT, even with a comprehensive accommodation plan, you must

design and administer an alternate assessment. To guide your use of alternate assessments, we have emphasized the following points:

- IEP teams are responsible for making the decision about participation in an alternate assessment based on a series of issues, utmost of which concerns the mismatch between the instructional level at which an individual student is working and the content and learning expectations characterized by the assessment. Decisions about participation should not be based on a student's disability category.

- IEP teams are responsible for conducting the alternate assessment, which at a minimum must involve a thorough and timely review of the student's achievement and progress within the academic standards framework (APIs in language arts, mathematics, science, and social studies) to which all students in Wisconsin are held accountable. The focus of the alternate assessment can cover areas in addition to those embodied by the state standards.

- A variety of assessment methods, including observations, records reviews, work samples, performance tasks, and developmental or diagnostic tests can be used to collect evidence that provides the basis for the IEP review. The results of these assessments should be summarized in writing, documented in the IEP, and stored for review by others, in particular the student's parents and future teachers.

- Inclusive accountability practices and federal law suggest that the assessment results for each student who participates in an alternate assessment be reported with the same frequency, level of detail, and time as the results of students participating in the regular assessment. The primary reporting method of alternate assessment results for the public in Wisconsin will be information from schools about the number of students who are functioning at the Prerequisite Skills level. More detailed reports about students' achievements and progress should be provided to parents, but not aggregated in a summary because the types of tasks and assessment methods used are likely to be quite variable, thus not directly comparable.

- Alternate assessment, like any other assessment, must be recent, reliable, and a representative sample of a student's skills and abilities. When these conditions are met and the content of the assessment is aligned with the state's content standards framework, the results can be interpreted with confidence.

Fair Testing Practices Require Efforts from Many People

Research suggests that teachers spend as much as a third of their time involved in some type of assessment. Teachers are continually making decisions about the most effective means of interacting with their students. These decisions are usually based on information they have gathered from observing their students' behavior and performances on learning tasks in the classroom and on standardized test results (Witt, et al., 1998).

Many individuals have a vested interest in student learning and assessment information about such learning. Clearly, teachers, students, and parents should have great interest in the results of student assessments. School administrators and community leaders also voice keen interest in assessment results that document students' performances. No single assessment technique or testing procedure, however, can serve all these potential users. Thus, the purpose of one's assessment must be clear, for it influences assessment activities and, consequently, the interpretation of any results.

Teachers have two main purposes for assessing students: (a) to form specific decisions about a student or a group of students and (b) to guide their own instructional planning and subsequent activities with students. Teachers use assessment results for specific decisions, including diagnosing student strengths and weaknesses, grouping students for instruction, identifying students who might benefit from special services, and evaluating students' progress against state standards of performance and proficiency. Teachers also use assessment activities and results to inform students about teacher expectations. In other words, the assessment process can provide students with information about the performance necessary to achieve success in a given classroom and grade. Tests become a critical link in teaching when teachers provide students with clear feedback about results. Assessments likewise provide teachers valuable feedback about how successful they have been in achieving their instructional objectives, and thus help them chart the sequence and pace of future instructional activities.

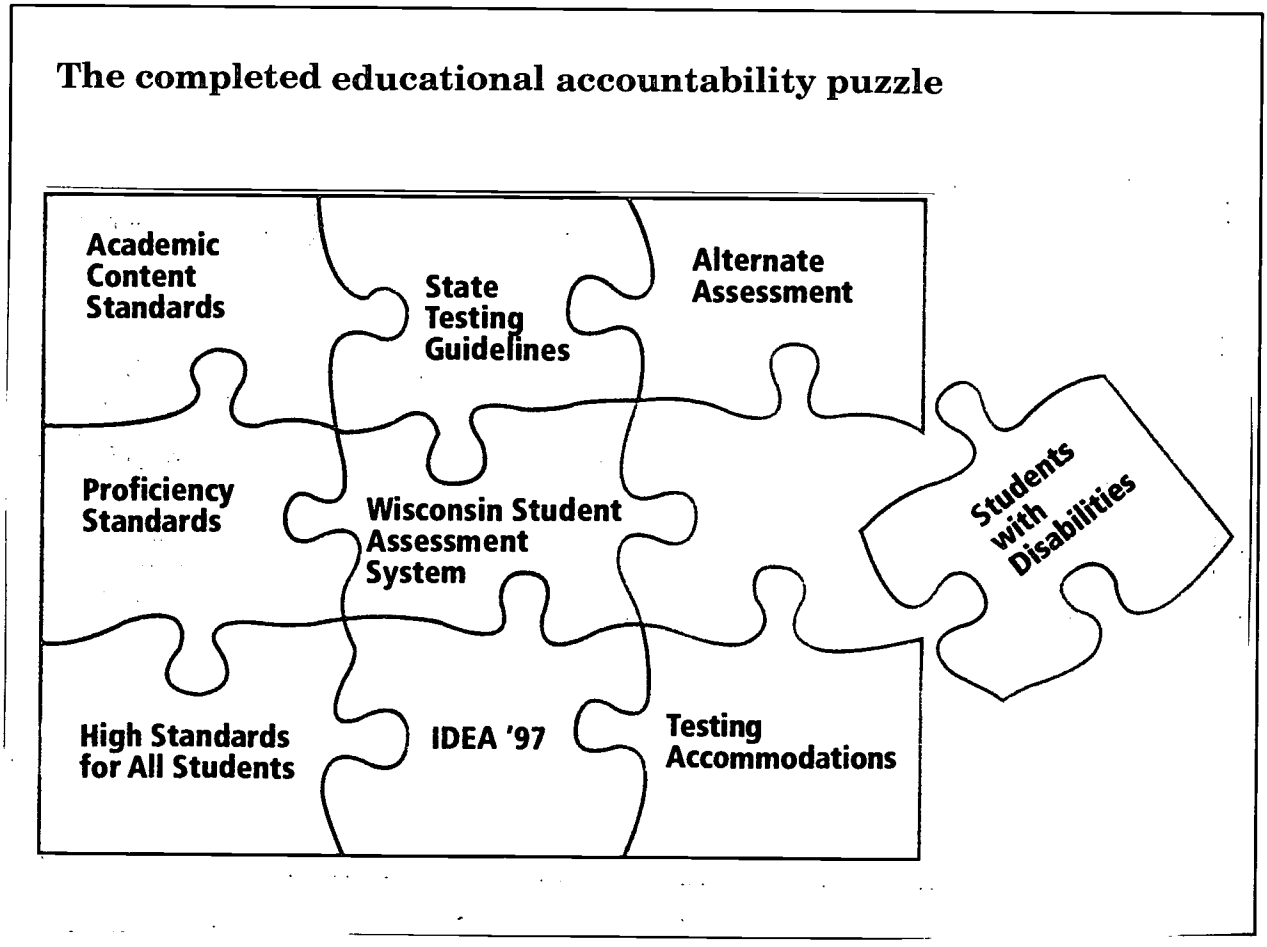
Students also are decision makers and use classroom assessment information to influence many of their decisions. For example, many

students set personal academic expectations for themselves based on teachers' assessments of prior achievement. Feedback they receive from teachers about their performances on classroom and standardized tests can directly affect students' decisions about strengths and weaknesses, interests, study activities, and possible career plans.

Teachers' assessment activities and decisions affect parents as well as students. For example, many parents communicate educational and behavior expectations to their children. Some parents also plan educational resources and establish home study environments to assist their children. Feedback from teachers about daily achievement, classroom tests, annual standardized tests, and statewide assessments often significantly influence parents' perceptions of their child and his or her teachers. Testing results also provide parents and others in the community with information about the school's performance. That is, does the school prepare students for the basic skills of reading, writing, and calculating? In sum, results from assessments of children's learning can significantly influence parents' attitudes about their children and schooling.

Clearly, the enterprise of assessing students often is very important in the lives of teachers, students, and many parents. Recognizing this, a joint committee on testing practices from major educational and psychological organizations developed a *Code of Fair Testing Practices in Education* (American Educational Research Association, 1988). This code contains standards for educational test developers and users in four areas: developing and selecting tests, interpreting scores, striving for fairness, and informing test-takers. The code is meant for use by the general public and is included in its entirety as Appendix F. With its focus on fairness and appropriate interpretation of test scores, the code serves as an appropriate conclusion to this book on educational assessment and the inclusion of all students in assessment programs.

Figure 5.1



BEST COPY AVAILABLE

Completing the Educational Accountability Puzzle

As stressed throughout this book, assessing all students is an important and sometimes challenging undertaking that requires knowledge of testing practices, test content, legal guidelines, and technical aspects of tests, as well as a clear understanding of students' learning objectives and instructional programs. If educators in Wisconsin are going to actualize the requirements of IDEA '97 and the potential of standards-based education for ALL students, then ALL educators will need a strong understanding of the state's standards, the content of the tests covered in WSAS, the state's testing guidelines, the valid use of testing accommodations, the valid use of alternate assessments, and how to communicate these assessment results to educational stakeholders.

As indicated early in this book, there are at least nine pieces to the educational accountability puzzle in Wisconsin (see Figure 5.1). As a result of reading this book and talking with colleagues about assessment activities like those required by the WSAS, we hope you are now prepared to facilitate the meaningful participation of all students in statewide and district assessments. If so, you understand how the pieces of the accountability puzzle fit together!

References

American Educational Research Association. *Code of Fair Testing Practices in Education.* Washington, DC: American Educational Research Association, 1988.

Witt, J.C., et al. *Assessment of At-Risk and Special Needs Children, 2nd Edition.* Boston: McGraw-Hill, 1998.

Appendixes 6

A. *Standards for Teacher Competence in Educational Assessment of Students*

B. *Calculating the Standard Error of Measurement*

C. *DPI Guidelines for Appropriate Testing Procedures*

D. *DPI Guidelines to Facilitate the Participation of Students with Special Needs in State Assessments 1999-2000*

E. *DPI Information Update Bulletin No. 98.14: Division for Learning Support: Equity and Advocacy Information (Guidelines for Complying with the Assessment Provisions of the Individuals with Disabilities Education Act)*

F. *Code of Fair Testing Practices in Education*

Standards for Teacher Competence in Educational Assessment of Students

Developed by the American Federation of Teachers, National Council on Measurement in Education, and the National Education Association

The professional education associations began working in 1987 to develop standards for teacher competence in student assessment out of concern that the potential educational benefits of student assessments be fully realized. The Committee[1] appointed to this project completed its work in 1990 following reviews of earlier drafts by members of the measurement, teaching, and teacher preparation and certification communities. We encourage parallel committees of affected associations to develop similar statements of qualifications for school administrators, counselors, testing directors, supervisors, and other educators in the near future. These statements are intended to guide the preservice and inservice preparation of educators, the accreditation of preparation programs, and the future certification of all educators.

A standard is defined here as a principle generally accepted by the professional associations responsible for this document. Assessment is defined as the process of obtaining information that is used to make educational decisions about students, to give feedback to the student about his or her progress, strengths, and weaknesses, to judge instructional effectiveness and curricular adequacy, and to inform policy. The various assessment techniques include, but are not limited to, formal and informal observation, qualitative analysis of pupil performance and products, paper-and-pencil tests, oral questioning, and analysis of student records. The assessment competencies included here are the knowledge and skills critical to a teacher's role as educator. It is understood that there are many competencies beyond assessment competencies which teachers must possess.

By establishing standards for teacher competence in student assessment, the associations subscribe to the view that student assessment is an essential part of teaching and that good teaching cannot exist without good student assessment. Training to develop the competencies covered in the standards should be an integral part of preservice preparation. Further, such assessment training should be widely available to practicing teachers through staff development programs at the district and building levels.

The standards are intended for use as:

- a guide for teacher educators as they design and approve programs for teacher preparation,
- a self-assessment guide for teachers in identifying their needs for professional development in student assessment,
- a guide for workshop instructors as they design professional development experiences for inservice, and
- an impetus for educational measurement specialists and teacher trainers to conceptualize student assessment and teacher training in student assessment more broadly than has been the case in the past.

The standards should be incorporated into future teacher training and certification programs. Teachers who have not had the preparation these standards imply should have the opportunity and support to develop these competencies before the standards enter into the evaluation of these teachers.

The Approach Used to Develop the Standards

The members of the associations that supported this work are professional educators involved in teaching, teacher education, and student assessment. Members of these associations are concerned about the inadequacy with which teachers are prepared for assessing the educational progress of their students, and thus sought to address this concern effectively. A committee named by the associations first met in September 1987 and affirmed its commitment to defining standards for teacher preparation in student assessment. The committee then undertook a review of the research literature to identify needs in student assessment, current levels of teacher training in student assessment, areas of teacher activities requiring competence in using assessments, and current levels of teacher competence in student assessment. The members of the committee used their

collective experience and expertise to formulate and then revise statements of important assessment competencies. Drafts of these competencies went through several revisions by the Committee before the standards were released for public review. Comments by reviewers from each of the associations were then used to prepare a final statement.

The Scope of a Teacher's Professional Role and Responsibilities for Student Assessment

There are seven standards in this document. In recognizing the critical need to revitalize classroom assessment, some standards focus on classroom-based competencies. Because of teachers' growing roles in education and policy decisions beyond the classroom, other standards address assessment competencies underlying teacher participation in decisions related to assessment at the school, district, state, and national levels.

The scope of a teacher's professional role and responsibilities for student assessment may be described in terms of the following activities. These activities imply that teachers need competence in student assessment and sufficient time and resources to complete them in a professional manner.

Activities Occurring Prior to Instruction

(a) understanding students' cultural backgrounds, interests, skills, and abilities as they apply across a range of learning domains and/or subject areas;

(b) understanding students' motivations and their interests in specific class content;

(c) clarifying and articulating the performance outcomes expected of pupils; and

(d) planning instruction for individuals or groups of students.

Activities Occurring During Instruction

(a) monitoring pupil progress toward instructional goals;

(b) identifying gains and difficulties pupils are experiencing in learning and performing;

(c) adjusting instruction;

(d) giving contingent, specific, and credible praise and feedback;

(e) motivating students to learn; and

(f) judging the extent of pupil attainment of instructional outcomes.

Activities Occurring After the Appropriate Instructional Segment (e.g., lesson, class, semester, grade)

(a) describing the extent to which each pupil has attained both short- and long-term instructional goals;

(b) communicating strengths and weaknesses based on assessment results to students, and parents or guardians;

(c) recording and reporting assessment results for school-level analysis, evaluation, and decision making;

(d) analyzing assessment information gathered before and during instruction to understand each student's progress to date and to inform future instructional planning;

(e) evaluating the effectiveness of instruction; and

(f) evaluating the effectiveness of the curriculum and materials in use.

Activities Associated with a Teacher's Involvement in School Building and School District Decision Making

(a) serving on a school or district committee examining the school's and district's strengths and weaknesses in the development of its students;

(b) working on the development or selection of assessment methods for school building or school district use;

(c) evaluating school district curriculum; and

(d) other related activities.

Activities Associated With a Teacher's Involvement in a Wider Community of Educators

(a) serving on a state committee asked to develop learning goals and associated assessment methods;

(b) participating in reviews of the appropriateness of district, state, or national student goals and associated assessment methods; and

(c) interpreting the results of state and national student assessment programs.

Each standard that follows is an expectation for assessment knowledge or skill that a teacher should possess in order to perform well in the five

areas just described. As a set, the standards call on teachers to demonstrate skill at selecting, developing, applying, using, communicating, and evaluating student assessment information and student assessment practices. A brief rationale and illustrative behaviors follow each standard.

The standards represent a conceptual framework or scaffolding from which specific skills can be derived. We will need to work to make these standards operational even after they have been published. It is also expected that experience in the application of these standards should lead to their improvement and further development.

Standards for Teacher Competence in Educational Assessment of Students

1 Teachers should be skilled in choosing assessment methods appropriate for instructional decisions.

Skills in choosing appropriate, useful, administratively convenient, technically adequate, and fair assessment methods are prerequisite to good use of information to support instructional decisions. Teachers need to be well-acquainted with the kinds of information provided by a broad range of assessment alternatives and their strengths and weaknesses. In particular, they should be familiar with criteria for evaluating and selecting assessment methods in light of instructional plans.

Teachers who meet this standard will have the conceptual and application skills that follow. They will be able to use the concepts of assessment error and validity when developing or selecting their approaches to classroom assessment of students. They will understand how valid assessment data can support instructional activities such as providing appropriate feedback to students, diagnosing group and individual learning needs, planning for individualized educational programs, motivating students, and evaluating instructional procedures. They will understand how invalid information can affect instructional decisions about students. They will also be able to use and evaluate assessment options available to them, considering among other things, the cultural, social, economic, and language backgrounds of students. They will be aware that different assessment approaches can be incompatible with certain instructional goals and may impact quite differently on their teaching.

Teachers will know, for each assessment approach they use, its appropriateness for making decisions about their pupils. Moreover, teachers will know where to find information about and/or reviews of various assessment methods. Assessment options are diverse and include text- and curriculum-embedded questions and tests, standardized criterion-referenced and norm-referenced tests, oral questioning, spontaneous and structured performance assessments, portfolios, exhibitions, demonstrations, rating scales, writing samples, paper-and-pencil tests, seatwork and homework, peer- and self-assessments, student records, observations, questionnaires, interviews, projects, products, and others' opinions.

2 Teachers should be skilled in developing assessment methods appropriate for instructional decisions.

While teachers often use published or other external assessment tools, the bulk of the assessment information they use for decision making comes from approaches they create and implement. Indeed, the assessment demands of the classroom go well beyond readily available instruments.

Teachers who meet this standard will have the conceptual and application skills that follow. Teachers will be skilled in planning the collection of information that facilitates the decisions they will make. They will know and follow appropriate principles for developing and using assessment methods in their teaching, avoiding common pitfalls in student assessment. Such techniques may include several of the options listed at the end of the first standard. The teacher will select the techniques which are appropriate to the intent of the teacher's instruction.

Teachers meeting this standard will also be skilled in using student data to analyze the quality of each assessment technique they use. Since most teachers do not have access to assessment specialists, they must be prepared to do these analyses themselves.

3 Teachers should be skilled in administering, scoring, and interpreting the results of both externally-produced and teacher-produced assessment methods.

It is not enough that teachers are able to select and develop good assessment methods: they also must be able to apply them properly. Teachers should be skilled in administering, scoring, and interpreting results from diverse assessment methods.

Teachers who meet this standard will have the conceptual and application skills that follow. They will be skilled in interpreting informal and formal teacher-produced assessment results, including pupil's performances in class and on homework assignments. Teachers will be able to use guides for scoring essay questions and projects, stencils for scoring response-choice questions, and scales for rating performance assessments. They will be able to use these in ways that produce consistent results.

Teachers will be able to administer standardized achievement tests and be able to interpret the commonly reported scores: percentile ranks, percentile band scores, standard scores, and grade equivalents. They will have a conceptual understanding of the summary indexes commonly reported with assessment results: measures of central tendency, dispersion, relationships, reliability, and errors of measurement.

Teachers will be able to apply these concepts of score and summary indixes in ways that enhance their use of the assessments that they develop. They will be able to analyze assessment results to identify pupils' strengths and errors. If they get inconsistent results, they will seek other explanations for the discrepancy or other data to attempt to resolve the uncertainty before arriving at a decision. They will be able to use assessment methods in ways that encourage students' educational development and that do not inappropriately increase students' anxiety levels.

4 Teachers should be skilled in using assessment results when making decisions about individual students, planning teaching, developing curriculum, and school improvement.

Educators use assessment results to make educational decisions at several levels: in the classroom about students, in the community about a school and a school district, and in society, generally, about the purposes and outcomes of the educational enterprise. Teachers play a vital role when participating in decision making at each of these levels and must be able to use assessment results effectively.

Teachers who meet this standard will have the conceptual and application skills that follow. They will be able to use accumulated assessment information to organize a sound instructional plan for facilitating students' educational development. When using assessment results to plan and/or evaluate instruction and curriculum, teachers will interpret the results correctly and avoid common misinterpretations, such as basing decisions on scores that lack curriculum validity. They will be informed about the results of local, regional, state, and national assessments and about their appropriate use for pupil, classroom, school, district, state, and national educational improvement.

5 Teachers should be skilled in developing valid pupil grading procedures which use pupil assessments.

Grading students is an important part of professional practice for teachers. Grading is defined as indicating both a student's level of performance and a teacher's valuing of that performance. The principles for using assessments to obtain valid grades are known and teachers should employ them.

Teachers who meet this standard will have the conceptual and application skills that follow. They will be able to devise, implement, and explain a procedure for developing grades composed of marks from various assignments, projects, in-class activities, quizzes, tests, and/or other assessments that they may use. Teachers will understand and be able to articulate why the grades they assign are rational, justified, and fair, acknowledging that such grades reflect their preferences and judgments. Teachers will be able to recognize and to avoid faulty grading procedures such as using grades as punishment. They will be able to evaluate and to modify their grading procedures in order to improve the validity of the interpretations made from them about students' attainments.

6 Teachers should be skilled in communicating assessment results to students, parents, other lay audiences, and other educators.

Teachers must routinely report assessment results to students and to parents or guardians. In addition, they frequently are asked to report or to discuss assessment results with other educators and with diverse lay audiences. If the results are not communicated effectively, they may be misused or not used. To communicate effectively with others on matters of student assessment, teachers must be able to use assessment terminology appropriately and must be able to articulate the meaning, limitations, and implications of assessment results. Furthermore, teachers will sometimes be in a position that will require them to defend their own assessment procedures and their interpretations of them. At other times, teachers may need to help the public to interpret assessment results appropriately.

Teachers who meet this standard will have the conceptual and application skills that follow. Teachers will understand and be able to give appropriate explanations of how the interpretation of student assessments must be moderated by the student's socioeconomic, cultural, language, and other background factors. Teachers will be able to explain that assessment results do not imply that such background factors limit a student's ultimate educational development. They will be able to communicate to students and to their parents or guardians how they may assess the student's educational progress. Teachers will understand and be able to explain the importance of taking measurement errors into account when using assessments to make decisions about individual students. Teachers will be able to explain the limitations of different informal and formal assessment methods. They will be able to explain printed reports of the results of pupil assessments at the classroom, school district, state, and national levels.

7 Teachers should be skilled in recognizing unethical, illegal, and otherwise inappropriate assessment methods and uses of assessment information.

Fairness, the rights of all concerned, and professional ethical behavior must undergird all student assessment activities, from the initial planning for and gathering of information to the interpretation, use, and communication of the results. Teachers must be well-versed in their own ethical and legal responsibilities in assessment. In addition, they should also attempt to have the inappropriate assessment practices of others discontinued whenever they encounter them. Teachers also should participate with the wider educational community in defining the limits of appropriate professional behavior in assessment.

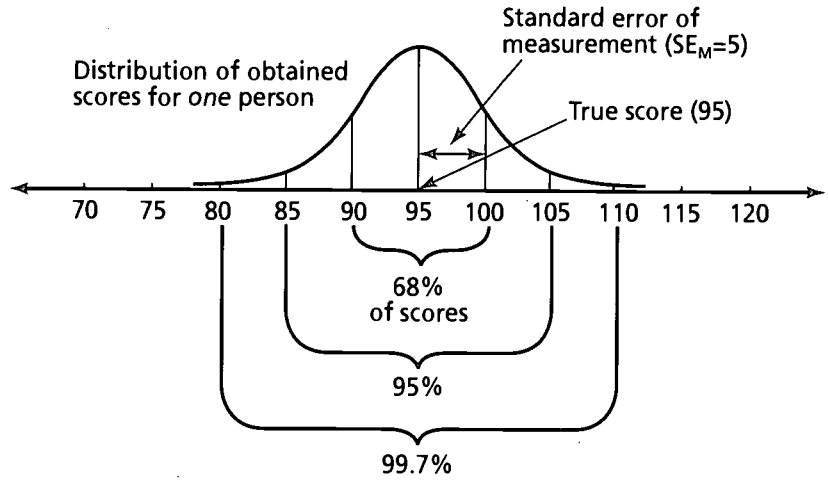
Teachers who meet this standard will have the conceptual and application skills that follow. They will know those laws and case decisions which affect their classroom, school district, and state assessment practices. Teachers will be aware that various assessment procedures can be misused or overused resulting in harmful consequences such as embarrassing students, violating a student's right to confidentiality, and inappropriately using students' standardized achievement test scores to measure teaching effectiveness.

---

[1]The Committee that developed this statement was appointed by the collaborating professional associations: James R. Sanders, (Western Michigan University) chaired the Committee and represented NCME along with John R. Hills (Florida State University) and Anthony J. Nitko (University of Pittsburgh). Jack C. Merwin (University of Minnesota) represented the American Association of Colleges for Teacher Education; Carolyn Trice represented the American Federation of Teachers; and Marcella Dianda and Jeffrey Schneider represented the National Education Association.

Calculating the Standard Error of Measurement

Hypothetical Distribution Illustrating the Standard Error of Measurement



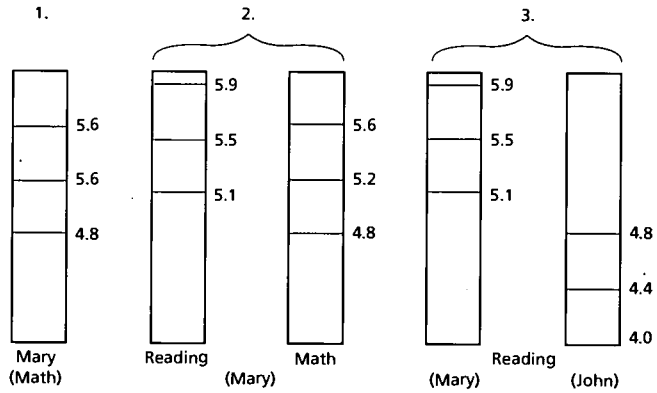
Theoretical Explanation of the Standard Error of Measurement

1. It is assumed that each person has a *true score* on a particular test, a hypothetical value representing a score free of error (true score + 95 on the diagram).
2. If a person could be tested repeatedly (without memory, practice effects, or other changes), the average of the obtained scores would *be approximately normally distributed* around the true score (see diagram).
3. From what is known about the normal distribution curve, approximately 68 percent of the obtained scores would fall within one standard error of measurement of the person's true score; approximately 95 percent of the scores would fall within two standard errors; and approximately 99.7 percent of the scores would fall within three standard errors.
4. Although the true score can never be known, the standard error of measurement can be applied to a person's obtained score to set "reasonable limits" for locating the true score (e.g., an obtained score of 97 ± 5 = 92 to 102).
5. These "reasonable limits" provide *confidence bands* for interpreting an obtained score. When the standard error of measurement is small, the confidence band is narrow (indicating high reliability), and thus we have greater confidence that the obtained score is near the true score.

GUIDELINES

Practical Applications of the Standard Error of Measurement in Test Interpretation

A confidence band one standard error above and below the obtained score is commonly used in test profiles to aid in interpreting individual scores and in judging whether differences between scores are likely to be "real differences" or differences caused by chance.



1. *Interpreting an individual score.* The confidence band indicates "reasonable limits" within which to locate the true score (Mary's math score probably falls somewhere between 4.8 and 5.6).
2. *Interpreting the difference between two scores from a test battery.* When the ends of the bands overlap, there is no "real difference" between scores (Mary's scores in reading and math show no meaningful difference).
3. *Interpreting the difference between the scores of two individuals on the same test.* When ends of bands do not overlap, there is a "real difference" between scores (Mary's reading score is higher than John's).

DPI Guidelines for Appropriate Testing Procedures

Wisconsin Student Assessment System
Knowledge and Concepts Examinations (WKCE)
at Grades 4, 8, and 10

December 1998

INTRODUCTION

Appropriate and ethical testing practices are not always universally understood or followed. People sometimes violate good testing practices because they are not informed about what is appropriate in testing. To help school staff who administer the Wisconsin Student Assessment System (WSAS) Knowledge and Concepts Examinations (WKCE) at elementary (grade 4), middle (grade 8), and high school (grade 10), have a common understanding of appropriate practices, the Department of Public Instruction has prepared these guidelines.

This paper covers general principles and standards as presented by a number of organizations and studies and is applicable to all types of assessments, including standardized multiple-choice tests as well as performance assessments. Topics covered in this paper include:

A. Test Security
B. Ethics in Testing
C. Testing Conditions
D. Post-Test Activities/Procedures

A. TEST SECURITY

February 1999 Memo from Office of Educational Accountability
Test Security Agreement

1. What Is Meant by Test Security?

Tests used in the WKCE statewide testing program are secure, proprietary instruments published and copyrighted by a testing company, not the state. Any disclosure or dissemination of actual test items to any person may be considered a copyright violation and may severely undermine the value of the test and adversely affect the validity of test results. The confidentiality of test questions and answers is paramount in maintaining the integrity and validity of the test. Therefore, the Department of Public Instruction (DPI) and all Wisconsin educators must take every step to assure the security of the test instruments.

2. Why Is Test Security Important?

Test security is important to:

- make valid inferences on student and school performances as required by law;
- guard against limiting the curriculum to content covered on the test;
- give accurate measures of students' abilities; and
- keep the integrity of the test and testing situation intact.

3. Who is Responsible for Test Security?

Everyone who works with assessment, communicates test results to others, and/or receives testing information is responsible for test security. This includes:

- Staff of the Department of Public Instruction
- District Administrators and both certified and non-certified school staff
- District Assessment Coordinators (DACs)
- School Assessment Coordinators (SACs)
- Students, parents, and the community at large
- Staff at the Cooperative Educational Service Agencies (CESAs)

4. Can District Staff Review the Test?

School and district staff should be familiar with the testing procedures and schedule before testing. However, district and school staff may arrange a review of the test instrument only after the test administration has concluded.

Test review by staff will help familiarize them with the test content and format and assist them in understanding and using test results as well as in their curriculum development efforts. When reviewing the test content with the staff, the DAC/SAC must make sure that all test books are numbered before distribution. At the end of the review session, staff must collect and account for all test books. No staff member should be allowed to leave the review session until all test books are counted and secured.

The DAC/SAC must take the following precautionary steps at the time of test review with the staff:

- Use actual test books. DO NOT use any reproduced copies of the test books.
- Number all test books before entering the review session.
- Ask staff to sign a security agreement acknowledging ethical practices, copyright, and proprietary restrictions before beginning the test review session.
- Distribute the numbered copies of the test books to staff at the review session.
- Work with staff as a group; do not allow individuals to retain a copy of the test.
- **DO NOT** allow any staff member to make copies of the test or any test items, take notes, or otherwise reproduce the test or test items.
- Concentrate on the review of the Objectives Performance Report and the Item Analysis Summary Report when discussing test content.
- **DO NOT** allow any individual to leave the test review session before all numbered test books are collected and accounted for.

5. Is the Public Allowed to Review the Tests?

Review of the test by parents and other private citizens must be a guarded matter. It must follow a formal security procedure. Any interested person may request to review the test but only after the test administration has concluded. Wisconsin law stipulates that "The state superintendent shall make available upon request, within 90 days after the date of administration, any examination required to be administered under this subsection [s. 118.30, Wis. Stats.]."

Tests administered under WSAS are copyrighted, secured instruments. They are the product of a costly contract with the state. The DAC must coordinate the public review of the test instrument in a way that does not compromise the integrity and security of the test, test items, and test results.

The following are some of the safeguards the DAC/SAC should take when a reviewer reviews the tests:

- Have the reviewer sign a confidentiality agreement prior to reviewing the test.
- **DO NOT** permit the reviewer to make any photocopies or other reproductions of the test, take notes, or copy test items.
- Allow the reviewer to review only materials that **DO NOT** identify individual students.
- **DO NOT** allow the reviewer to review material that identifies individual students.
- Be prepared, or secure the availability of a qualified staff member, to explain the purpose of the test, answer questions about test content, and explain the meaning of test results to the reviewer.
- The reviewer must be accompanied at all times and should not be left alone with the test.

6. Pre-Test Security
It is essential that all test materials remain secure. That is, when the tests are not being used for testing, testing materials should be kept in a LOCKED STORAGE area. Access to these materials only should occur with the knowledge and expressed permission of the DAC.

District Assessment Coordinators should work closely with school assessment coordinators, and together they direct the management of WKCE. Their number one responsibility is to ensure test security throughout the testing process in order to protect not only the integrity of the test, but also to protect principals and teachers from any appearance of impropriety. On a daily basis, DACs and SACs should make sure that all test materials are placed in locked storage when not in use in a testing session. They also must make sure that students do not share information about test content when the test is administered to same grade-level students at different times. If they note any deviation, they should take immediate action to correct it. Depending on the severity of a deviation in security, it may be necessary to advise the Department's Office of Educational Accountability.

7. What Are Some Examples of Test Security Violations?
Educators, students, or others can commit test security violations. Some examples of test security violations by **educators** include, but are not limited, to:

- leaving students unsupervised during testing;
- leaving test materials in an unsecured place;
- photocopying or keeping a personal copy of the test;
- taking notes about test questions and using them or a close paraphrase to prepare students for testing;
- offering "hints" that indicate an answer or help eliminate answer choices;
- rephrasing the test questions;
- editing (changing) student answers after completion of the test by erasing any wrong answers and writing in the correct ones;
- extending testing time beyond regulations for students other than those with documented disabilities per their IEP or for students covered by Sec. 504 per their IAP or for certain limited-English proficiency (LEP) students;
- providing test accommodations for students with disabilities that are not included in the student's IEP or IAP that are not specified in accommodations offered to an LEP student;
- allowing students to go back to previous sections in the test booklet to check their work; and
- allowing students to go back to the current section in the test booklet to change their answers after allowed testing time has expired.

Some examples of test security violations by **students** include, but are not limited to:

- illegally obtaining a test booklet to study or to let others study;
- securing a marked test booklet or "crib sheet" from a teacher or another student;
- copying or "stealing" answers from another student during testing;
- sharing specific test information with other students in the same grade who are scheduled to take the same test at a later time; and
- taking a test during the make-up period and asking a student who has already completed the test to disclose test questions and/or answers.

8. Possible Consequences/Sanctions for Compromising Test Security
Administrators, certified and non-certified school staff, students, and parents must adhere to ethical procedures in testing. The local school board, the Department, and/or the court system can investigate violation of these procedures and take appropriate sanctions.

The school faculty, conscientious students, their parents or other family members, and persons in the community may report test security violations. Erasure analysis, unusual score gains, or other irregularities also may detect test security violations.

Potential sanctions for **educator** violation of security measures may include:

- suspension or acceptance of voluntary surrender of license;
- suspension without pay or a written reprimand;

- termination of contract, acceptance of resignation, or retirement;
- civil legal liability for copyright violations;
- legal prosecution;
- public embarrassment; and
- others as determined by local school boards.

Potential sanctions for **student** violation of security measures may include:

- invalidation of test results;
- invalidation of specific test questions or subtests, or invalidation of pass/fail proficiency results;
- suspension or expulsion from school or other disciplinary actions according to the local code of conduct;
- suspension or exclusion from participating in school extra-curricular activities, such as sports, plays, school-sponsored social functions, etc., as dictated by the local school board policies;
- denial of appointment to a school membership team, such as mathematics or debate team;
- removal from an elected office, such as president of the student council, etc., as dictated by the local school board policies; and
- others as determined by local school boards.

B. ETHICS IN TESTING

Aside from security issues, the most significant consideration for appropriate and ethical testing practices in pretest activities relates to preparing students for the test in ways that allow for a valid interpretation of the test results. A WKCE test score is an estimate of the student achievement in the content areas of mathematics, science, social studies, English language arts, and reading.

It is important to be reasonably certain that if the student has done well on the WKCE, she or he understands the content sufficiently to perform well on similar tests and to apply that understanding (knowledge) in real life. If a student is coached or taught only the content specific to a given test, his or her scores may not be valid indicators of what the student knows and can do. The result simply will be a measure of how well the student has been taught the specific content on the test.

1. Wisconsin Model Academic Standards and Test Alignment

The WKCE allows students to demonstrate their knowledge and skills using selected-response and constructed-response items in one test instrument. They also include performance assessment in the form of a writing essay. In April 1998, CTB/McGraw-Hill, the Department's contractor for WKCE, conducted a workshop for the Department to match items on the TerraNova (WKCE) to the Wisconsin Model Academic Standards in Reading, Language Arts, Mathematics, Science, and Social Studies, adopted in January 1998 by an Executive Order issued by the Governor. The purpose of conducting the match was to determine whether individual TerraNova test items assess the Model Academic Standards and the extent to which the standards are addressed by TerraNova items. One may obtain a copy of the alignment findings by contacting the Department's Office of Educational Accountability.

It is important to note that no single assessment can measure domains as large as those identified in the Wisconsin Model Academic Standards or in locally adopted standards. The WKCE measures only part of what students need to know. The Department supports district teaching efforts that focus on the breadth and depth of materials encompassed by state and local content standards rather than efforts that are solely and narrowly focused on the items or content of the WKCE.

2. Classroom Instruction and the WKCE Content

Students may receive instruction, experience, and practice in the objectives that the WKCE samples. These objectives have been widely distributed to district and school staff in Wisconsin. However, this does not mean that a school or district should narrow its curriculum to fit the objectives covered by the examinations or that teachers should focus mainly on these objectives. Teachers should cover these objectives along with many other objectives that are in the curriculum but not measured on the test. The WKCE spans the content taught over several academic years and are not the sole responsibility of a particular grade-level teacher.

3. Preparing Teachers to Administer the Test

Teachers should carry out test administration procedures in a way that is consistent with prescribed, standardized procedures in order to give every student an equal opportunity to succeed and to allow for making valid inferences and interpretation of test results. When conducting test administration training sessions, the DAC/SAC should rely entirely on information found in the Directions for Test Administration manual, provided as part of the WKCE test materials. The DAC/SAC should not use actual test books when training the staff.

The responsibilities of the test administrator are to:

a) be familiar with test administration directions before entering any testing session;
b) plan for the distribution and collection of materials;
c) plan student seating arrangement, making sure that spacing between students prevents them from sharing answers;
d) adhere strictly to standardized testing procedures;
e) for students with disabilities under IDEA or students covered by Sec. 504, follow the accommodation provisions outlined in the student's IEP or IAP and provide accommodations for students with Limited English Proficiency (LEP) consistent with their LEP status, as described in the DPI Guidelines to Facilitate the Participation of Students with Special Needs in State Assessments;
f) ensure that adequate and complete sets of materials are available to all students;
g) provide an adequate testing environment, free from interruption and public address announcements;
h) schedule make-up sessions for absentees; and
i) ensure all security procedures are followed at all times.

4. Preparing Students to Take the Test

THE GOAL OF TEACHING IS TO INCREASE LEARNING RATHER THAN TO INCREASE TEST SCORES. Therefore, teachers should direct students' attention and effort to learning the entire scope of the curriculum, not just the limited knowledge and skills measured by the WKCE.

The Department does not encourage school staff to buy, develop, or promote the usė of extensive test practice materials that closely parallel the WKCE's items or tasks. Staff must adhere to the following ethical test preparation procedures:

- Student learning should cover the entire scope of the curriculum. Teaching students the entire subject domain is ethical; teaching to the test is not.
- Students may have one or two short practice sessions to familiarize them with the test format one or two days before the administration of the actual test. A Practice Activities Test is provided and ought to be administered to fourth-grade students. Additional practice items for fourth-grade students and test practice items for students in eighth and tenth grades may be found in the 1997-98 Wisconsin School Performance Report Results for Districts and Schools Within Districts, Volumes I, II, and III, as well as in the Teacher's Guide to TerraNova. Sample WKCE test items are also included in our website at [http://www.dpi.state.wi.us/oea/profitem.html].
- Instruction, experience, and practice should not be limited to the content that the examinations will sample.
- A reasonable notice of the upcoming examination schedule should be given to students, teachers, and parents.
- All students must be encouraged and prepared to participate in the examination. All eligible students must participate in the testing.

5. Reasonable Notice and Full Participation

All concerned, including teachers, students, and parents, must receive reasonable notice of the WKCE. However, this notice should not be used to discourage students from participating in the assessment, particularly if these students are members of groups whose test scores have been historically low. Educators should plan adequate make-up sessions so that all students have an opportunity to fully participate in the assessment.

Educators should make participation plans in WKCE for students with disabilities under IDEA on an individual basis and specify the plan in the child's Individualized Educational Program (IEP).

Although all students with disabilities under IDEA also are covered by Section 504 of the Vocational Rehabilitation Act, there are a limited number of students who are not considered students with disabilities under IDEA but who are covered by Section 504. Students qualifying only under Sec. 504 must receive the necessary individual accommodations in testing as specified in the student's Sec. 504 Individualized Accommodation Plan (IAP). It is possible, although extremely rare, that some of these Sec. 504 students may not participate in the tests.

Teachers must base decisions regarding inclusion of limited-English proficient (LEP) students on s. PI 16.01, Wis. Admin. Code. Qualified school staff shall determine, on an individual basis, whether an LEP student will participate in the assessment and will specify the type of accommodations they will provide, if necessary. Students who fall in categories 4 and 5, based on s. PI 13.03 (3), Wis. Admin. Code criteria, may be included in the assessment. DPI encourages the inclusion of these students in the assessment and teachers should justify their exclusion in writing.

Students with disabilities under IDEA who are appropriately excluded from the WKCE should complete alternate assessments. DPI published guidance concerning these alternate assessments in October 1998: *Division for Learning Support: Equity and Advocacy Information Update Bulletin No. 98.14.*

DPI Guidelines to Facilitate the Participation of Students with Special Needs in State Assessments contain further information concerning testing of all students with special needs, including students with disabilities, LEP students, and students covered by Sec. 504 or the Vocational Rehabilitation Act of 1973. The Department will provide additional guidance concerning alternate assessments. If a student's parents request their school board to excuse the student from taking the WKCE, the school board must excuse the student from testing.

Providing all students and their parents/guardians with a copy of the Student/Parent Pre-Test Guide prior to the administration of WKCE will ensure reasonable notice.

6. Effect of Ethics on Test Results

Although not acceptable, it is very tempting for some school staff to teach too closely to the actual test questions in order to achieve high test scores. Temptations increase in a testing situation where the stakes are high and where sanctions may be attached to test results. However, a test, no matter how well designed, only can measure a small part of the overall curriculum.

The inferences made about a student who does well on the WKCE indicates that the student has learned the larger domain from which that test content has been sampled, not solely the content knowledge included in the test. If teachers limit their classroom instruction to skills measured on the test, they have violated this assumption and, therefore, can consider students proficient only in the particular skills covered on the specific test. These students may not do well on other questions or tasks covering the larger domain.

Similarly, it is very important to precisely follow standardized test administration procedures. If the WKCE are administered in inappropriate, non-standardized ways, the results will not be comparable to those produced under standardized testing conditions. For example, if a teacher helps students by paraphrasing and explaining the test items and another teacher adheres strictly to the guidelines by repeating the initial instructions, the scores of students in the two classrooms cannot be interpreted in the same manner. In either case, this would result in an inaccurate representation of student learning in Wisconsin.

Strict adherence by school staff to the test standardization procedures and to the guidelines presented in this paper will ensure that the test results are accurate and reflect student learning in our state.

97

BEST COPY AVAILABLE

C. TESTING CONDITIONS

1. Testing Procedures

Test administrators must strictly follow the written test administration procedures included in the Directions for Test Administration, which is provided to districts as part of the WKCE materials. These procedures include planning for the test, organizing the classroom, preparing students to take the test, completing student-identification information, timing of testing sessions, reading instructions to students, and collecting test booklets after each testing session. Failure to follow the specified procedures jeopardizes the validity and integrity of the test results.

2. Testing Environment

Testing conditions should be comfortable and similar for all students. To the extent possible, the conditions should reflect the school's instructional environment. School Assessment Coordinators and test administrators must ensure that announcements are not made on the public address system during testing sessions, lighting is adequate, chairs and desks are available, and "QUIET" signs are posted. This will permit students to do their best work. It is recommended that teachers conduct the testing session in small groups of classroom size, rather than in a large group and auditorium-type hall. This will help students in their concentration since instruction normally is given in smaller, class-size groups.

3. Testing Materials

Before students begin taking the test, test administrators must ensure that adequate and complete sets of test materials are available to all students, including test booklets, pencils, calculators, and manipulatives such as rulers, protractors, punch-out tools, and geometric shapes, as required.

4. Test Administration

- Test Directions

  Test administrators must be completely prepared and familiar with the test directions before entering any testing session. Administrators should anticipate and be ready to answer questions about the test. When reading test directions aloud, test administrators must ensure that all students understand what is expected of them. Students must have the opportunity to ask questions and understand how to mark their answers before they begin taking the test. However, test administrators MUST NOT answer questions about specific test items. They may only repeat the initial instructions about item format, scoring rules, and timing. They may also help students with test-taking mechanics but must be careful not to inadvertently give clues that indicate the correct answer or help eliminate some answer choices.

- Special Populations

  The Department is committed to including ALL students in testing. Special population students must participate in the WKCE and, when necessary, receive any necessary accommodations to ensure their participation. The majority of these students require minor or no accommodations. Accommodations in assessment for students with disabilities under IDEA, LEP students, and students covered by Sec. 504 of the Vocational Rehabilitation Act of 1973 should reflect the accommodations used in classroom instruction.

  In those cases where a student with disabilities under IDEA, even with accommodations, would be unable to demonstrate at least some of the knowledge and skills tested in WKCE, teachers must provide an alternate assessment to measure the student's performance. The Department of Public Instruction published guidance concerning these alternate assessments in October 1998: *Division for Learning Support: Equity and Advocacy Information Update Bulletin No. 98.14.* The Department will provide additional guidance concerning alternate assessments.

  The Individuals with Disabilities Education Act (IDEA) requires that children with disabilities be "included in general state and districtwide assessment programs, with appropriate accommodations, where necessary." It also requires that the state report "to the public with the same frequency and in the same detail as it reports on the assessment of nondisabled children" specific information about the participation of children with disabilities in assessment and their performance on

the assessment. The federal government will monitor the extent to which the state complies with these requirements.

Federal and state law require students with disabilities under IDEA to participate in the assessment program. Teachers must assign a grade level to students with disabilities. Generally, it would be appropriate for the district to use age-based guidelines. These guidelines should allow for some flexibility, recognizing that there is a range of student ages within any grade. There should not be so much flexibility, however, that it is possible to defeat the purpose of requiring student participation in the assessment program.

The Department recognizes that there will be some instances in which the nature or severity of the disability of a student under IDEA or the English proficiency of an LEP student may necessitate his/her participation in the statewide assessment system through an alternate assessment. Since it is the intent of the Department to include ALL students in the WKCE, educators should make the decision to exclude any student from testing only after careful evaluation of each student's ability and with written justification.

Students covered by Sec. 504 of the Vocational Rehabilitation Act of 1973, must receive the necessary accommodations in testing as specified in the student's Sec. 504 Individualized Accommodation Plan (IAP).

Title I students and students who receive free or reduced-price lunches are not, by such definitions alone, students with disabilities. Unless they also are identified as students with disabilities under IDEA, students covered by Sec. 504, or LEP students, testing students served by Title I should not receive accommodations. In such cases, the IEP, IAP, or LEP status will determine the type of accommodations needed.

5. Monitoring/Proctoring the Test Session

Test administrators must carefully monitor (proctor) the testing session to ensure that all students have the opportunity to succeed. It is not acceptable for test administrators to leave the room, visit with another person, read, or ignore what is happening in the testing session. Test administrators and proctors must be trained to follow the testing procedures and to understand the significance of their responsibilities. Test administrators must:

a) study the Directions for Test Administration manual thoroughly and be prepared to answer questions;
b) follow standardized test administration instructions, adhering strictly to standardized procedures and follow the written script, verbatim, without adding or deleting information;
c) ensure that all students understand the directions;
d) be sure that students know how to mark their answers and help students in their test-taking mechanics without inadvertently giving them hints that indicate an answer or help eliminate answer choices;
e) encourage students not to spend too much time on any test item—be careful not to imply that they should guess randomly, but tell them that if they do not know the correct answer they can eliminate some of the choices to help them find the correct answer;
f) encourage students to attempt to answer every item on the test and monitor their work to ensure that they do not skip or overlook any of the test questions;
g) ensure that students respond in the appropriate places in the test booklet;
h) direct students to mark only one response for each selected-response item and ask them to erase completely any responses they do not want;
i) ensure that students do not exchange or copy answers from each other;
j) encourage students who complete the test before regulation time to review their answers;
k) ensure that students are not disruptive and do not interfere with or distract each other;
l) ensure that students use only permitted test materials and devices;
m) follow the provisions of the IEP for students with disabilities under IDEA;
n) follow the provisions of the IAP for students covered under Sec. 504;
o) follow Department's guidelines for testing of LEP students;
p) encourage students to do their best on the test and to check their work;
q) do not engage in conversation with other staff while the testing session is in progress;

r) collect and check all materials when the testing session has concluded;
s) write a report about all deviations, irregularities, and anomalies that may have compromised the testing situation and give to your school principal or school assessment coordinator; and
t) keep track of absent students and plan make-up sessions.

D. POST-TEST ACTIVITIES/PROCEDURES

1. Collecting Test Materials and Completing the Report
When the testing session has concluded, the test administrator will collect and check all materials and follow test security procedures. The test administrator must account for all test materials and deliver them to the SAC immediately.

The test administrator must write a report of all incidents and events that may have compromised the testing situation and could have the potential of invalidating test scores. This includes disruptions, illnesses, cheating, refusal of students to complete the test, etc. Test administrators must submit the report to the school principal and/or school assessment coordinator who, in turn, will take it to the district assessment coordinator and/or the district administrator.

2. Use of Test Information
School and district staff must follow strict confidentiality measures to protect individual student test scores and maintain student privacy, as required by federal, state, and local laws. Only authorized personnel, i.e., the student, the student's parents or legal guardians, and the specific staff responsible for the student's education should have access to the student's scores.

Schools must make sure that test interpretation guides are provided to students and their parents and should use student scores in context with other relevant information about that student.

3. Making the Test Instruments Available to the Public
The Department will comply with the requirement of s. 118.30(3), Wis. Stats., by making the WKCE available to the public upon request, within 90 days after the date of administration. However, we will conduct such availability under limited and controlled conditions and we will base it on the ability of school staff to assure the security of the test contents and answers, confidentiality of individual student's test results, and capability of explaining the purpose of the test and the meaning of test results. These conditions include that the:

a) school district has filed the "Confidentiality Agreement" form, signed by the district administrator, with the Department prior to meeting with any member from the public for the purpose of viewing the WKCE;
b) school district will have the viewer sign a confidentiality agreement form, provided by the Department and printed on district letterhead, prior to viewing the test;
c) the test should be disclosed to a member of the public only upon request;
d) viewer will be permitted to view the test and test results without any violation of student privacy or jeopardizing student confidentiality issues;
e) viewer will be accompanied by a qualified school staff member at all times;
f) viewer will NOT be allowed to copy or take notes on any portion of the test; and
g) school district will have a qualified staff member available to explain the proper use of the test, the purposes of the WSAS program, and the meaning of test results.

Refer to Section A, Item 5, page 2, of this paper for more information on the public review of the WSAS Knowledge and Concepts Examinations.

Submit questions regarding the OEA to: oeamail@dpi.state.wi.us. Submit questions regarding the WKCE to: rajah.farah@dpi.state.wi.us.

Copyright: State of Wisconsin Department of Public Instruction
Phone: 1-800-441-4563 (U.S. only) / 608-266-3390
Submit questions or comments regarding this website to: webmaster@www.dpi.state.wi.us
Last Modified September 08, 1999

DPI Guidelines to Facilitate the Participation of Students with Special Needs in State Assessments 1999-2000

Related Websites, Bulletins, and Notices:

- Students with Disabilities and Statewide Assessment (DPI Special Education Team Web Site)
- Guidelines for Complying with the Assessment Provisions of the Individuals with Disabilities Education Act (Special Education Information Update Bulletin No. 98.14)
- Participation of Students with Special Needs in State Assessments (A powerpoint presentation for local training purposes, Spring 1999)

Introduction

Wisconsin has published academic content, performance, and proficiency standards for ALL students in the state. The Improving America's Schools Act (IASA) of 1994 requires states to administer high-quality student assessments that are aligned with the state's academic standards and provide coherent information about students' attainment of such standards. Wisconsin's academic standards are for all students, including students with special needs (students with limited English proficiency under Wis. Stats. s. 115.955(7) and Title VII of IASA, students with disabilities under Subchapter V of Wis. Stats. 115 and the Individuals with Disabilities Education Act (IDEA), and students covered by Sec. 504 of the Vocational Rehabilitation Act of 1973). Students with special needs must receive the same opportunity to acquire and demonstrate their academic performance as students without special needs.

In the state of Wisconsin, one way that students demonstrate their progress toward achieving the academic standards in reading, math, language arts, social studies, and science is via the Wisconsin Student Assessment System (WSAS). At present the WSAS includes the Wisconsin Reading Comprehension Test (WRCT) at third grade and the Wisconsin Knowledge and Concepts Examinations (WKCE) at fourth, eighth, and tenth grades. The purpose of this document is to provide guidelines for facilitating the participation of students with special needs in WSAS assessments (WRCT and WKCE).

As such, we intend this document to update and replace previously published DPI guidelines regarding the participation of students with special needs in the WRCT and WKCE.

Although the rationale for participating in WSAS assessments is the same for all students with special needs, there are different laws which affect participation decisions for each group. Thus, we present considerations for students with limited English proficiency first, followed by a discussion of considerations for students with disabilities under the IDEA and students receiving accommodations under Section 504 of the Vocational Rehabilitation Act of 1973.

Students with Limited English Proficiency (LEP)[1]

The IASA requires that state assessments allow for "the inclusion of limited English proficient students who shall be assessed, to the extent practicable, in the language and form most likely to yield accurate and reliable information on what such students know and can do, to determine such students' mastery of skill in subjects other than English" [Part A, subpart 1, sec. 1111 (b) (3) (F) (iii); 20 USC ss 6311 (b) (3) (F) (iii)]. Students with LEP should participate in the WRCT or WKCE as soon as they achieve an English proficiency level that allows them to demonstrate their knowledge and skills on these tests. The translation of large scale assessments into all of the languages spoken by students with LEP in Wisconsin is not viable. Thus, local alternate assessment offers the "best practices" solu-

tion for full inclusion of students with LEP at the early English proficiency levels. To be used for state accountability purposes, IASA requires that alternate assessments be aligned with state academic standards.

Decisions regarding participation in locally-developed, standards-based alternate assessments for individual students with limited English proficiency must be consistent with the federal IASA legislation and specifically based on PI 16.01, Wis. Admin. Code. PI 16 requires districts to adopt a policy establishing procedures for testing students with LEP, procedures for notifying parents of students with LEP, and any district-specific criteria used to determine participation of students with LEP in WKCE assessments or alternate assessment. For students with LEP, participation in the WKCE is not an "all or nothing" decision. Instead, there are multiple alternatives for facilitating the participation of a student with LEP. These alternatives reflect three broad options:

1. participation in the WKCE without accommodations, i.e., changes in the administration or format of the test that do not alter the test content or intent of the test,
2. participation in the WKCE with accommodations, or
3. participation in locally-developed, standards-based alternate assessments.

For the WKCE, teachers may use these options exclusively or in combination depending on the individual needs of the student. That is, they must make separate decisions regarding need for accommodations or alternate assessment for each content domain included in the test. For example, some students with LEP may not require any accommodations to participate in the WKCE. Other students with LEP, however, may need accommodations for some of the content domains, i.e., math, reading/language arts, science, or social studies, but not for others. Still other students may need accommodations for some areas within the WKCE and alternate assessment for one or more content domains. Finally, there will be a limited number of students with LEP for whom the WKCE will not be appropriate, and teachers will assess these students through alternate assessment only. It is important to note that students participating in alternate assessment are coded as "excluded" when WKCE data are reported to the state. All students coded as "excluded" from any WKCE content-area test are expected to take an alternate assessment in that content area.

For local educators to determine which of the above options is most appropriate for each student with LEP, a thorough, individualized English language proficiency assessment must first be conducted by qualified school staff (see PI 13.03, Wis. Admin. Code). This assessment should include reading, writing, speaking, and listening. The results of this assessment should be compared to the definition of English language proficiency levels recommended by the State Superintendent's Advisory Council on Bilingual/ESL Education [based on PI 13.03 (3) (a)-(e)]. This definition uses five levels of limited English proficiency.

Based on the requirements of PI 16 and the LEP definitions found in PI 13.03, *students with LEP do not participate in the WKCE if their English proficiency level is one, two or three (beginning through intermediate). DPI recommends that students at English proficiency levels four and five participate in all WKCE content domains, with appropriate accommodations.*

For the WRCT, DPI requires that students at English proficiency level one, two, three, and four (beginning through advanced intermediate) not participate in the test. Those students at English proficiency level five (advanced) are required to participate in the WRCT. It is important to note that, because the WRCT assesses specific language-based skills (reading comprehension) and is administered in an untimed format, the WRCT must be administered without accommodations to students at level five.

Students who reach "full English proficiency" are no longer students with LEP and should not be classified as LEP. These students must participate in WRCT or WKCE. They may not receive accommodations because a need for accommodations contradicts the definition of "fully English proficient."

Students with Disabilities as Defined Under the Individuals with Disabilities Education Act (IDEA)

The 1997 reauthorization of the Individuals with Disabilities Education Act (IDEA) and s.115.77, Stats., require participation of students with disabilities in state and districtwide assessments. Specifically, the IDEA stipulates, "children with disabilities are included in general state and districtwide assessment programs with accommodations, where necessary." In addition, the IDEA and s.115.787, Stats., require that children with disabilities for whom the standard state assessment is inappropriate receive alternate assessments. Several state and national reviews concerning alternate assessment suggest that approximately 10 to 20% of students with disabilities, or 1 to 2% of the total student population, will be assessed via an alternate assessment.

The student's IEP team, which includes the parent as an equal participant, must address all questions regarding the participation of a student with disabilities in general state and districtwide assessments. State and federal special education laws require that a student's IEP include "a statement of any individual modifications in the administration of state or districtwide assessments of student achievement that are needed in order for the child to participate in such assessment; and if the IEP team determines that the child will not participate in a particular state or districtwide assessment of student achievement (or part of such an assessment), a statement of why that assessment is not appropriate for the child; and how the child will be assessed." To make these determinations, the IEP team must be knowledgeable about the child's present level of educational performance and measurable annual goals, the general curriculum, the format and content of the state or district test, and the alignment between the curriculum and the academic content standards assessed by the state or districtwide assessment system.

Participation in the WRCT or WKCE for students with disabilities is not an "all or nothing" decision. Instead, there are multiple alternatives for facilitating the participation of a student with a disability. These alternatives reflect three broad options:

1. participation in WRCT or WKCE without accommodations, i.e., changes in the administration or format of the test that do not alter the test content or intent of the test,
2. participation in the WRCT or WKCE with accommodations, or
3. participation in locally-developed, standards-based alternate assessments.

For the WRCT, these options are mutually exclusive because the test only assesses one content domain (reading comprehension). For the WKCE, however, teachers may use these options exclusively or in combination depending on the individual needs of the student. That is, they must make separate decisions regarding need for accommodations or alternate assessment for each content domain included in the test. For example, some students with disabilities may not require any accommodations to participate in the WKCE. Other students with disabilities, however, may need accommodations for some of the content domains, i.e., math, reading/language arts, science, or social studies, but not for others. Still other students may need accommodations for some areas within the WKCE and alternate assessment for one or more content domains. Finally, there will be a limited number of students with disabilities for whom the WKCE will not be appropriate, and teachers will assess the performance of these students through an alternate assessment only. It is important to note that students participating in alternate assessment are coded as "excluded" when WSAS data are reported to the state. All students coded as "excluded" from any WKCE content-area test must take an alternate assessment in that content area.

The IEP team must make decisions regarding student participation in state assessment on an individual basis. As a result, the team bases this decision on a thorough review of child-specific data to assess the student's current educational performance relative to the academic performance standards for ALL students. This thorough review includes consideration of existing student records, including the most recent evaluation data, formal and informal evaluations conducted by team members, reports by parents and special education and/or general education teachers, classroom work samples, indepen-

dent educational evaluations, and any other information available to the IEP team. To make appropriate decisions regarding the student's need for accommodation and/or alternate assessment, the IEP team should consider the following:

1. Begin with the assumption that all students with disabilities will participate in the WRCT. For the WKCE, assume that the student will participate in all content domains, i.e., reading/language arts, math, science, and social studies.

2. Assess need for accommodation and/or alternate assessment based on the student's present level of educational performance, IEP goals, and the content and format of the WRCT or WKCE. For the WKCE, assessment of need for accommodation should be conducted independently for each content domain.

3. Consider the accommodations that the child receives in classroom assessments as possible accommodations for the WRCT or WKCE.

4. Select accommodations that do not invalidate the test, i.e., change the skills or content tested. If the necessary accommodations would invalidate the test, assess the student's knowledge and skills through alternate assessment. For example, an accommodation that included reading passages and/or items aloud to students would not be an acceptable accommodation if the purpose of the assessment is to measure reading skills. Thus, a student who would require this accommodation should participate in an alternate assessment for the WRCT or the reading/ language arts test of the WKCE.

5. Allow for alternate assessment only if a student would not be able to demonstrate some of the knowledge and skills on the WRCT or WKCE assessment *with appropriate accommodations.*

Based on the thorough review of the student's current educational performance relative to the academic standards, the IEP team determines how a child with a disability will participate in the WRCT or WKCE assessment. For those students who are identified as needing accommodations on the WRCT or WKCE assessment, the IEP team must specify which accommodations are necessary for the child to participate in the assessment.

The IEP team may determine that, even with accommodations, a child with a disability would be unable to demonstrate at least some of the knowledge and skills tested through the standardized assessment, and, as a result, they will assess the student's performance through alternate assessment. The thorough review undertaken to reach this decision can function as an alternate assessment if it is documented as part of the IEP process. It is important to note that to serve as an alternate assessment, the review must be recent, reliable, and representative of the student's present level of educational performance relative to the academic standards. In addition, to qualify as an alternate assessment the IEP team must conduct the review within a time frame that approximates the administration of the statewide standardized assessment. (The DPI suggests holding the IEP review three or four months prior to the administration of the WRCT or WKCE.) Additional information regarding the DPI's position on alternate assessment for children with disabilities under the IDEA is in DPI Bulletin 98.14. at http://www.dpi.state.wi.us/dpi/dlsea/een/bul98-14.html.

Students Covered by Section 504 of the Vocational Rehabilitation Act

Under Section 504 of the Vocational Rehabilitation Act of 1973, no student with a physical or mental impairment which substantially limits one or more major life activities, or has a record of such an impairment, or is regarded as having such an impairment, shall solely by reason of this impairment "be excluded from participation in, be denied the benefits of, or be subjected to discrimination."

Although all students with disabilities under IDEA/s.115.76 also meet the criteria for protection under Sec. 504, there are a limited number of students who are not considered students with disabilities under IDEA/s.115.76 but who do meet the criteria for protection under Sec. 504. Examples of these situations include students with health conditions (e.g., diabetes, asthma) or mobility impairment (e.g. paraplegia) which do not warrant special education placement. Students qualifying only under Sec. 504

criteria are entitled to accommodations and services necessary to benefit from all educational activities available to other students, including state (and district) assessment activities. For these Sec. 504 students, an Individualized Accommodation Plan (IAP) must document appropriate accommodations and services, including any accommodations necessary for participation in assessment activities.

The individuals responsible for developing a student's IAP are responsible for specifying accommodations necessary for participation in state assessments. Students receiving accommodations under Sec. 504 are eligible for the same range of accommodations as students with disabilities under IDEA/s.115.76 or students with limited English proficiency.

According to state law, teachers must administer WKCE and WRCT tests to all students enrolled in the grade. State laws provide for two exceptions to this requirement: certain students with disabilities under IDEA/s.115.76, Stats., and certain students with limited English proficiency under IASA/ s.115.955(7), Stats. For WKCE only, state law provides for an additional exception: students who are excused by their parent or guardian.

It is recognized, however, that it may not be possible to administer WKCE or WRCT tests to an extremely small number of Sec. 504 students, or other students, not described by these exceptions. According to the Office of Civil Rights, circumstances warranting a decision not to test a "Sec. 504-only" student would be extremely rare. One example of such a situation might be a Sec. 504-only student suffering from acute emotional disturbance such as one caused by recent trauma. This student is not necessarily a student with a disability under IDEA/s.115.76, Stats. The IAP team may reasonably conclude that participation in WKCE or WRCT during the respective testing window would be damaging to the student.

The Office of Civil Rights also has advised that it is highly unlikely that a school or district could justify not testing a Sec. 504-only student based on federal law unless the parent agrees that his/her child should not be tested. Under Wisconsin law, parents have the right to excuse their child from the WKCE but not from the WRCT. If the affected Sec.504-only student is excused by their parent or guardian from WKCE, the student should be coded as "excused by parent or guardian."

For reporting purposes, the "excluded" code is used only for certain students with disabilities under IDEA and under s.115.76, Stats., and certain students with limited English proficiency under IASA and under s.115.955(7), Stats. The data collection and reporting software does not permit the "excluded" code to be used for Sec. 504-only students. District and school coding of Sec. 504-only students who are not expected to be tested on WKCE or WRCT should be as follows:

- For the WKCE, use the Testing Status code "P" if the student was excused by a parent or guardian or "F" if the student was not tested due to the IAP team decision.
- For the WRCT, use the Testing Status code "F" for any student who was not tested due to the IAP team decision if submitting student demographic data electronically prior to testing. Otherwise, mark the "504 IAP Not Tested" box under "Reason Not Tested" on the School Header Form. Sec. 504-only students who are not expected to be tested on WKCE (Testing Status "P" or "F") or WRCT (Testing Status "F") are not required to take alternate assessments.

Additional Assessment Considerations for Students with Special Needs

The following assessment considerations reflect Chapter PI 16 of the Wisconsin Administrative Code which refers specifically to tests administered in the eighth and tenth grade.

Results from WKCE assessments, or any other single assessment, cannot be the sole criterion in exiting students from a bilingual-bicultural program or special education services. In addition, results from WKCE assessments cannot be the sole criterion in determining grade promotion, eligibility for courses or programs, eligibility for graduation, or eligibility for participation in post-secondary education opportunities such as the options listed under s.118.55, Stats.

Students whose performance is assessed through alternate assessments may not be penalized in grade promotion, eligibility for courses or programs, eligibility for graduation, or eligibility for post-secondary education opportunities.

At least 30 days before testing, parents of students with LEP must be notified in writing or orally *in their native language* of the district's intent to include or exclude the students and the reason for the decision. For parents of students with disabilities, parent notification will occur as part of the IEP teams' thorough review to determine how each child will participate in state and district assessments.

Parents of students with LEP must receive WKCE or alternate assessment results in their native language or any other means necessary so that parents understand the results of their child's assessment.

Although PI 16 currently refers only to state assessments administered in eighth and tenth grade, the DPI will review this chapter of the Administrative Code. The DPI recommends that educators apply the aforementioned considerations to all state assessments administered in grades 3, 4, 8, and 10.

---

[1]Wisconsin state statutes are currently being revised to reflect the official federal designation of Limited English Proficient (LEP). Prior to this amendment, Wisconsin used the term Limited English Speaking (LES). These terms refer to the same students.

[**IMPORTANT**: These DPI Guidelines to Facilitate the Participation of Students with Special Needs in State Assessments should continue to be used as the primary source of information regarding participation of students with disabilities in WKCE and WRCT. District and school staff are encouraged to read the book *Educational Assessment and Accountability for All Students: Facilitating the Meaningful Participation of Students with Disabilities in District and Statewide Assessment Programs* as a supplement to these guidelines but not in lieu of these guidelines. **This book was designed for more general use and does not specifically address certain situations addressed in these guidelines. When ambiguity, gaps, or perceived conflicts exist, follow the information contained in these guidelines.**]

9/16/99 specneed.doc

Learning Support/Equity and Advocacy Information Update No. 98.14, October 1998

TO: District Administrators, CESA Administrators, CCDEB Administrators, Directors of Special Education and Pupil Services, and Other Interested Parties

FROM: Juanita S. Pawlisch, Ph.D., Assistant Superintendent
Division for Learning Support: Equity and Advocacy

SUBJECT: Guidelines for Complying with the Assessment Provisions of the Individuals with Disabilities Education Act

Wisconsin has published academic content and performance standards for ALL students in the state. As of August 1, 1998, all districts must have adopted either the academic content standards proposed by the state or their own locally-developed standards (s. 118.30, Wis. Stats.). General curriculum in all schools is to be aligned with the academic content standards adopted by the district. Because recent state and federal special education legislation emphasize access to the general curriculum for students with disabilities, educational goals on students' individualized education programs (IEPs) also must be based on the academic content standards.

Federal and state special education legislation also requires that all students with disabilities participate in statewide and districtwide assessments. At present, the statewide assessment system, the Wisconsin Student Assessment System (WSAS), includes the Wisconsin Reading Comprehension Test (WRCT) at third grade and the Wisconsin Knowledge and Concepts Examinations (WKCE) at fourth, eighth, and tenth grades. Specifically, the Individuals with Disabilities Education Act (IDEA) states, "children with disabilities are included in general state and districtwide assessment programs with accommodations, where necessary." Although this bulletin is not intended to provide guidance on when or how to provide assessment accommodations, the department has funded projects that will produce written guidelines addressing these issues. The IDEA, however, also directs state educational agencies to develop guidelines for the provision of alternate assessments to children with disabilities for whom the standard statewide assessment is inappropriate. The purpose of this bulletin is to provide alternate assessment guidelines for statewide assessments. It is the responsibility of local school districts to establish comparable guidelines for districtwide assessments.

Regarding the importance of including students with disabilities in statewide and districtwide assessments, the Department of Public Instruction (DPI) concurs with the rationale provided by the U.S. Department of Education:

> Given the emphasis on assessment in recent educational reform efforts, including state and federal legislation linking assessment and school accountability, it is of utmost importance that students with disabilities be included in the development and implementation of assessment activities. Too often, in the past, students with disabilities have not fully participated in state and district assessments only to be shortchanged by the low expectations and less challenging curriculum that may result from exclusion.
>
> Given the benefits that accrue as a result of assessment, exclusion from assessments based on disability generally would not only undermine the value of the assessment but also violate Section 504.... Similarly, Title II of the Americans with Disabilities Act (ADA) of 1990 provides that no qualified individual with a disability shall, by reason of such disability, be excluded from participation in or be denied the benefits of

the services, programs, or activities of a public entity, or be subjected to discrimination by such entity....

For the *small* (emphasis added) number of students whose IEPs specify that they should be excluded from regular assessments, including some students with significant cognitive impairment, participation in regular assessments is not appropriate.

It is important to note that several state and national reviews concerning alternate assessment suggest that approximately 10 to 20% of students with disabilities, or 1 to 2% of the total student population, would participate via an alternate assessment.

Pursuant to state and federal special education law, all questions regarding the participation of an individual student with disabilities in general statewide and district assessments must be addressed by the IEP team. In the 1997 reauthorization of the IDEA, Congress required that the IEP team must include: "the parents of the child; at least one regular education teacher of the child; (and) at least one special education teacher (or special education provider) of the child." In addition, the IEP team must include "a representative of the local education agency who...is knowledgeable about the general education curriculum." In the state of Wisconsin, knowledge of the general curriculum requires knowledge of the academic content standards. The composition of the IEP team was established to ensure that parents understand and participate in decisions regarding their child's performance; the IEP team has access to broad information relating to the child's performance; and the required connection is made between the child's performance and the academic content standards for the general education curriculum.

Federal and state special education law explicitly directs IEP teams to include in each student's IEP "a statement of the child's present levels of educational performance, including how the child's disability affects the child's involvement and progress in the general education curriculum." The IEP also must include a statement of "measurable annual goals, including benchmarks or short-term objectives related to meeting the child's needs...to enable the child to be involved in and progress in the general curriculum." The review and analysis necessary to develop these IEP statements also are necessary for determining how the student will participate in statewide or districtwide assessment systems.

Special education law also requires that the student's IEP include "a statement of any individual modifications in the administration of state or districtwide assessments of student achievement that are needed in order for the child to participate in such assessment; and if the IEP team determines that the child will not participate in a particular state or districtwide assessment of student achievement (or part of such an assessment), a statement of why that assessment is not appropriate for the child; and how the child will be assessed." To make these determinations, the IEP team must be knowledgeable about the child, including statements of present levels of educational performance and measurable annual goals for the child; in addition, the IEP team must be knowledgeable about the general curriculum, the format of the state or district test, and the alignment between the curriculum and the academic content standards assessed by the state or district assessment system.

How a student participates in a statewide or districtwide assessment system is not an "all or nothing" decision. Instead, there are multiple alternatives for facilitating the participation of a student with a disability. These alternatives reflect combinations of three broad options: participation in the standard assessment with no accommodations, participation in the standard assessment with accommodations, and participation through an alternate assessment. Depending on the needs of the student, these options may be used exclusively or in combination. For example, some students with disabilities may not require any accommodations to participate in the WKCE. Other students with disabilities, however, may need accommodations for some of the content domains (i.e., math, language arts, science or social studies) but not for others. Still other students may need accommodations for some areas within WKCE and alternate assessment for one or more skill domains. Finally, there will be a limited num-

ber of students for whom the standardized test will not be appropriate, and these students will participate in the assessment system through an alternate assessment only.

The only appropriate justification for a student not to participate in the WRCT or WKCE is a decision by the IEP team that, even with accommodations, the student would be unable to demonstrate at least some of the knowledge and skills tested through the standardized assessment. A student with a disability for whom the standardized assessment is not appropriate, however, must still be provided with the opportunity to demonstrate his or her knowledge and skills. Thus, these students must participate in the assessment system through an alternate assessment. The WKCE and WRCT use four proficiency levels to characterize student performance: minimal performance, basic, proficient, and advanced. Beginning with the 1999-2000 school year, the DPI will include a reporting category of prerequisite skill for those students who participate in the statewide assessment system through an alternate assessment.

The IEP team decision regarding student assessment must be based upon a thorough review of child-specific data. A thorough review requires consideration of existing student records, including the most recent evaluation data, formal and informal evaluations conducted by team members, reports by parents and special education and/or general education teachers, classroom work samples, independent educational evaluations, and any other information available to the IEP team.

Based on this thorough review of the student's current educational performance relative to the academic content standards, the IEP team may decide that a child with a disability will participate in the statewide assessment system through an alternate assessment. The review undertaken in this decision-making process can, if documented in the IEP as part of the IEP process, function as an alternate assessment because it measures the individual student's actual current level of performance in each of the general curricular areas assessed by the statewide assessment. It is important to note that to serve as an alternate assessment, the review must be comprehensive, recent, and representative of the student's present level of educational performance. In addition, to qualify as an alternate assessment the IEP team review must be conducted within a time frame that approximates the administration of the statewide standard assessment. The department suggests holding the IEP review within three or four months of the WKCE or WRCT.

As stated in the first paragraph of this bulletin, when a child with a disability participates in an alternate assessment, the assessment should be based on the academic content standards for all students. To assist IEP teams in making these connections, the DPI is publishing "alternate performance indicators" (APIs). These APIs are an extension of the academic content standards assessed through the WKCE or WRCT, and they provide some examples of educational goals for students with disabilities that are aligned to the state standards.

When a student with a disability participates in the state assessment system through an alternate assessment, the parent of the student may want more detailed information about their child's performance relative to the academic content standards. The APIs also are intended to assist IEP teams in communicating with parents and educators about a student's current level of performance relative to the academic content standards. The APIs can serve as a framework for constructing a detailed review for those students participating in the statewide assessment system through alternate assessment. A review using the APIs, however, is only one of many tools for linking alternate assessments and Wisconsin's academic standards. School districts may choose to develop other frameworks for alternate assessments or choose to utilize assessments available from a commercial or public vendor.

Included in this bulletin is a glossary of key terms related to the participation of students with disabilities in state assessments. The intent of this glossary is to provide a common language for communications between educators, administrators, parents, and the DPI regarding the participation of students with disabilities in alternate assessments.

During the second semester of the 1998-99 school year, DPI- and CESA-funded assessment projects will be sponsoring inservice programs focusing on issues related to including students with disabilities in statewide and districtwide assessments. Information will be provided on these opportunities in the coming months. Written information also will be made available through the department regarding the selection of test accommodations, design of alternate assessments, and reporting of assessment results. In addition, to assist educators with aligning IEP goals and the academic content standards for all students, the DPI is publishing a guidebook entitled "A Guide for Understanding and Developing IEPs."

If you have comments or questions regarding this bulletin, please contact Stephanie Petska, Director, Special Education Team, at (608) 266-1781.

GLOSSARY OF TERMS

1. Wisconsin Student Assessment System: At present, the Wisconsin Student Assessment System (WSAS) includes statewide assessments at four grade levels. The majority of students (approximately 98% of the total student population) will participate in the WSAS through standardized assessments. These standardized assessments include the Wisconsin Reading Comprehension Test (WRCT) at third grade and the Wisconsin Knowledge and Concepts Examinations (WKCE) at fourth, eighth, and tenth grade. For approximately 2% of the total population the standardized tests would not be appropriate. Thus, these students participate in the WSAS through alternate assessments.

2. Test Accommodations: A change in the administration or response format of a test. Accommodations do not change the purpose or content of a test; rather, they are used to eliminate distortions in test scores that result from a disability as opposed to variations in knowledge or skill level.

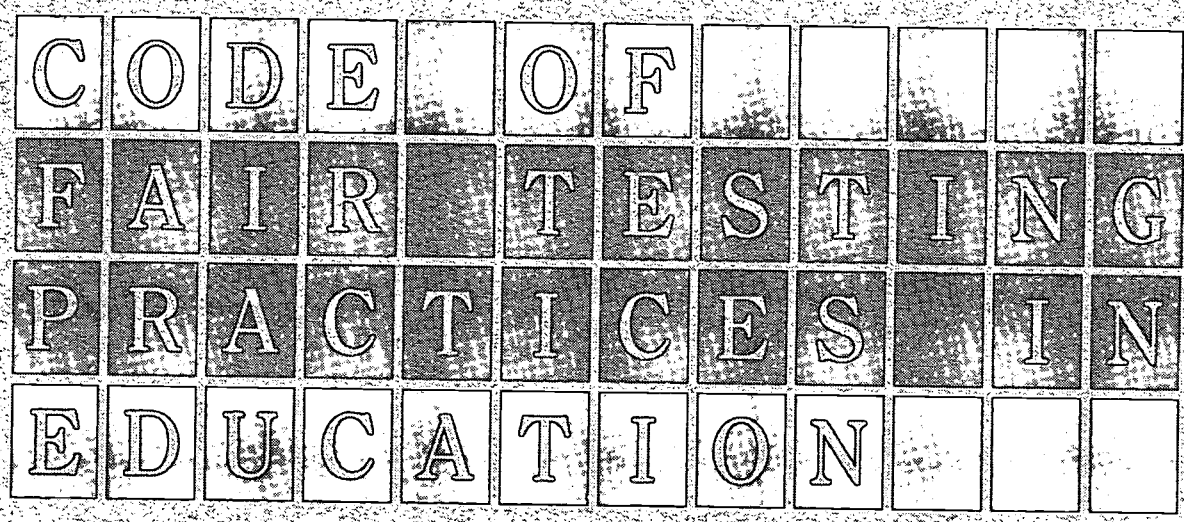
3. Alternate Assessment: An assessment used in place of the standardized assessment administered by the state or district. (For the state assessment system, the standardized assessment is the WRCT or WKCE.) Data are collected via alternative methods such as observations, interviews, record reviews, tests, etc., because students cannot take the standard form of assessment even with appropriate accommodations.

4. Alternate Performance Indicators: Alternate performance indicators (APIs) are descriptions of specific knowledge and skills that, when demonstrated by a student, serve as meaningful predictors of some of the fundamental competencies represented in the state's content and performance standards. APIs have been developed by Wisconsin educators in each of five academic content areas (reading, language arts, mathematics, science, and social studies) for use with students with disabilities. These performance indicators were designed as extensions of the state's existing academic content and performance standards, and they are accompanied by sample activities that educators could use to determine the status of a student's present knowledge and skills. APIs are a resource for teachers as they align a student's performance and their classroom assessments with the state content and performance standards. They are only one possible resource to assist educators in the assessment of students and the determination of their current level of proficiency.

For students with disabilities, the APIs were created for use by the IEP team (including parents) in their assessment of a student's "current level of performance" or the extent to which the student may need modification or replacement of the regular education program or curriculum. APIs may also be helpful to the IEP team in the development of annual goals. They may serve as indicators for progress over time.

This information update can also be accessed through the Internet: http://www.dpi.state.wi.us/dpi/dlsea/een/bulindex.html.

Code of Fair Testing Practices in Education



Prepared by the Joint Committee on Testing Practices

The Code of Fair Testing Practices in Education states the major obligations to test takers of professionals who develop or use educational tests. The Code is meant to apply broadly to the use of tests in education (admissions, educational assessment, educational diagnosis, and student placement). The Code is not designed to cover employment testing, licensure or certification testing, or other types of testing. Although the Code has relevance to many types of educational tests, it is directed primarily at professionally developed tests such as those sold by commercial test publishers or used in formally administered testing programs. The Code is not intended to cover tests made by individual teachers for use in their own classrooms.

The Code addresses the roles of test developers and test users separately. Test users are people who select tests, commission test development services, or make decisions on the basis of test scores. Test developers are people who actually construct tests as well as those who set policies for particular testing programs. The roles may, of course, overlap as when a state education agency commissions test development services, sets policies that control the test development process, and makes decisions on the basis of the test scores.

The Code has been developed by the Joint Committee on Testing Practices, a cooperative effort of several professional organizations, that has as its aim the advancement, in the public interest, of the quality of testing practices. The Joint Committee was initiated by the American Educational Research Association, the American Psychological Association and the National Council on Measurement in Education. In addition to these three groups, the American Association for Counseling and Development/Association for Measurement and Evaluation in Counseling and Development, and the American Speech-Language-Hearing Association are now also sponsors of the Joint Committee.

This is not copyrighted material. Reproduction and dissemination are encouraged. Please cite this document as follows:

Code of Fair Testing Practices in Education. (1988) Washington, D.C.: Joint Committee on Testing Practices. (Mailing Address: Joint Committee on Testing Practices, American Psychological Association, 750 First Avenue, NE, Washington, D.C., 20002-4242.)

BEST COPY AVAILABLE

Code of Fair Testing Practices in Education

The Code presents standards for educational test developers and users in four areas:

A. Developing/Selecting Tests
B. Interpreting Scores
C. Striving for Fairness
D. Informing Test Takers

Organizations, institutions, and individual professionals who endorse the Code commit themselves to safeguarding the rights of test takers by following the principles listed. The Code is intended to be consistent with the relevant parts of the *Standards for Educational and Psychological Testing* (AERA, APA, NCME, 1985). However, the Code differs from the Standards in both audience and purpose. The Code is meant to be understood by the general public; it is limited to educational tests; and the primary focus is on those issues that affect the proper use of tests. The Code is not meant to add new principles over and above those in the Standards or to change the meaning of the Standards. The goal is rather to represent the spirit of a selected portion of the Standards in a way that is meaningful to test takers and/or their parents or guardians. It is the hope of the Joint Committee that the Code will also be judged to be consistent with existing codes of conduct and standards of other professional groups who use educational tests.

A Developing/Selecting Appropriate Tests*

Test developers should provide the information that test users need to select appropriate tests.

Test users should select tests that meet the purpose for which they are to be used and that are appropriate for the intended test-taking populations.

Test Developers Should:

1. Define what each test measures and what the test should be used for. Describe the population(s) for which the test is appropriate.
2. Accurately represent the characteristics, usefulness, and limitations of tests for their intended purposes.
3. Explain relevant measurement concepts as necessary for clarity at the level of detail that is appropriate for the intended audience(s).
4. Describe the process of test development. Explain how the content and skills to be tested were selected.
5. Provide evidence that the test meets its intended purpose(s).
6. Provide either representative samples or complete copies of test questions, directions, answer sheets, manuals, and score reports to qualified users.
7. Indicate the nature of the evidence obtained concerning the appropriateness of each test for groups of different racial, ethnic, or linguistic backgrounds who are likely to be tested.
8. Identify and publish any specialized skills needed to administer each test and to interpret scores correctly.

Test Users Should:

1. First define the purpose for testing and the population to be tested. Then, select a test for that purpose and that population based on a thorough review of the available information.
2. Investigate potentially useful sources of information, in addition to test scores, to corroborate the information provided by tests.
3. Read the materials provided by test developers and avoid using tests for which unclear or incomplete information is provided.
4. Become familiar with how and when the test was developed and tried out.
5. Read independent evaluations of a test and of possible alternative measures. Look for evidence required to support the claims of test developers.
6. Examine specimen sets, disclosed tests or samples of questions, directions, answer sheets, manuals, and score reports before selecting a test.
7. Ascertain whether the test content and norms group(s) or comparison group(s) are appropriate for the intended test takers.
8. Select and use only those tests for which the skills needed to administer the test and interpret scores correctly are available.

*Many of the statements in the Code refer to the selection of existing tests. However, in customized testing programs test developers are engaged to construct new tests. In those situations, the test development process should be designed to help ensure that the completed tests will be in compliance with the Code.

BEST COPY AVAILABLE

B Interpreting Scores

Test developers should help users interpret scores correctly.	Test users should interpret scores correctly.

Test Developers Should:

9. Provide timely and easily understood score reports that describe test performance clearly and accurately. Also explain the meaning and limitations of reported scores.
10. Describe the population(s) represented by any norms or comparison group(s), the dates the data were gathered, and the process used to select the samples of test takers.
11. Warn users to avoid specific, reasonably anticipated misuses of test scores.
12. Provide information that will help users follow reasonable procedures for setting passing scores when it is appropriate to use such scores with the test.
13. Provide information that will help users gather evidence to show that the test is meeting its intended purpose(s).

Test Users Should:

9. Obtain information about the scale used for reporting scores, the characteristics of any norms or comparison group(s), and the limitations of the scores.
10. Interpret scores taking into account any major differences between the norms or comparison groups and the actual test takers. Also take into account any differences in test administration practices or familiarity with the specific questions in the test.
11. Avoid using tests for purposes not specifically recommended by the test developer unless evidence is obtained to support the intended use.
12. Explain how any passing scores were set and gather evidence to support the appropriateness of the scores.
13. Obtain evidence to help show that the test is meeting its intended purpose(s).

C Striving for Fairness

Test developers should strive to make tests that are as fair as possible for test takers of different races, gender, ethnic backgrounds, or handicapping conditions.	Test users should select tests that have been developed in ways that attempt to make them as fair as possible for test takers of different races, gender, ethnic backgrounds, or handicapping conditions.

Test Developers Should:

14. Review and revise test questions and related materials to avoid potentially insensitive content or language.
15. Investigate the performance of test takers of different races, gender, and ethnic backgrounds when samples of sufficient size are available. Enact procedures that help to ensure that differences in performance are related primarily to the skills under assessment rather than to irrelevant factors.
16. When feasible, make appropriately modified forms of tests or administration procedures available for test takers with handicapping conditions. Warn test users of potential problems in using standard norms with modified tests or administration procedures that result in non-comparable scores.

Test Users Should:

14. Evaluate the procedures used by test developers to avoid potentially insensitive content or language.
15. Review the performance of test takers of different races, gender, and ethnic backgrounds when samples of sufficient size are available. Evaluate the extent to which performance differences may have been caused by inappropriate characteristics of the test.
16. When necessary and feasible, use appropriately modified forms of tests or administration procedures for test takers with handicapping conditions. Interpret standard norms with care in the light of the modifications that were made.

BEST COPY AVAILABLE

D Informing Test Takers

Under some circumstances, test developers have direct communication with test takers. Under other circumstances, test users communicate directly with test takers. Whichever group communicates directly with test takers should provide the information described below.

Test Developers or Test Users Should:

17. When a test is optional, provide test takers or their parents/guardians with information to help them judge whether the test should be taken, or if an available alternative to the test should be used.

18. Provide test takers the information they need to be familiar with the coverage of the test, the types of question formats, the directions, and appropriate test-taking strategies. Strive to make such information equally available to all test takers.

Under some circumstances, test developers have direct control of tests and test scores. Under other circumstances, test users have such control. Whichever group has direct control of tests and test scores should take the steps described below.

Test Developers or Test Users Should:

19. Provide test takers or their parents/guardians with information about rights test takers may have to obtain copies of tests and completed answer sheets, retake tests, have tests rescored, or cancel scores.

20. Tell test takers or their parents/guardians how long scores will be kept on file and indicate to whom and under what circumstances test scores will or will not be released.

21. Describe the procedures that test takers or their parents/guardians may use to register complaints and have problems resolved.

Note: The membership of the Working Group that developed the Code of Fair Testing Practices in Education and of the Joint Committee on Testing Practices that guided the Working Group was as follows:

Theodore P. Bartell
John R. Bergan
Esther E. Diamond
Richard P. Duran
Lorraine D. Eyde
Raymond D. Fowler
John J. Fremer
(Co-chair, JCTP and Chair, Code Working Group)
Edmund W. Gordon
Jo-Ida C. Hansen
James B. Lingwall
George F. Madaus
(Co-chair, JCTP)
Kevin L. Moreland
Jo-Ellen V. Perez
Robert J. Solomon
John T. Stewart
Carol Kehr Tittle
(Co-chair, JCTP)
Nicholas A. Vacc
Michael J. Zieky
Debra Boltas and Wayne Camara of the American Psychological Association served as staff liaisons.

Additional copies of the Code may be obtained from the National Council on Measurement in Education, 1230 Seventeenth Street, NW, Washington, D.C. 20036. Single copies are free.

BEST COPY AVAILABLE

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

NOTICE

REPRODUCTION BASIS

☑ This document is covered by a signed "Reproduction Release (Blanket) form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

☐ This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").