



A COMPREHENSIVE K-3 READING ASSESSMENT PLAN

Guidance for School Leaders



**CENTER ON
INSTRUCTION**

A COMPREHENSIVE K-3 READING ASSESSMENT PLAN

Guidance for School Leaders

Joseph K. Torgesen

Florida Center for Reading Research
Florida State University

The Center on Instruction is operated by RMC Research Corporation in partnership with the Florida Center for Reading Research at Florida State University; RG Research Group; the Texas Institute for Measurement, Evaluation, and Statistics at the University of Houston; and the Vaughn Gross Center for Reading and Language Arts at the University of Texas at Austin.

The contents of this document were developed under cooperative agreement S283B050034 with the U.S. Department of Education. However, these contents do not necessarily represent the policy of the Department of Education, and you should not assume endorsement by the federal government.

Editorial design, and production support provided by RMC Research Corporation.

The Center on Instruction requests that no changes be made to the content or appearance of this product.

Preferred citation:

Torgesen, J. K. (2006) *A comprehensive K-3 reading assessment plan: Guidance for school leaders*. Portsmouth, NH. RMC Research Corporation, Center on Instruction

2006

To download a copy of this document, visit www.centeroninstruction.org.



**CENTER ON
INSTRUCTION**



INTRODUCTION

Anyone who has worked as an elementary school teacher or principal understands the value of reliable and valid assessments of early reading progress. Timely, reliable assessments indicate which children are falling behind in critical reading skills so teachers can help them make greater progress in learning to read. Reliable and valid assessments also help monitor the effectiveness of instruction for all children; without regularly assessing children's progress in learning to read, we cannot know which children need more help and which are likely to make good progress without extra help. Because scientific studies have repeatedly demonstrated the value of regularly assessing reading progress^{1,2}, a comprehensive assessment plan is a critical element of an *effective school-level plan for preventing reading difficulties*. The general principles outlined in this guide, such as the early identification of students who are struggling in learning to read³, are all based on scientific findings, but the detailed recommendations for implementation derive from practical experiences in helping many school leaders implement successful plans.

OBJECTIVES OF A COMPREHENSIVE ASSESSMENT PLAN

A comprehensive assessment plan for the early elementary grades has four main objectives:

1. To **identify** students at the beginning of the year who are "at risk" for reading difficulties and who may need extra instruction or intensive interventions if they are to progress toward grade-level standards in reading by the end of the year.
2. To **monitor** students' progress during the year to determine whether "at risk" students are making adequate progress in critical reading skills and to identify any students who may be falling behind.
3. To **collect** information about students that will be helpful in planning instruction to meet their most critical learning needs.
4. To **assess** whether the instruction provided by classroom teachers and intervention specialists is sufficiently powerful to help all students achieve grade-level reading standards by the end of each year from kindergarten through third grade.

DIMENSIONS OF READING SKILL TO ASSESS

The Report of the National Reading Panel⁴ identified five critical components of reading skill that children must master as they progress from non-readers in kindergarten to proficient readers at the end of third grade. They are phonemic awareness, phonics, fluency, vocabulary, and reading comprehension. A comprehensive reading assessment plan should monitor the development of each component in a manner appropriate to each grade level.

Kindergarten

Kindergarten students require sensitive assessments of their growth in phonemic awareness, phonics skills (knowledge of letters and beginning phonemic decoding ability), and vocabulary. Their reading skills are rarely sufficiently developed to usefully assess text reading fluency and reading comprehension. Sometimes listening comprehension is assessed instead of reading comprehension to identify students whose language processing skills place them at risk for difficulties comprehending text once they can read words fluently and accurately.

Grade 1

It is important to continue monitoring students' development of phonemic awareness in first grade because struggling students may continue to have difficulty in this area. The development of accurate and fluent phonemic decoding skills should also be monitored in first grade, since these foundational skills for reading accuracy undergo major development in this period. As soon as students can begin to read connected text with reasonable accuracy, their development of oral reading fluency should be monitored. Oral measures of young children's reading fluency are much more reliable⁵ than measures of silent reading fluency. Oral reading fluency's importance as an index of reading growth extends from first through third grades. Continued growth in vocabulary should also be assessed, and reading comprehension can be reliably assessed in most students by the end of first grade.

Grade 2

Second graders may need continued monitoring of their phonemic decoding ability, especially for multi-syllable words, particularly in schools with high proportions of poor and minority students, who have traditionally been at risk for difficulties with the early mastery of these skills. Continued monitoring of



reading fluency is critical through second grade, since students must make strong growth in this skill to maintain grade-level reading proficiency. A comprehensive assessment plan should also measure second graders' vocabulary and reading comprehension.

Grade 3

The primary dimensions of reading growth that should be monitored in third grade are reading fluency, vocabulary, and reading comprehension.

TYPES OF ASSESSMENTS IN A COMPREHENSIVE PLAN

The four objectives outlined above can be achieved through four types of assessments during the school year: screening, progress monitoring, diagnostic, and outcome tests. They correspond roughly to the four objectives, but all can contribute to helping to plan effective instruction.

Screening Tests

Briefly administered, screening tests provide an initial indication of which students are entering the school year "at risk" for reading difficulties because

Informal Reading Inventories

Informal reading inventories are often used to gain a level of detail about students' specific skills and knowledge that is not typically provided by formal screening, progress monitoring, or diagnostic assessments. For example, some informal inventories provide information about the specific letter-sounds a student may have mastered, or the types of words he or she can accurately decode using phonemic decoding strategies, or the types of errors students make most frequently when reading orally. However, information about test reliability and validity is not usually provided for informal reading inventories.

School leaders and teachers should examine the specific information each element of their comprehensive assessment plan provides, and determine the most efficient way to gain the information necessary for planning classroom instruction and making decisions about allocating school-level resources. The goal is to gain enough information about student progress to make effective decisions while minimizing the time spent administering assessments. With a fully implemented comprehensive assessment plan such as the one described here, there may be less need for informal reading inventories than in the past. Much of the information these inventories provide can be gathered through careful observation during instruction. Informal reading inventories might be used much as formal diagnostic tests are: only when there is a well defined need for additional information that will be directly helpful in making instructional decisions.

they are lagging in the development of critical reading skills. Valid and reliable screening tests can help teachers differentiate their instruction based on what students already know and can do.

Progress Monitoring Tests

Also brief, progress monitoring tests are given periodically to determine whether students are making adequate progress. There are two types of progress monitoring tests; both are important to a comprehensive assessment plan. The “curriculum-embedded” test⁶, in common use for many years, assesses the extent to which students have learned the material taught in the current unit of the core reading curriculum. This type of test helps the teacher identify which students have mastered the material and whether the class is ready to move on to the next unit. These tests are included in core reading program materials or may be constructed by teachers if a published core reading program is not used. There is usually no information about the reliability or validity of this type of test, but teachers frequently find them useful because they provide a relatively detailed assessment of the extent to which students have learned what they were just taught.

Establishing Progress Monitoring Benchmarks

To be valid, benchmark scores used with progress monitoring tests must be based on longitudinal studies in which children’s performance at one point (e.g., in oral reading fluency at the beginning of second grade) is linked to their success in meeting a critical reading goal (e.g., grade-level performance in reading comprehension at the end of second grade) at some other point. Analyses of the relationships between scores on benchmark tests and outcome tests in a large number of students are conducted in an attempt to identify the level of performance on the first test that is associated with a high probability (i.e., 80%) of success on the second test. In the same analyses, performance levels that predict poor performance on the second test are also usually identified. It is important to recognize that performance benchmarks do not predict subsequent performance perfectly. Rather, they establish a certain level of confidence that students are developing adequately, or inadequately, in a given reading skill that is critical to overall reading success. If students are very strong, or very weak, in other skills that are also critical to overall reading success, the prediction from the progress monitoring measure may be in error. Because progress monitoring tests assess skills that are critical to overall reading growth, if students perform at a level seriously below the benchmark for a given time of year, it is an indication that they need immediate interventions to accelerate their development in skills that are low.



The second type of progress monitoring test has a shorter history of use in American schools. Sometimes referred to as “general” or “external” progress monitoring tests, they measure critical reading skills such as phonemic awareness, phonics, fluency, vocabulary, or comprehension, but are not tied to any specific reading curriculum.⁷ Rather, through extensive development research, these tests establish performance targets, or “benchmarks” for different points in the school-year (i.e., beginning, middle, and end) that predict success in meeting grade-level reading standards by the end of the year. When administered at the end of the school year, these tests also identify students who will likely have trouble meeting grade-level standards at the end of the next school year unless they receive extra help.

For example, a general progress monitoring test might establish an oral reading fluency target, or “benchmark” of 69 correct words per minute by February of second grade—a target associated with a high probability of meeting the end of the year grade-level standard on a measure of reading comprehension. Another example would be a benchmark of being able to blend three-phoneme words by the end of kindergarten in order to be prepared for success in learning phonemic decoding skills during first grade. General progress monitoring tests provide performance targets teachers can aim for in order to ensure that their students are on track for meeting grade-level reading standards by the end of the school year. Examples of widely used general progress monitoring tests are the Dynamic Indicators of Basic Early Literacy Skills (DIBELS),⁸ the Texas Primary Reading Inventory (TPRI),⁹ and the Phonological Awareness Literacy Screening (PALS)¹⁰ tests. Curriculum-based measurement is a general term that is frequently used as an umbrella term for general progress monitoring tests. The National Center on Student Progress Monitoring (<http://www.studentprogress.org/>) provides extensive information

Formative and Summative Assessment

The term formative assessment has been widely used to describe assessments that serve essentially the same purpose as the progress monitoring tests described here. Basically, both serve to provide information about student progress in order to make “mid course” corrections or improvements to instruction. Typically, formative assessment is contrasted with summative assessment, or the assessment of a final outcome or product. Summative assessment is synonymous with the term outcome assessment used in this guide.

about the use of progress monitoring assessment to guide instruction. They have also conducted evaluative reviews of various progress monitoring tests and have made these available on their website.

Diagnostic Tests

Relatively lengthy, diagnostic tests provide an in-depth, reliable assessment of important component skills in reading. Their major purpose in the early elementary grades is to provide information for planning more effective instruction. Diagnostic tests should be given when there is a clear expectation that they will offer new, or more reliable, information about a child's reading difficulties that can be used to help plan more powerful instruction.

Diagnostic Tests and Diagnostic Information

It is important to distinguish between *diagnostic tests* and *diagnostic information*. Diagnostic information is any knowledge about a child's skills and abilities that is useful in planning instruction. It can come from student work, teacher observations, or other tests, as well as diagnostic tests. For example, if a child performs poorly on a test of reading comprehension at the end of second grade, it would be useful to know if he or she is impaired in reading fluency or accuracy, knowledge of word meanings, general background knowledge, or use of efficient comprehension strategies. Any information gathered about the child's knowledge and skill in the components of reading comprehension is *diagnostic information* that could be used to direct instructional interventions.

In another example, if a child were struggling to acquire fluent and efficient phonemic decoding skills (phonics), it would be useful to have reliable information about his or her level of phonemic awareness and letter-sound knowledge, since both are required to understand and use the alphabetic principle in reading. If the child were relatively strong in phonemic awareness, but had a poorly developed knowledge of letter-sound relationships, this information could be used to focus intervention work.

Diagnostic tests are one important way to obtain *diagnostic information* that can help guide interventions for students who are experiencing difficulty learning to read. However, reliable and valid *diagnostic information* can come from sources other than formal *diagnostic tests*.¹¹



Reducing the Need for Diagnostic Testing

If schools are implementing screening, external progress monitoring, and outcome assessments in a reliable and valid way, the need for additional testing using formal diagnostic instruments should be reduced. For example, reliable and valid screening measures are available in K-3 for phonemic awareness, phonics, reading fluency, and vocabulary. There are also reliable and valid measures to monitor progress throughout the year in phonemic awareness, letter knowledge, phonics, and reading fluency. If these components are reliably assessed at the beginning of the year and several times during the year with screening and progress monitoring instruments, the resulting diagnostic information may prevent the need for additional assessment with formal diagnostic tests.

For example, if a school used reliable and valid screening tests for phonemic awareness, phonics, and vocabulary at the beginning of first grade, a certain percentage of children would be identified as “at risk” because of low performance on these measures. The question becomes, should these “at-risk” students be administered an additional diagnostic test to learn more about a broader range of components than were tested on the screening measures? The answer would be, *only if this information could be used to plan additional instruction for the “at-risk” students*. The screening measure would already provide information for three major components of reading that can be measured reliably at the beginning of first grade. Based on *diagnostic information* from the screening measures, interventions in critical components of reading could begin immediately, rather than waiting for additional information generated by diagnostic tests. The argument for not doing additional diagnostic testing in this case is that it would not likely add any critical information for planning effective interventions, and might delay the start of necessary interventions for these “at-risk” students.

Using Diagnostic Tests with At-Risk Students

Whether an additional diagnostic measure should be given after a student has been identified as at risk by a screening or progress monitoring measure depends on two things. First, the reliability with which each critical reading component has been assessed is key: If there is some question about whether the child performed poorly because the test was improperly administered, or the child was having a “bad day,” a diagnostic test could be used to *confirm*

the finding about the need for additional instruction. (Less expensively, a different form of the screening or progress monitoring measure could be readministered.) Second, if the screening or progress monitoring battery did not assess all the dimensions of reading or language skill relevant to planning an effective intervention, a diagnostic assessment could help fill any remaining gaps in understanding the child's knowledge and skill.

A number of situations might arise in which knowledge beyond that provided in a screening or external progress monitoring measure would be useful in planning instruction. For example, in some instructional programs, a program-specific placement test is used to help place the child at exactly the right spot in the program's instructional sequence. Further, the child's teacher might find it useful to know precisely which letter-sound correspondences a child knows, or in which "sight words" he or she is fluent. However, neither type of information is typically provided by standardized diagnostic tests. Rather, this information is gained through a program-specific placement test, or less formal teacher-administered tests.

In summary, the screening, progress monitoring, and outcome elements of a comprehensive assessment plan often provide valid and reliable *diagnostic information* about a child's instructional needs. Because they are time-consuming and expensive, complete *diagnostic reading tests* should be administered far less frequently than the other assessments, although specific subtests from diagnostic instruments might be used to provide information in areas not assessed by screening, progress monitoring, or outcome assessments. For example, if progress monitoring measures are not reliably assessing vocabulary, and a child is still struggling with reading comprehension at mid-year, the teacher might seek a mid-year diagnostic assessment of vocabulary to assess the child's skills on this key component of reading comprehension. School leaders should continually ask if the value to teachers of the information from formal diagnostic tests in planning instruction merits the time spent administering such tests.

Reading Outcome Tests

Given at the end of the year, reading outcome tests are frequently group-administered tests of important reading outcomes such as reading comprehension. These tests are important because they give school leaders and teachers feedback about the overall effectiveness of their reading program.



As part of a comprehensive plan, they should be administered at the end of every year from kindergarten through third grade, although the kindergarten tests may differ greatly from those administered at the end of 1st, 2nd, and 3rd grades, once children have begun to acquire skills in reading comprehension. Longitudinal studies of reading have shown that students are much more likely to meet grade-level standards in reading at the end of third grade if they have met those standards in each preceding year (grades K-2).¹² Thus, outcome tests at the end of grades K-2 are useful to school leaders to ensure that instruction in each grade is sufficiently powerful to keep most students on track for successful performance when they take important reading accountability measures at the end of third grade.

BEGINNING IMPLEMENTATION: AN IMPORTANT CAVEAT

The principles described in the three sections above represent an “ideal” comprehensive assessment plan, particularly if the goal is to assess each identified reading component with screening, progress monitoring, and outcome assessments. In practice, schools will need to be selective about the components they measure and the assessments they use. For example, no widely used formal progress monitoring measures for vocabulary and reading comprehension currently meet reasonable standards of reliability and validity. Both vocabulary and reading comprehension assessments take time to measure reliably: brief assessments of reading comprehension may be so unreliable that they do not provide dependable information about young children’s growth from one assessment to the next.¹³ Reading comprehension is typically assessed at the end of 1st, 2nd, and 3rd grades through group-administered tests that take approximately 30-45 minutes. Although these tests could also be given at the beginning and mid-year, they are expensive. Vocabulary also tends to be tested more as an outcome assessment at the end of the year or as a screening variable at the beginning of the year than as a general progress monitoring element during the year. However, each of these elements, vocabulary and reading comprehension, are frequently assessed in curriculum-embedded tests, which can be used by teachers to directly assess whether students have learned the material that has just been taught in the current instructional unit.

IMPLEMENTING A COMPREHENSIVE ASSESSMENT PLAN

This section covers when to conduct the assessments described above and with which students, how to select assessments, who should administer the assessments, resources schools will need to carry out a comprehensive assessment plan, and considerations for managing data effectively.

Which Students to Assess, and When

Screening tests are typically administered to all students at the beginning of the year. Information from the outcome assessment of the previous year may provide useful screening information at the beginning of the new year; however, this information will not typically be available for all students, so some form of screening assessment must be available for at least some entering students. If students with severe disabilities are mainstreamed in the class and cannot respond to the format of the test, alternate testing should be arranged for them.

General progress monitoring tests are typically given at least three times a year. For the tests currently in widest use, the first assessment occurs when school starts, and thus becomes the screening test. General progress monitoring tests are then usually given at mid-year and again at the end of the year, when they contribute to the overall end-of-year outcome assessment of reading competence. Some schools only administer these progress monitoring tests to students who performed below grade level on the previous year's outcome measure, or were designated as "at risk" by the screening test at the beginning of the year. Although this strategy requires that fewer children be assessed, it does risk failing to identify students who begin to fall behind in reading growth during the year. These students' needs for extra instructional support in their current grade would be identified by a reliable mid-year progress monitoring assessment.

Students receiving reading interventions should take general progress monitoring assessments more than three times a year in order to determine whether the interventions are having the desired effect in accelerating their reading growth.¹⁴ A rule of thumb for "at risk" students receiving interventions is to monitor more severe problems more frequently. Weekly or biweekly testing is recommended for students with severe problems, while monthly monitoring is appropriate for students with less severe problems.



Curriculum-embedded progress monitoring tests should also be given whenever the teacher needs guidance on how well students have mastered the content or skill in the current unit of instruction. The time between assessments may vary depending on the curriculum being used or the topics being covered.

Diagnostic tests are administered only when specific questions arise about instruction for individual students that cannot be answered from teacher observations, student work, and other forms of assessment (i.e., screening, progress monitoring, or outcome assessments). They should only be given when there is a clear expectation that they will provide information useful in planning more powerful instruction. Diagnostic tests are also sometimes required when evaluating students for placement in special education programs.

Reading **outcome tests** are administered as close to the end of the year as practical to allow information from them to help make decisions about students for the coming year, and they should be given to all students for whom the test format is appropriate. Obviously, students with severe physical or mental disabilities or who are English language learners may need some form of alternate assessment, but the percentage of students excluded from the standard outcome assessment should be very small. Even though students with some forms of disability may not be expected to perform as highly as students without disabilities, they should still be expected to show reasonable progress on outcome assessments from year to year.

How to Select Assessments

In an earlier section of this document, we outlined the elements of reading growth that should be assessed in a comprehensive reading assessment plan for grades K-3. Beyond knowing what should be assessed, the two most important considerations in guiding the selection of specific tests should be evidence about their reliability and validity. These terms are discussed more completely at the end of this document, but for now it is useful to understand that reliability refers to how consistently a test measures a skill or ability, and validity refers to the extent to which a test actually measures the skill or ability in question. A test can be reliable without being valid, but to be a valid measure of any construct, a test must also be reasonably reliable.

If tests will be used to make important decisions about individual students, the tests should meet reasonable standards of reliability and validity. For example, if students are assigned to receive intensive interventions on the basis of their performance on a screening or progress monitoring test, it is important that these tests reliably measure critical reading skills. Further, if information from the tests is to be used to help plan instruction within the interventions, then the tests used to assign students to particular groups should provide valid measurement of critical skills. Part of the process of selecting tests for use within a comprehensive assessment plan should always include examining the test manuals for information about the test's reliability and validity for the way it will be used within the overall assessment plan.

The reading skill measured by a particular test, as well as its reliability and validity, are the major *scientific* considerations involved in selecting tests for use within a K-3 assessment plan. However, other considerations may also play a role, such as the initial cost of the test, the cost of individual test forms, and the amount of training required to administer the test. Best practice is to choose tests with sufficient evidence of reliability and validity that can also be administered and interpreted in a reliable and valid way by those who will administer and use the test data for making instructional decisions.

Who Should Administer The Assessments?

Screening and progress monitoring tests can be administered by anyone—teachers, paraprofessionals, retired teachers, school counselors, media specialists, art teachers, etc.—trained to administer them correctly. Schools typically either ask teachers to administer the tests to their own students, or create a school-level assessment team to administer the tests. An advantage of having teachers administer the tests is that they may acquire information from directly observing the way students respond that goes beyond a test's basic score. Experience in administering the tests also helps teachers better understand the dimensions of reading skill that are being assessed. This strategy typically does not require additional money (except perhaps for test forms), but it does take time that might otherwise be spent teaching.

School-level assessment teams of four to eight people (depending on school size) could include non-instructional personnel, paraprofessionals, or others such as retired teachers, who are trained to administer tests to students in grades K-3. The assessment team conducts all the assessments and sees



that the scores are entered into the data management system so that they are available to teachers. An advantage of an assessment team is that the tests are likely to be administered more consistently across all classes. A schoolwide assessment team also disrupts instruction less than using teachers to administer the tests. For example, if a progress monitoring assessment requires 10 minutes per student, then a teacher would need to spend slightly more than three hours doing nothing but administering the tests. Another advantage of an assessment team approach is that fewer people need to be trained to administer the tests. Some schools blend approaches, using teachers to administer the tests to some of their students, while the school-level team assesses the rest of the students.

Diagnostic tests are usually administered by an educational diagnostician or school psychologist or by a teacher or reading coach with extensive training in their administration and interpretation.¹³ Some diagnostic tests require that the person administering them have credentials in school or clinical psychology. The diagnostic tests that are most useful in planning instruction assess various dimensions of reading and language skill, and can usually be administered by a wider range of personnel than intelligence tests.

Group-administered, year-end **outcome tests** are usually administered by classroom teachers, often with proctoring help from someone outside the classroom.

Organizing Resources For A Comprehensive Assessment Plan

Rather than specify the personnel required to implement a comprehensive assessment plan, it seems more helpful to identify the essential tasks required to implement such a plan; schools can assign the tasks as their circumstances and resources permit. The following tasks must be routinely accomplished each year:

1. A master testing schedule should specify the weeks during which screening, progress monitoring, and outcome assessments will be administered. Assessments should be given at a reasonably uniform time to all students to facilitate the use of the data in instructional decision-making and planning.
2. All testing materials must be ordered or reproduced in time to reach those who will do the testing.

-
3. All teachers or members of the school-level assessment team need to receive adequate training in administering the tests. It is important to remember that teachers may not be used to administering tests according to standard guidelines, yet these standard guidelines make test data interpretable across students and across testing intervals.
 4. One person needs to be designated to do the necessary follow-up and coordination to ensure that the testing is accomplished by all teachers, or across all students, during the time periods specified in the master testing schedule.
 5. A plan for scoring all tests must be developed and executed.
 6. A plan for entering and summarizing test data is necessary; typically, individual student scores will need to be transferred to a classroom, grade-level, or school file.

Managing Data Effectively

In order to use testing data most effectively, a school-level comprehensive assessment plan needs a school-level data management plan or resource. Although teachers can certainly use the data they obtain from testing their students without a formal data management resource, monitoring student performance over time and making school-level decisions are greatly facilitated by an efficient data management plan.

A number of Web-based data management resources allow schools to enter data locally and provide data summaries and individual student charting that are helpful in interpreting test data. These services typically charge a small fee (e.g., one dollar per student per year), but they add significantly to the ease with which student data can guide both classroom and school-level decisions.

Some school districts have a district-level data management program they can offer schools, or individual schools have developed their own resources using programs like Microsoft Excel. Another approach is to use free resources such as the data management program *Chart Dog*¹⁶ to manage and summarize student data.

The larger point is that finding an efficient way to manage and use the data from a comprehensive assessment plan is as important as gathering the data in the first place. In order to make important instructional decisions (e.g., does the student need school-level intervention resources? Does the teacher need extra



support or professional development in a given area?), more than one person (i.e., teacher, grade-level team leader, reading coach, assistant principal, principal) will need access to student data and reports. Some decisions can be based on individual student data, but others may require summaries of data at the classroom or grade level. Investing in an efficient data management tool is critical to the long-term success of a comprehensive assessment plan.

Key Terms— Reliability and Validity

Reliability

A test's **reliability** is the degree to which it provides a dependable, consistent measurement of some trait or ability.¹⁷ A reliable test is likely to produce similar estimates of a student's ability no matter who gives the test (assuming they are well trained), or when it is administered (assuming testing is conducted at a reasonable time of day). A test's reliability is expressed as a number between 0 and 1, with .80 falling at the lower margin of acceptability, and .90 being the most desirable standard.^{18,19} A test's reliability can be calculated a number of ways; internal consistency reliability typically produces the highest estimates, while test-retest reliability often produces slightly lower estimates.

Validity

In the simplest terms, tests are said to be **valid** if they measure the trait or ability they say they are measuring. Unfortunately, it is easier to define validity than to demonstrate conclusively that a given test is valid, or to describe the level of validity with a single number, as in the case of test reliability. This is because a test's validity depends on the purpose for which it is used. In discussing a test's validity, it is always important to keep its purpose in mind. Most current textbooks on educational testing^{17,18,19} describe three important types of validity: 1) content description; 2) criterion prediction; and 3) construct identification.

Content description validity simply refers to the extent and consistency with which the test items cover a representative sample of the knowledge or ability being measured. This type of validity is usually established by expert judgment and statistical analyses that show the items are consistent in the way they measure the knowledge or skill being tested.

Criterion prediction validity is usually established by determining whether performance on the test in question predicts outcomes in the way it should. For example, a screening test for phonemic awareness and letter knowledge at the beginning of kindergarten should predict a student's ability to decode words phonemically at the end of the year. By the same token, a test of phonemic decoding ability in the middle of first grade should predict oral reading fluency by the end of the year. If these predictive relationships cannot be demonstrated, then something is wrong, either with the theory of reading development on which the tests are based, or with the tests themselves (i.e., perhaps they do not measure the ability with sufficient reliability to predict later development). The authors of screening and progress monitoring tests, in particular, should provide evidence that performance on these tests is usefully related to important outcome measures in reading.

Construct identification validity is the most complex form of validity, and is usually demonstrated by a convergence of evidence from several sources. For example, based on current theories of reading development, scores from a valid test of oral reading fluency should show: 1) regular development with age; 2) differences among groups of students that traditionally show different patterns of development in reading (e.g., differences in socio-economic levels, differences between students who are, or are not, classified as learning disabled); 3) responsiveness to intensive interventions that have been shown to affect reading fluency; and, 4) appropriate relationships with other reading skills (i.e., a significant relationship with reading comprehension). Many types of evidence are usually assembled to demonstrate a test's construct identification validity.



REFERENCES

- ¹Fuchs, L., & Fuchs, D. (1999). Monitoring student progress toward the development of reading competence: A review of three forms of classroom-based assessment. *School Psychology Review, 28*, 659-671.
- ²Shinn, M. (1998). *Advanced applications of curriculum-based measurement*. New York: Guilford.
- ³Torgesen, J. K. (2004). Avoiding the devastating downward spiral: The evidence that early intervention prevents reading failure. *American Educator, 28*, 6-19.
- ⁴National Reading Panel (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. National Institute of Child Health and Human Development, Washington, D.C.
- ⁵Fuchs, L. S., Fuchs, D., Hosp, M. K., & Jenkins, J. R. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading, 5*, 239-256.
- ⁶Diamond, L. (2005). Assessment-driven instruction: A systems approach. *Perspectives, Fall*, 33-37
- ⁷Deno, S. L. (2003). Developments in curriculum based measurement. *The Journal of Special Education, 37*, 184-192)
- ⁸Official DIBELS home page <http://dibels.uoregon.edu/>
- ⁹Official TPRI home page <http://www.tpri.org/>
- ¹⁰Official PALS home page <http://pals.virginia.edu/>
- ¹¹Nitko, A. J. (2001). *Educational assessment of students* (3rd Ed.). Englewood Cliffs, NJ: Prentice- Hall/Merrill Education.
- ¹²Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology, 80*, 437-447.
- ¹³Fuchs, L., Fuchs, D., & Maxwell, L. (1988). The validity of informal reading comprehension measures. *Remedial and Special Education, 9*, 20-28.

-
- ¹⁴Shinn, M. (2002). Best practices in using curriculum-based measurement in a problem solving model. In Thomas A. & Grimes J. (Eds.) *Best Practices in School Psychology IV*. Bethesda MD: National Association of School Psychologists.
- ¹⁵American Psychological Association (1999). *The Standards for Educational and Psychological Testing*. American Psychological Association: Washington, D.C.
- ¹⁶Chartdog url:
<http://www.interventioncentral.org/htmldocs/tools/chartdog/chartdog.shtml>
- ¹⁷Anastasi, A., & Urbina, S. (1997). *Psychological testing*. (7th Ed.). Upper Saddle River, NJ: Prentice-Hall.
- ¹⁸Aiken, L. R. (1994). *Psychological testing and assessment*. Needham Heights, MA: Allyn & Bacon.
- ¹⁹Salvia, J. & Ysseldyke, J. E. (1998). *Assessment* (7th Ed.) Boston: Houghton Mifflin.



CENTER ON
INSTRUCTION