acadience™

Acadience™ Reading K–6
Technical Manual

acadience™reading

Roland H. Good, III, PhD

Ruth A. Kaminski, PhD

Elizabeth N. Dewey

Joshua Wallin

Kelly A. Powell-Smith

Rachael Latimer

Dynamic Measurement Group, Inc.

https://acadiencelearning.org/

Table of Contents

Acknowledgments

The program of research and development that has culminated in Acadience Reading K–6 has been a collaborative effort among many dedicated contributors. The talents and efforts of literally thousands of individuals contributed to the successful completion of this edition of Acadience Reading, including Dynamic Measurement Group research scientists and staff, research colleagues from across the country, educators and school personnel, children and parents, and publishing partners.

Throughout a four-year program of research and focused effort, these individuals provided skill, expertise, time, and unlimited energy for the research and development of Acadience Reading. Listings of those people who contributed to the research and development of Acadience Reading are included in this manual and in the *Assessment Manual*. There are, in addition, numerous unnamed children, teachers, and school personnel to whom we owe our special thanks. These people volunteered their time to participate in the research and provided invaluable feedback and suggestions on the measures. We are deeply indebted to each and every individual for his or her contribution.

We deeply appreciate the contributions of Kelli D. Cummings, who was instrumental in developing the initial structure and outline for this technical manual, as well as making substantial contributions to many of the research studies reported here. We also deeply appreciate all of the work and support of Trent Atkins in conducting the Acadience Reading Oral Reading Fluency Readability Study reported in this manual. He was the site coordinator of the study and recruited participants, trained data collectors, supervised all assessments, supervised data entry, and assisted with the reporting of the results.

Roland H. Good III, Ruth A. Kaminski, Elizabeth N. Dewey, Joshua Wallin, Kelly A. Powell-Smith, and Rachael J. Latimer

February, 2011

Research Scientists
Chantal Dufour-Martel, Ph.D.
Nancy Knutson, Ph.D.
Kathleen Petersen, Ph.D.
Stephanie Stollar, Ph.D.

Graphic Designer
Karla Wysocki

Support Staff
Daniel Cohn
Laura Collins
Carol Gassman
Michele Heckel
Jeff Heriot
Sarah Laszlo
Maya O'Neil

Research Assistants
Katherine Bravo Aguayo
Alisa Dorman
Rebecca Freund
Annie Hommel
Douglas Rice

Field Site Coordinators
Ann Marie Anderson
Theresa Fletcher
Lana Gerber
Mary Giboney
Allison Hardin
Debra Hawkins
Tammy Hillebrand
Carol Keskeny
Sara Krebs
Kristen MacConnell
Tina McMullen
Amy Murdoch
Kristin Orton
Lori Pinchot

Alecia Rahn-Blakeslee
Karla Reed
Janet Reynolds
Janet Richards
Christina Schmid
Sue Spiker
Lisa Habedank Stewart
Peggy Taylor
Carol Trumbo
Lillian Vardon
Jennifer Watson-Kilgrow
Christi Whitter

Chapter 1: **Foundations and History**

Acadience Reading K–6 is a set of measures used to assess early literacy and reading skills for students from kindergarten through sixth grade.

You can use Acadience Reading to:

- identify students who may be at risk for reading difficulties;
- help teachers identify areas to target instructional support;
- monitor at-risk students while they receive additional, targeted instruction; and
- examine the effectiveness of your school's system of instructional supports.

Acadience Reading is designed to be an efficient, cost-effective tool used to help make decisions about reading instruction, to help the teacher provide support early, and to prevent the occurrence of later reading difficulties. Acadience Reading assesses *basic early literacy skills*, or the essential skills that every child must master to become a proficient reader (National Reading Panel, 2000; National Research Council, 1998).

Theoretical Framework

The Basic Early Literacy Skills

- ***Phonemic Awareness:*** Hearing and using sounds in spoken words.
- ***Phonics:*** The system of letter-sound relationships that serves as the foundation for decoding words in print.
  - *Alphabetic Principle and Basic Phonics:* The concept that printed letters correspond to the sounds of spoken words. Knowing the most common sounds of consonants and vowels and sounding out phonetically regular VC and CVC words.
  - *Advanced Phonics and Word Attack Skills:* Knowing all of the sounds for letters and letter combinations and sounding out written words.
- ***Accurate and Fluent Reading of Connected Text:*** Reading stories and passages easily and confidently with few mistakes.
- ***Vocabulary and Language Skills:*** Understanding and correctly using a variety of words.
- ***Reading Comprehension:*** The cognitive process during which a reader integrates multiple complex skills (e.g., language, prior knowledge, code, context, etc.) to understand and gain meaning from text.

Assessing student performance on the basic early literacy skills, which are also known as core components or foundational skills, can help distinguish children who are on track to become successful readers from children who are likely to struggle and require additional support to be successful. These skills are the

basic building blocks that every child must master in order to become a proficient reader (Adams, 1990; National Reading Panel, 2000; National Research Council, 1998). These skills also can be improved with effective instruction (Kame'enui, Carnine, Dixon, Simmons, & Coyne, 2002; Simmons & Kame'enui, 1998; Torgesen, et al., 1999).

The Acadience Reading measures are designed to be indicators of the basic early literacy skills. An indicator is a brief, efficient index that provides a fair degree of certainty about a larger, more complex system or process. For example, a pediatrician measures a child's height and weight as a quick and efficient indicator of that child's physical development. Similarly, each Acadience Reading measure is a quick and efficient indicator of how well a child is doing in learning a particular basic early literacy skill (see Table 1.1). As indicators, Acadience Reading measures are not intended to be comprehensive, in-depth assessments of each and every component of a basic early literacy skill. Instead, they are designed to measure key components that are representative of that skill area and to be predictive of overall reading competence. The only Acadience Reading measure that is not designed to be an indicator of a basic early literacy skill is Letter Naming Fluency (LNF). For young students, LNF provides a powerful prediction of the difficulty the student will experience in learning the basic early literacy skills. Thus, LNF is used in identifying students who may need additional support but is not used to provide instructional goals.

Table 1.1 Alignment of Acadience Reading Measures with Basic Early Literacy Skills

Basic Early Literacy Skills	Acadience Reading Measures
Phonemic Awareness	First Sound Fluency (FSF) Phoneme Segmentation Fluency (PSF)

Alphabetic Principle and Basic Phonics	Nonsense Word Fluency (NWF)
	    –Correct Letter Sounds
	    –Whole Words Read

Advanced Phonics and Word Attack Skills	Oral Reading Fluency (ORF) –Accuracy

Accurate and Fluent Reading of Connected Text	Oral Reading Fluency (ORF)
	    –Correct Words Per Minute
	    –Accuracy

Reading Comprehension	Maze
	Oral Reading Fluency (ORF)
	    –Correct Words Per Minute
	    –Retell Total/Quality of Response

Vocabulary and Language Skills	Word Use Fluency—Revised (Available as an experimental measure. Email info@acadiencelearning.org for more information.)

Oral Reading Fluency (ORF) is a complex measure that serves as an indicator of many different skills. In addition to measuring the student's fluency and automaticity in reading of connected text, ORF examines the student's accuracy, which provides an indicator of advanced phonics and word attack skills. ORF is also a good indicator of reading comprehension for most students, and when combined with Retell and Maze provides a robust and powerful indicator of comprehension. ORF and Maze also require adequate vocabulary and language skills to comprehend the content of the passages.

The model in Figure 1.1 (on page 4) shows the relationships among the basic early literacy skills, the Acadience Reading measures, and the timeline for achieving benchmark goals for each measure. The basic early literacy skills are represented by the rounded boxes at the top of the figure (e.g., phonemic awareness, phonics). The arrows connecting the rounded boxes show how the early literacy skills relate to one another and lead to reading comprehension. The arrows from the rounded boxes to the boxes in the middle level show the linkage between the basic early literacy skills and the Acadience Reading measures. The lines between the Acadience Reading measures and the timeline at the bottom indicate the target time of the benchmark goals for that measure.

In this model, (a) automaticity with the code in combination with (b) vocabulary and language skills provide a necessary foundation for reading comprehension. If the student does not have adequate skills in either area, the development of reading comprehension is likely to be compromised.

The model is intended to highlight the primary, most powerful, and instructionally relevant relationships. Other, secondary relations between core components are not included in this figure for clarity. For example, in addition to the relationship between phonemic awareness and phonics, there is also a reciprocal relationship between phonics and phonemic awareness. The model emphasizes this set of relationships in a prevention-oriented framework where phonemic awareness skills can be developed very early and can provide a foundation for successful phonics instruction.

Two caveats are important to note with respect to Figure 1.1. First, the figure is intended to assist in organizing the developmental progression of skills and the linkage to the Acadience Reading indicators and timeline. Although the core components are portrayed as distinct rounded boxes, the skills are tightly intertwined in proficient reading. Phonemic awareness and phonics skills, for example, might be taught and practiced in isolation in a designed curriculum, but instruction is not complete until the skills are integrated. A complete understanding of how words are portrayed in written English requires the integration of all core components into a coherent whole. Second, the role of systematic and explicit instruction is critical throughout this model. Acquisition and mastery of an earlier skill by itself is unlikely to result in achievement of the subsequent skill. However, a foundation of an earlier-developed skill, combined with systematic and explicit instruction in the subsequent skill, is likely to result in successful achievement.

Figure 1.1 Model of Basic Early Literacy Skills, Acadience Reading Indicators, and Timeline

Note: For Nonsense Word Fluency, the first number is the Correct Letter Sounds goal, and the second number is the Whole Words Read Goal. For Oral Reading Fluency, the first number is the Words Correct goal, the second number is the Accuracy goal, and the third number is the Retell goal. For example, at the end of second grade, a student should be able to read 87 words with 97% accuracy and retell of 27 words relevant to the passage.

* Word Use Fluency–Revised (WUF–R) is available as an experimental measure. Email info@acadiencelearning.org for more information.

** Kindergarten through third grade benchmark goals are illustrated; benchmark goals for fourth through sixth grade are also available.

The Importance of Fluency

Acadience Reading assesses reading fluency and automaticity, which, when measured together, are the best indicators of reading performance. Reading fluency is "accurate reading of connected text at a conversational rate with appropriate prosody" (Hudson, Lane, & Pullen, 2005, p. 702). Readers still show improvement in how quickly they read, even long after they have become accurate, thus demonstrating that continued exposure and over-learning are necessary for word recognition to become automatic (Logan, 1988, 1997). Measuring fluency is not limited to oral reading in connected text; fluency in phonemic awareness and understanding of the alphabetic principle should be measured as well, because without fluent knowledge of letters and sounds, young children cannot apply them "on the fly" in connected text when they really matter.

General Outcome Measures

Acadience Reading was developed based on measurement principles from Curriculum-Based Measurement (e.g., Deno & Mirkin, 1977; Deno, 1985; Deno & Fuchs, 1987), and General Outcome Measurement (GOM, Fuchs & Deno, 1991). The Acadience Reading measures were designed to be economical and efficient indicators of a student's progress toward achieving a general outcome such as reading or phonemic awareness, and to be used for both benchmark assessment and progress monitoring. With General Outcome Measures (GOM), student performance on a common task is sampled over time to assess growth and development toward meaningful long-term goals. GOMs measure key skills that are representative of important outcomes such as reading competence. The GOM approach is different from another commonly used formative assessment approach called Mastery Monitoring in which test content is drawn directly from the content taught (e.g., end-of-unit tests in a curriculum). For further discussion of the differences between GOM and Mastery Monitoring, please see Kaminski, Cummings, Powell-Smith, & Good, 2008.

As GOMs, the Acadience Reading measures were designed to be economical and efficient indicators of students' skills, and they include the following features:

- They are standardized assessments, which means they are administered and scored exactly the same way every time with every student. An assessment must be standardized in order to compare results across students or across time, or to compare student scores to a target goal.
- They include alternate forms of approximately equal difficulty, so that student progress can be measured over time.
- They are brief and repeatable, so that students can be assessed efficiently and frequently.
- They are reliable, which means they provide a relatively stable assessment of the skill across time, different forms, and different assessors.
- They are valid, which means they are measuring the essential early literacy skills they are intended to measure.
- They are sensitive to student growth over relatively short periods of time.

Purposes of Acadience Reading Testing

Acadience Reading was designed for formative assessment, or ongoing assessment that is used to adapt teaching to meet student needs, and is used for two primary types of formative assessment: Benchmark Assessment and Progress Monitoring. Unlike high-stakes testing, which is used for decisions that have substantial consequences for students, such as retention or placement in special education, formative assessment is considered low-stakes testing because the results are used for making modifications to instruction to enhance student learning

(Kaminski & Cummings, 2007). Test items or copies of the Acadience Reading assessments should never be used for student instruction or practice in the classroom or at home.

Having students practice the tests may result in artificially high scores, which could prevent those students from receiving the instruction they need to make adequate progress. Such practices compromise the validity and utility of Acadience Reading as measurement tools. Table 1.2 summarizes appropriate uses of Acadience Reading.

For further information on the appropriate use of Acadience Reading, please see the position papers from the Acadience Reading authors on Dynamic Measurement Group's website (https://acadiencelearning.org/).

Acadience Reading is used for two primary types of formative assessment, Benchmark Assessment and Progress Monitoring.

Table 1.2 Uses of Acadience Reading

	Appropriate Uses	*Inappropriate Uses*
Student Level	• Identify students who may be at risk for reading difficulties • Help identify areas to target instructional support • Monitor at-risk students while they receive additional, targeted instruction • Research	• Label, track, or grade students • Make decisions regarding retention and promotion
Systems Level	• Examine the effectiveness of a school's system of instructional supports • Research	• Evaluate teachers • Make decisions about funding • Make decisions about rewards for improved performance or sanctions for low performance

Benchmark Assessment

Benchmark assessment refers to testing all students within a school or grade three times per year for the purpose of screening the students to identify those who may be at risk for reading difficulties. Benchmark assessment also provides school-wide information to evaluate and improve the system of curriculum and instruction. Benchmark assessment is always conducted using grade-level material. The measures administered for benchmark assessment vary by grade and time of year, and they include those measures that are most relevant for making instructional decisions at that time.

Progress Monitoring

Progress monitoring refers to testing conducted more frequently for students who may be at risk for future reading difficulty. Progress monitoring is completed using Acadience Reading measures that correspond to the skill areas in which students are receiving instruction, and is designed to ensure that they are making adequate progress. Progress monitoring can be conducted using grade-level or out-of-grade materials, depending on the student's level of skill and instructional needs. Decisions about the skill areas and levels to monitor are made at the individual student level.

The Outcomes-Driven Model

Acadience Reading measures were developed to provide teachers with information they need to make decisions about instruction. The authors of Acadience Reading advocate a data-based decision-making model referred to as the Outcomes-Driven Model, because the data are used to make decisions to improve student outcomes by matching the amount and type of instructional support with the needs of individual students. Figure 1.2 illustrates the five steps of the Outcomes-Driven Model.

Figure 1.2 The Outcomes-Driven Model
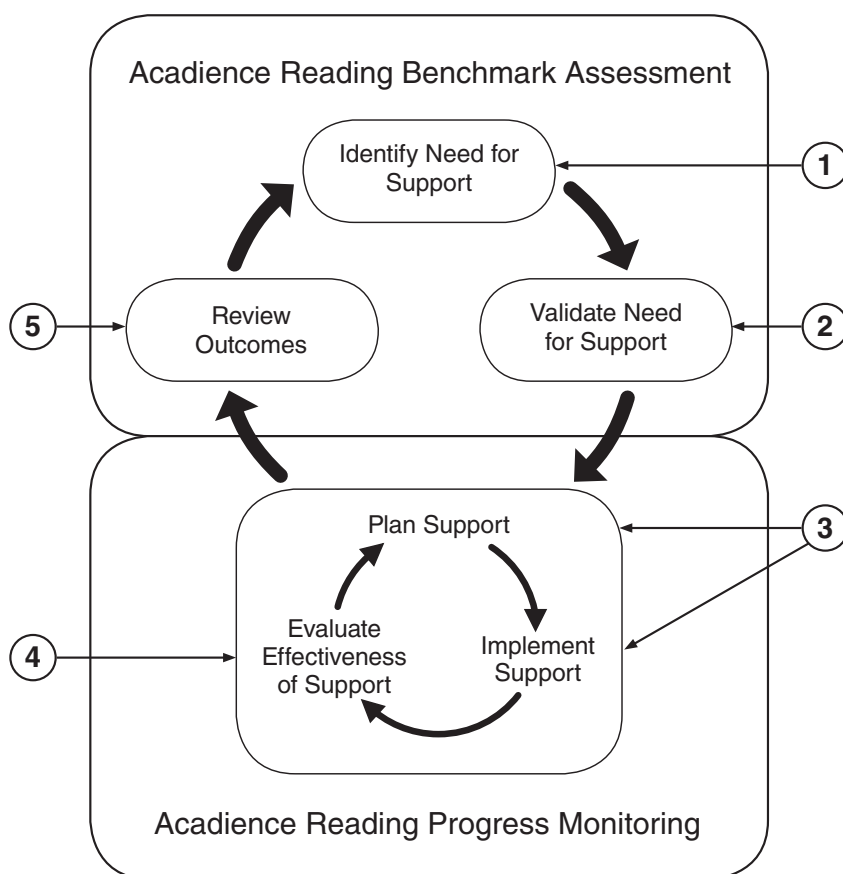
The Outcomes-Driven Model is based on foundational work with a problem-solving model (see Deno, 1989; Shinn, 1995; Tilly, 2008) and the initial application of the problem-solving model to early literacy skills (Kaminski & Good, 1998). The general questions addressed by a problem-solving model include: What is the problem? Why is it happening? What should be done about it? Did it work? (Tilly, 2008). The Outcomes-Driven Model was developed to address these questions, but within a prevention-oriented framework designed to preempt early reading difficulty and ensure step-by-step adequate progress toward outcomes that will result in established, adequate reading achievement.

The steps illustrated in Figure 1.2 repeat each semester as a child progresses through the grades. At the beginning of the semester, the first step is to identify students who may need additional support. At the end of the semester, the final step is to review outcomes, which also facilitates identifying students who need additional support for the next semester. The middle-of-year benchmark assessment is used to review outcomes from the first semester and identify need for support for the second semester. By following these steps, educators can ensure that students who are on track to become proficient readers continue to make adequate progress, and

that those students who are not on track receive the support they need to become proficient readers. The five steps of the Outcomes-Driven Model are:

Step 1: Identify need for support early. This process occurs during benchmark assessment and is also referred to as universal screening. The purpose is to identify those students who may need additional instructional support to achieve benchmark goals. The benchmark assessment also provides information regarding the performance of all children in the school with respect to benchmark goals. All students within a school or grade are tested with Acadience Reading three times per year on grade-level material. The testing occurs at the beginning, middle, and end of the school year.

Step 2: Validate need for support. The purpose of this step is to be reasonably confident that an individual student needs or does not need additional instructional support. Before making individual student decisions, it is important to consider additional information beyond the initial data obtained during benchmark testing. Teachers can always use additional assessment information and knowledge about a student to validate a score before making decisions about instructional support. If there is a discrepancy in the student's performance relative to other information available about the student, or if there is a question about the accuracy of a score, the score can be validated by retesting the student using alternate forms of the Acadience Reading measures or additional diagnostic assessments as necessary.

Step 3: Plan and implement support. In general, for students who are meeting the benchmark goals, a good, research-based core classroom curriculum should meet their instructional needs, and they will continue to receive benchmark assessment three times per year to ensure they remain on track. Students who are identified as needing support are likely to require additional instruction or intervention in the skill areas where they are having difficulties.

Step 4: Evaluate and modify support as needed. Students who are receiving additional support should have their progress monitored more frequently to ensure that the instructional support provided is helping them make adequate progress toward important literacy goals. Students should be monitored on the measures that provide an indicator of the skill areas where they are having difficulties and where they are receiving additional instructional support. Progress monitoring may occur once per month, once every two weeks, or as often as once per week. In general, students who need the most intensive instruction are monitored most frequently.

Step 5: Review outcomes. Each benchmark assessment (semester) provides an opportunity to review outcomes and ensure adequate progress for each individual student and for all students in the school-wide system. By looking at the benchmark assessment data for all students, schools can ensure that their system of instructional supports—both the core curriculum and additional interventions—are meeting the needs of all children. If a school identifies areas of instructional support that are not working as desired, the school can use the data to modify the school-wide system and improve outcomes.

The use of Acadience Reading within the Outcomes-Driven Model is consistent with the most recent reauthorization of the Individuals with Disabilities Education Improvement Act (IDEA, 2004), which allows the use of a Response to Intervention (RtI) approach to identify children with learning disabilities. In an RtI approach, early intervention is provided to students who are at risk for the development of learning difficulties. Data are gathered to determine which students are making adequate progress with the instruction or intervention provided and which students are in need of more intensive support (Fuchs & Fuchs, 2006).

Interpreting Acadience Reading K–6 Data: Frames of Reference

There are four frames of reference in providing meaning for Acadience Reading scores: (a) criterion-referenced benchmark goals and cut points for risk; (b) individually referenced interpretations; (c) local norm-referenced interpretations; and (d) system-wide, norm-referenced interpretations. While all frames of reference provide valuable information about a student, the authors of Acadience Reading generally regard the criterion-referenced information as most important, followed by the individually referenced information and then the local norm-referenced information.

These four frames of reference can be used to interpret results on individual scores and on the Reading Composite Score. The Reading Composite Score is a combination of multiple Acadience Reading scores and provides the best overall estimate of the student's reading proficiency. For more information about the Reading Composite Score as well as worksheets to calculate it, see Appendix 6 of the *Acadience Reading Assessment Manual* (Good, et al., 2011).

Criterion-Referenced Interpretations: Understanding Benchmark Goals and Cut Points for Risk

Acadience Reading benchmark goals are empirically derived, criterion-referenced target scores that represent adequate reading progress. A benchmark goal indicates a level of skill where the student is likely to achieve the next Acadience Reading benchmark goal or reading outcome. Benchmark goals for Acadience Reading are based on research that examines the predictive validity of a score on a measure at a particular point in time, compared to later Acadience Reading measures and external outcome assessments. If a student achieves a benchmark goal, then the odds are in favor of that student achieving later reading outcomes if the student receives research-based instruction from a core classroom curriculum.

The *cut points* for risk indicate a level of skill below which the student is unlikely to achieve subsequent reading goals without receiving additional, targeted instructional support. Students with scores below the cut point for risk are identified as likely to need *intensive support*. Intensive support refers to interventions that incorporate something more or something different from the core curriculum or supplemental support. Intensive support might entail:

- delivering instruction in a smaller group,
- providing more instructional time or more practice,
- presenting smaller skill steps in the instructional hierarchy,
- providing more explicit modeling and instruction, and/or
- providing greater scaffolding

Because students needing intensive support are likely to have individual and sometimes unique needs, their progress is monitored frequently and their intervention is modified dynamically to ensure adequate progress.
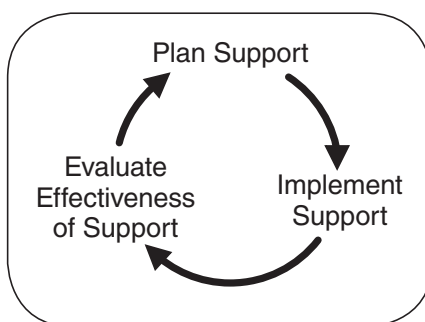
These progress monitoring steps from the Outcomes-Driven Model (see Figure 1.2, page 7) provide an intervention feedback loop. By planning, implementing, and evaluating the effectiveness of support in an ongoing loop, the intervention can be modified dynamically to meet the student's needs.

Students are likely to need *strategic support* when their scores are between the benchmark goal and the cut point for risk. In this range, a student's future performance is harder to predict. Strategic instructional support is carefully targeted additional support in the skill areas where the student is having difficulty. These students should be monitored regularly to ensure they are making adequate progress, and they should receive increased or modified support if necessary to achieve subsequent reading goals.

To gain a better understanding of what Acadience Reading results mean in a local context, districts and schools can examine the linkages between the Acadience Reading benchmark goals and cut points for risk and their own outcome assessments, such as state-level criterion-referenced tests. By comparing Acadience Reading measures to an outcomes assessment (e.g., Buck & Torgesen, 2003; Wilson, 2005), and by calculating conditional probabilities (e.g., "80% of students at benchmark on ORF at the end of third grade met the Proficient level on the state criterion-referenced test."), schools can determine how the Acadience Reading benchmark goals compare to their own external criteria.

A score at or above the benchmark goal indicates that the odds are in the student's favor of achieving the next goal, but it is not a guarantee. For example, if students at or above the benchmark goal have an 85% chance of meeting the next goal, that means that 15% of students in the benchmark range may not achieve that goal. Some students who achieve scores at or above the benchmark goal may still need supplemental support to achieve the next goal. It is important to attend to other indicators of risk when planning support for students, such as attendance, behavior, motivation, vocabulary and language skills, and other related skill areas.

The Acadience Reading benchmark goals and cut points for risk can be found in Appendix A.

Table 1.3 provides interpretations of student performance with respect to the benchmark goals and cut points for risk. Additional information is provided in Appendix A.

Table 1.3 Student Performance Interpretations

Score level	Likely need for support to achieve subsequent early literacy goals	Interpretation
At or Above Benchmark *scores at or above the benchmark goal*	Likely to Need Core Support	The odds are in the student's favor (approximately 80–90%) of achieving subsequent early literacy goals. The student is making adequate progress in reading and is likely to achieve subsequent reading benchmarks with appropriate and effective instruction. The student needs continuing effective curriculum and instruction.
Below Benchmark *scores below the benchmark goal and at or above the cut point for risk*	Likely to Need Strategic Support	The odds of achieving subsequent early literacy goals are roughly 40–60% for students with skills in this range. Students with scores in this range typically need strategic, targeted instructional support to ensure that they make adequate progress and achieve subsequent reading benchmarks.
Well Below Benchmark *scores below the cut point for risk*	Likely to Need Intensive Support	The odds of achieving subsequent early literacy goals are approximately 10%–20% for students whose performance is below the cut point for risk. The student is unlikely to achieve subsequent reading benchmarks unless provided with substantial, intensive instructional support.

Individually Referenced Interpretations: Analyzing Student Growth and Progress Over Time

In addition to information on where a student is performing relative to the benchmark goals and cut points for risk, Acadience Reading also allows interpretations based on where the student's skills are relative to their past performance. For example, even though a student's Oral Reading Fluency score of 45 words correct per minute might be below the cut point for risk, the score of 45 might represent substantial progress compared to previous scores. For individually referenced interpretations, Acadience Reading results are used to examine individual student performance over time. Evaluating student growth is essential in determining whether the student is making adequate progress toward later goals. Examining student growth (i.e., progress monitoring) is also essential in Response-to-Intervention (RtI) models of service delivery and educational decision-making. Progress monitoring helps the teacher decide whether the instructional support the student is receiving is adequately addressing the student's needs, or whether changes should be made to that support.

Local Norm-Referenced Interpretations: Comparing Students Districtwide

Local norms allow a school or district to compare an individual student's performance to other students in the district. Local norms have the important advantage of being representative of the student's district. Another important advantage is that local norms can be updated yearly. If a district's population changes over time, local norms from the current year will continue to be representative of that population. Although local norms are representative of the district, they are not necessarily representative of the national population. If the average achievement in a given school is below the national average achievement score, all percentile ranks would be affected. For example, the score at the 40th percentile in a low-performing district may be at the 20th percentile in a high-performing district. Local normative comparisons also can be problematic when a small number of

students is included. All students in the district should be included when determining local norms, but small districts may not have enough students for stable local normative comparisons. Most data management services for Acadience Reading data will provide local norms.

Local norms can be valuable for a district when making decisions about providing additional support for students. Districts have the flexibility of choosing a level, based on local norms, below which students are provided with additional instructional support. Districts can make this choice based on any pertinent considerations, including financial and staff resources. If a district is able to provide support to 50% of students, students may be selected for support who are at the 50th percentile or lower on Acadience Reading. If a district is only able to provide additional support to 15% of students, students can be selected who are at the 15th percentile or lower on Acadience Reading. By using districtwide local norms, students with equivalent needs in different schools can be provided with support.

For norm-referenced interpretations with Acadience Reading, the following descriptors for levels of performance are provided. The performance descriptors are intended to describe the current level of skill for the student in comparison to other students in the district. They are not intended as statements about what the student is capable of learning with appropriate effective instruction.

Table 1.4 Levels of Performance

Percentile Ranges	Performance Descriptors. Compared to other students in the school or district, the student's performance is:
98th percentile and above	Upper Extreme
91st to 97th percentile	Well-Above Average
76th to 90th percentile	Above Average
25th to 75th percentile	Average
9th to 24th percentile	Below Average
3rd to 8th percentile	Well-Below Average
2nd percentile and below	Lower Extreme

National Norm-Referenced Interpretations: Comparing Students in a Larger Context

National norms are available from Acadience Data Management. National norms allow a school or district to compare a student's performance to other students across the nation. A disadvantage of system-wide norms is that they may not be representative of the characteristics of students in a particular district. For example, a local district may have a very high proportion of English language learners. While the national norms may include English language learners, the proportion may or may not be representative of the local district. It is important for district and school leaders to obtain information about the norm sample and assess its relevance to their particular demographic prior to making decisions about students or overall district performance.

The primary value of national normative information is to provide an alternative perspective on student performance. When the national norms are based on a large and nationally representative sample of students, they can provide an indication of national student achievement in early reading. For instance, if 120 words correct on ORF at the end of third grade is at the 50th percentile in local district norms and is at the 60th percentile on national norms, then the average achievement in the district is above the national average. Similarly, at an individual student level,

a student might be at the 55th percentile compared to local norms but might be at the 5th percentile compared to national norms. In this context, the student might appear to be making adequate progress, but the national normative information clarifies that the student is still of concern in a larger context. Considering local norms and national norms can provide a balanced perspective on the student's skills and needs.

For more information about national norms, see:

Gray, J. S., Warnock, A. N., Kaminski, R. A., & Good, R. H. (2018). *Acadience Reading National Norms 2014–2015* (Technical Report No. 23). Eugene, OR: Dynamic Measurement Group. Available: https://acadiencelearning.org/.

The Importance of Response Patterns

In addition to interpreting scores from a criterion-referenced, individually referenced, local norm-referenced, or system-wide norm-referenced perspective, the pattern of behavior that the student displays on the assessment is also important. Acadience Reading measures are designed to be indicators of basic early literacy skills. If the student achieves a score above the benchmark goal but does so in a way that indicates that the early literacy skill has not been mastered, the student may still need additional support to be on track. For example, if a student reaches the benchmark goal on Phoneme Segmentation Fluency (PSF) but does so by rapidly segmenting words in an onset-rime pattern (/m/ /ap/, /str/ /eat/), that student may not be as likely to reach the next goal as a student who achieves the benchmark goal by correctly segmenting phonemes (/m/ /a/ /p/, /s/ /t/ /r/ /ea/ /t/) (See Appendix B on page 135 for a pronunciation guide that shows how individual phonemes are represented on PSF). For this reason, each measure includes a checklist of common, instructionally relevant response patterns. Teachers and other specialists who interpret Acadience Reading results to provide instruction for students should review the types of responses for students in their classes. This information, in addition to the raw scores, can dramatically guide instructional strategies.

How Does Acadience Reading K–6 Improve on Earlier Versions of These Measures?

Empirically equated oral reading passages. All oral reading passages went through an extensive readability analysis and field-testing with actual students. Based on this empirical testing, the best-performing passages (in terms of reliability and comparability in student results) were selected for inclusion in Acadience Reading and then organized in triads in such a way as to ensure that student performance was comparable.

Materials designed for ease of use. Measures were explicitly designed and field-tested such that they can be administered and scored with ease. Wait rules, discontinue rules, and reminder prompts are embedded into the administration directions. Scoring booklets are large enough to be easily readable, and an early-reader font is used for kindergarten through second-grade materials.

Empirically field-tested directions. All of the directions that are read to the student and the reminder prompts were designed and tested so that they are explicit and facilitate student understanding of the task.

Stratification. A stratified random sampling procedure was used to improve the equivalence of the forms and to more evenly distribute items of different difficulty. This procedure increases the consistency of scores from one form to another. With stratified random sampling, items of similar difficulty appear in the same places on every form. For example, on NWF there were six difficulty/word-type categories that were distributed by design identically on each form. For instance, the first item is always an easier item, a word

with a three-letter CVC pattern where both consonants occur frequently in English. For each form, the actual test items were then randomly selected from the appropriate category.

Response patterns. Measures include lists of common response patterns that the assessor can mark to help in planning instruction. These lists are located within the scoring booklets for better accessibility.

Table 1.5 below summarizes the key features of the Acadience Reading measures.

Table 1.5 Key Features of Acadience Reading Measures

Measures	Description
First Sound Fluency (FSF)	• FSF provides an early indicator of phonemic awareness. FSF is easy to administer and eliminates concerns related to the use of pictures when assessing initial sounds. FSF includes production items with continuous timing. • Stratification of test items based on whether the word begins with a continuous sound, a stop sound, or a blend. • Explicit directions and reminders to facilitate student understanding of the task.
Letter Naming Fluency (LNF)	• Materials with integrated reminders to enhance the administration of the measure. • Font that is familiar to younger children. • Stratification of test items to increase equivalence and consistency of scores from one form to another. • Explicit directions and reminders to facilitate student understanding of the task. • A checklist of common response patterns to facilitate linkages to instruction.
Phoneme Segemntation Fluency (PSF	• Materials with integrated reminders to enhance the administration of the measure. • Score form layout that facilitates scoring. • Stratification of test items to increase equivalence and consistency of scores from one form to another. • Explicit directions and reminders to facilitate student understanding of the task. • A checklist of common response patterns to facilitate linkages to instruction.
Nonsense Word Fluency (NWF)	• Materials with integrated reminders to enhance the administration of the measure. • In addition to scoring for Correct Letter Sounds (CLS), scoring for Whole Words Read (WWR) to measure the critical target skill of reading the words as whole words. • Font is familiar to younger children. • Stratification of test items to increase equivalence and consistency of scores from one form to another. • An even distribution of vowels, with each row of five items including one word with each vowel. • Explicit directions and reminders facilitate student understanding of the task and clarify that the preferred responses are whole words. The student is permitted to provide individual letter sounds or to sound out the word while learning the skills. • A checklist of common response patterns to facilitate linkages to instruction.

Table 1.5 Key Features of Acadience Reading Measures, cont.

Measures	Description
Oral Reading Fluency (ORF)	• Field-tested empirically equated passages with consistent difficulty within each grade level. • Materials with integrated reminders to enhance the administration of the measure. • Font is more familiar to younger children in first- and second-grade passages. • Explicit directions and reminders to facilitate student understanding of the task. When administering three passages during benchmark assessment, shortened directions are provided for the second and third passages to increase efficiency. • A checklist of common response patterns to facilitate linkages to instruction.
Retell	• Included as a component of the Oral Reading Fluency measure to indicate that the end-goal of reading is to read for meaning. • Materials with integrated reminders to enhance the administration of the measure. • Explicit directions and reminders to facilitate student understanding of the task. • A checklist of common response patterns to facilitate linkages to instruction.
Maze	• Maze provides an added indicator of comprehension in grades 3 through 6. • Can be administered in groups or individually. • Explicit directions and reminders to facilitate student understanding of the task.
Word Use Fluency–Revised (WUF-R)	• Available as an experimental measure. (Email info@acadiencelearing.org for more information)

History and Development of Acadience Reading K–6 Research and Development

Initial research and development of the Acadience Reading measures[1] was conducted in the late 1980s and early 1990s. The Acadience Reading program of research built on the measurement procedures from Curriculum-Based Measurement, or CBM (e.g., Deno & Mirkin, 1977; Deno, 1985; Deno & Fuchs, 1987), and General Outcome Measurement, or GOM (Fuchs & Deno, 1991). The Acadience Reading measures were designed to be economical and efficient indicators of a student's progress toward achieving a general outcome such as reading or phonemic awareness, and to be used for both benchmark assessment and progress monitoring.

Initial research on these measures focused on examining their technical adequacy for these primary purposes (Good & Kaminski, 1996; Kaminski & Good, 1996). The early versions of the measures authored by Roland Good and Ruth Kaminski were first published under the name DIBELS® in 2002. Since then, the measures have gained widespread use for monitoring progress in acquisition of early literacy skills. Prior to 2002, these measures were made available to research partners. An ongoing program of research over the past three decades has

continued to document the reliability and validity of the Acadience Reading measures as well as their sensitivity in measuring changes in student performance over time.

Acadience Reading is the result of an expanding knowledge base in the fields of reading and assessment, continuing research and development, and feedback from users of these assessments. From 2006 to 2010, initial research and field-testing of the Acadience Reading measures occurred in 90 schools across the United States. A series of studies over that time period examined the reliability, validity, and utility of the measures. From 2010 to 2018, the measures underwent continued validation and refinement. See this manual for a description of the technical adequacy data on Acadience Reading. Additional technical adequacy data are also available on our website under Publications and Presentations (https://acadiencelearning.org/).

---

[1]Acadience™ Reading K–6 is the new name for the DIBELS Next® assessment. Acadience is a trademark of Dynamic Measurement Group, Inc. (DMG). The DIBELS Next copyrighted content is owned by DMG. The DIBELS® and DIBELS Next registered trademarks were sold by DMG to the University of Oregon (UO) and are now owned by the UO.

Chapter 2: **Acadience Reading K–6 Measures—Descriptions, Design Specifications, and Content Validity**

In this chapter, we describe each of the Acadience Reading measures and how they were constructed. The design specifications in this chapter relate directly to the content validity of the measures. The Acadience Reading Benchmark Administration Timeline (Figure 2.1 on page 20) shows the measures that are administered at each benchmark assessment period.

First Sound Fluency

Grade: Kindergarten

Indicator of: Phonemic Awareness

First Sound Fluency (FSF) is a brief, direct measure of a student's fluency in identifying the initial sounds in words. The ability to isolate the first sound in a word is an important phonemic awareness skill that is highly related to reading acquisition and reading achievement (Yopp, 1988). The ability to isolate and identify the first phoneme in a word is an easier skill than segmenting all the sounds in words or manipulating phonemes in words, thus FSF is used as a measure of developing phonemic awareness at the beginning and middle of kindergarten.

Using standardized directions, the assessor says a series of words one at a time to the student and asks the student to say the first sound in the word. On the scoring page, the assessor circles the corresponding sound or group of sounds the student says. Students receive either 2 points for saying the initial phoneme of a word (e.g., saying the /s/ sound as the first sound in the word *street*) or 1 point for saying the initial consonant blend, consonant plus vowel, or consonant blend plus vowel (e.g., /st/, /str/, or /strea/ for street). A response is scored as correct as long as the student provides any of the correct responses listed for the word. The total score is the sum of the points the student receives in 1 minute.

To make FSF more sensitive for use with young students, the measure uses differential scoring, which allows students to receive partial credit for demonstrating beginning skills in phonemic awareness. A student who may not be able to isolate the initial phoneme would receive partial credit for providing the first group of sounds in the word, showing emerging understanding that words are made up of sounds. Although partial credit is given, the goal is for the student to be able to correctly say the first phoneme of each word. This is the preferred response and is given the most points.

To ensure that students understand the task and to maximize the performance of young students who may not have had any prior exposure to instruction in phonemic awareness, three practice items are included. The practice items provide increasing levels of support, including modeling (e.g., "listen to me say…") and leading the correct response (e.g., "say it with me"). By design, the first two practice items start with the same sound, /m/. In the first practice item, isolation of the /m/ sound at the beginning of a word is modeled. In the second practice item, the student is asked to isolate the beginning sound in a word that also starts with /m/. In the third practice item, the student is asked to generalize the skill of isolating beginning sounds to a word that does not start with /m/.

Figure 2.1 Acadience Reading Benchmark Administration Timeline
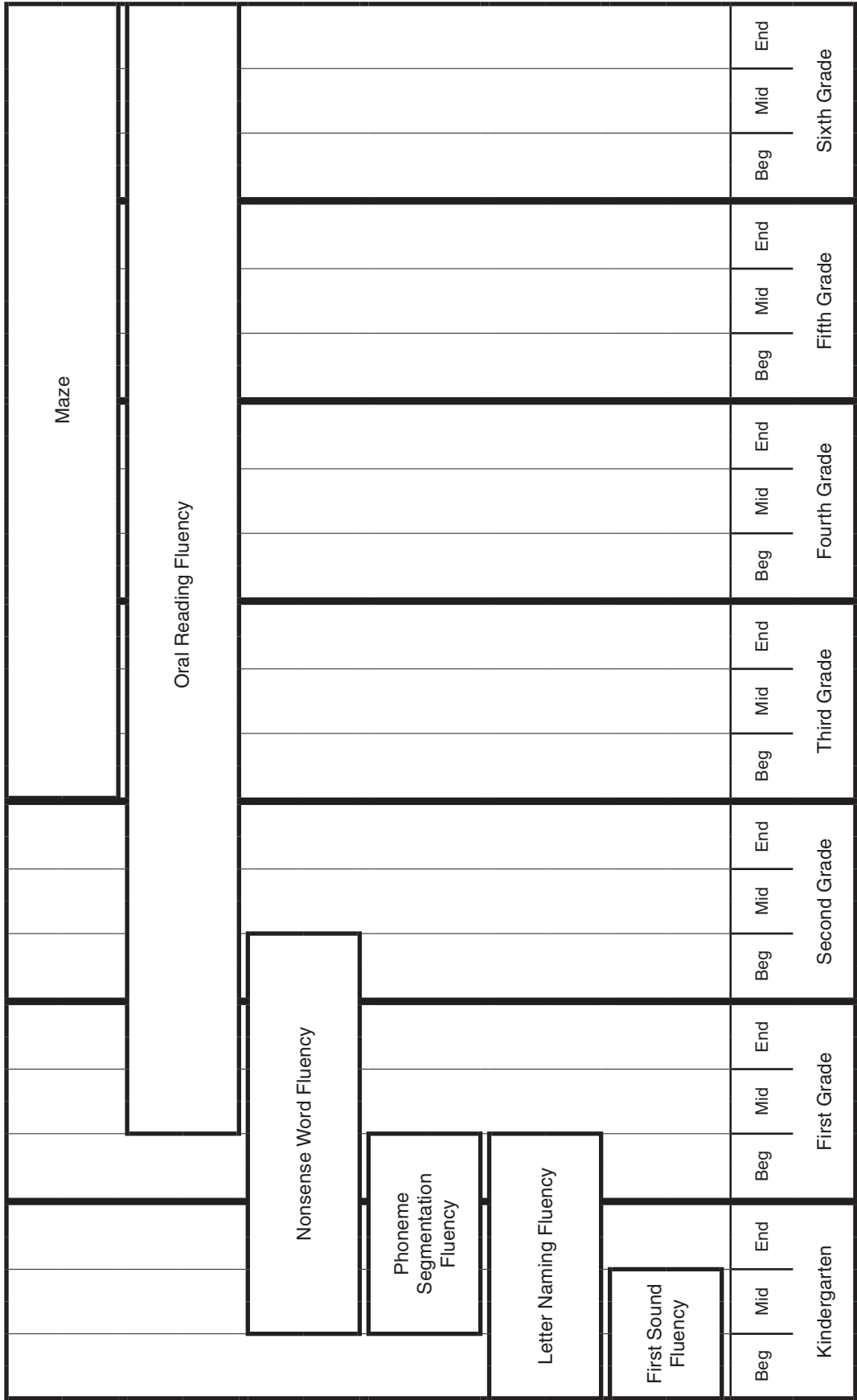
Test Construction

Items for all FSF forms were selected from a word pool consisting of single-syllable words. Initial work on this word pool was derived from a study of preschool measures of early literacy (Kaminski, Baker, Chard, Clarke, & Smith, 2006). Words were excluded if they were deemed inappropriate (e.g., *rob, knife*) or if they began with the initial phonemes /b/, /d/, /p/, or /g/ followed by the /u/ sound (e.g., *duck*), as such words cannot be scored differentially due to confusion with the schwa sound. The final word pool consisted of 861 words, three of which were used as example items and so do not appear as test items. The words were then broken into three difficulty levels:

Difficulty Category	Number and Percent of Items per Form	Total Items in Word Pool
Initial continuous sound (e.g., /s/, /m/) followed by a vowel sound	23%, seven items per form	234
Initial stop sound (e.g., /b/, /t/) followed by a vowel sound	27%, eight items per form	265
Initial blend (e.g., /st/)	50%, 15 items per form	362

Each form consists of 30 items. Before creating the individual forms, a stratified sequence of the different difficulty categories was developed. Of the 30 items in the sequence, the first 28 items were divided into seven groups of four. Each group of four included one word with an initial continuous sound, one word with an initial stop sound, and two words with an initial blend. Within the groups of four, the order of the categories was randomized, except for the first group, which started with an initial continuous sound, then an initial stop sound, then two words with blends. The 29th category in the sequence was a word with an initial stop sound, and the 30th category was a word that started with a blend. Once the sequence was determined, that stratification was applied to all forms so that the same difficulty categories appear in the same locations on every form.

Each word on a form was then randomly selected from the words that matched the specified difficulty category.

Letter Naming Fluency

Grade: Kindergarten–First Grade
Indicator of: Not directly linked to a basic early literacy skill

Letter Naming Fluency (LNF) is a brief, direct measure of a student's fluency with naming letters. LNF assesses a student's ability to recognize individual letters and say their letter names. Using standardized directions, the assessor presents a page of upper- and lower-case letters arranged in random order and asks the student to name the letters. The assessor marks letter names that are read incorrectly or skipped. The total score is the number of correct letter names that the student says in one minute.

The purpose of LNF is to measure students' automaticity with letter naming. Fluency in naming letters is a strong and robust predictor of later reading achievement (Adams, 1990). All letters are included on the LNF materials, but they appear in random order.

LNF is an indicator of risk rather than an instructional target. While the ability to recognize and name letters in preschool and at the beginning of kindergarten is a strong predictor of later reading achievement (e.g., Badian, 1995; Walsh, Price, & Gillingham, 1988), studies have failed to show that teaching letter names to students enhances their reading ability (e.g., Ehri, 1983) and, in fact, have demonstrated that successful learning of

letter-sound correspondences that leads to reading acquisition can occur without knowledge of letter names (Bruck, Genesee, & Caravolas, 1997; Mann & Wimmer, 2002). Because learning letter names is not a powerful instructional target, benchmark goals are not provided for LNF. LNF is a strong predictor, however, so it is included as a part of the Reading Composite Score in kindergarten and early first grade.

Test Construction

All upper- and lower-case letters in the English alphabet were used. The 26 upper-case and 26 lower-case letters were divided into three categories based on relative difficulty, with 18 letters in the easy category and 17 letters each in the medium and hard categories. A randomly selected letter from the easy category was used as the first test item, and then 17 triads were constructed, with a triad including one randomly selected letter from each category: easy, medium, and hard. The ordering of letters by triads of easy, medium, and hard letters was done to more evenly space the difficulty levels. The first triad was placed with the easy letter first, the medium letter second, and the hard letter third. For the other 16 triads, the order of the difficulty categories was randomized within the triad. The process was then repeated, to include another set of 26 upper-case and 26 lower-case letters, providing 104 test items. Displaying a full set of 52 letters (26 upper-case and 26 lower-case) first, and then displaying another full set of 52 letters meant that the same letter would not appear in close proximity. The only difference in procedure for the second set of 52 letters was that the order of difficulty categories in the first triad were also randomized. The letters were displayed in 11 rows of 10 letters each. To prevent the last row from only having four letters, the first six letters from the beginning of the form were repeated at the end of the form, for a total of 110 test items.

Each form was constructed using the same process, but the location of the difficulty categories was re-randomized each time.

Phoneme Segmentation Fluency

Grade: Kindergarten–First Grade
Indicator of: Phonemic Awareness

Phoneme Segmentation Fluency (PSF) is a brief, direct measure of phonemic awareness. PSF assesses the student's fluency in segmenting a spoken word into its component parts or sound segments. Using standardized directions, the assessor says a word and asks the student to say the sounds in the word. The assessor underlines each correct sound segment of the word that the student says. A correct sound segment is any different, correct part of the word the student says. The total score is the number of correct sound segments that the student says in one minute. For example, if the assessor says the word *fish* and the student says /f/ /i/ /sh/, the student has completely and correctly segmented the word into all of its component sounds and the score is three correct sound segments. If the student says /f/ /ish/, the score is two correct sound segments.

Partial credit is given for partial segmentation. A student who is developing phonemic awareness may not yet segment words completely into individual sounds but may segment parts of words. For example, a student who says the first sound of the word *sun* (/s/) receives 1 point. A student who says the onset and rime (/s/ / un/) receives 2 points and a student who completely and correctly segments all of the individual phonemes in the word (/s/ /u/ /n/) receives 3 points. Note that consonant blends have two or more phonemes that should be produced separately for a student to receive full credit. For example, for the word *trap,* a student who says /tr/ /a/ /p/ receives partial credit of 3 points, and one who says /t/ /r/ /a/ /p/ receives the full 4 points. Allowing partial credit in scoring increases the sensitivity of the measure, thus making it possible to measure growth

from partial to complete segmentation. Although partial credit is given, the preferred response is for students to completely segment words at the phoneme level by the end of kindergarten.

Test Construction

The word pool for Phoneme Segmentation Fluency comes from *The Educator's Word Frequency Guide* (Zeno, Ivens, Millard, & Duvvuri, 1995), where either the first or second grade U value (the relative frequency of occurrence) was 20 or higher. Words were then excluded if they were not found in the *Oxford Advanced Learner's Dictionary* (Hornby, Wehmeier, McIntosh, & Turnbull, 2005), were proper nouns, had more than one syllable, had a single phoneme, had six or more phonemes, included apostrophes, or were inappropriate. The final word pool included a total of 1132 items, three of which were used as example items and so do not appear as test items. The words were then broken into four difficulty levels:

Difficulty Category	Number and Percent of Items per Form	Total Items in Word Pool
Easiest—no r-controlled vowels, no consonant blends, two or three phonemes	67%, 16 items per form	501
Less Easy—One difficulty feature consisting of an r-controlled vowel or a single, two-consonant blend, but not both; no three-consonant blends; two to four phonemes	25%, six items per form	491
More Difficult—two difficulty features; no three-consonant blends; two to four phonemes	4%, one item per form	30
Most Difficult—three-consonant blends or five phonemes	4%, one item per form	110

Each form consists of 24 items. Before creating the individual forms, a stratified sequence of the different difficulty categories was developed. The order of appearance of the "Easiest" and "Less Easy" categories was random, except the first two items on a form were selected from the "Easiest" category. Since only one item each from the "More Difficult" and "Most Difficult" categories appeared on each form, the "More Difficult" category was randomly placed in the first half of the form, and the "Most Difficult" category was randomly placed in the second half of the form. Once the sequence was determined, that stratification was applied to all forms, so that the same difficulty categories appear in the same locations on every form. The item stratification used for PSF ensures that every form has the same number of items from each difficulty category, and that those difficulty categories will appear in the same place on every form.

Each word on a form was then randomly selected from the words that matched the specified difficulty category.

Nonsense Word Fluency

Grade: Kindergarten–Second Grade

Indicator of: Alphabetic Principle and Basic Phonics

Nonsense Word Fluency (NWF) is a brief, direct measure of the alphabetic principle and basic phonics. It assesses knowledge of basic letter-sound correspondences and the ability to blend letter sounds into consonant-vowel-consonant (CVC) and vowel-consonant (VC) words. The test items used for NWF are phonetically regular make-believe (nonsense or pseudo) words. To successfully complete the NWF task, students must rely on their knowledge of letter-sound correspondences and how to blend sounds into whole words. One reason that nonsense word measures are considered to be a good indicator of the alphabetic principle is that "pseudowords have no lexical entry, [and thus] pseudo-word reading provides a relatively pure assessment of students' ability to apply grapheme-phoneme knowledge in decoding" (Rathvon, 2004, p. 138).

Following a model and a practice item, the student is presented with a sheet of randomly ordered VC and CVC nonsense words (e.g., *dif, ik, nop*). Standardized directions are used to ask the student to read the make-believe words the best they can, reading either the whole word or saying any sounds they know. For example, if the stimulus word is *tof*, the student could say /t/ /o/ /f/ or "tof." The assessor underlines each correct letter sound produced either in isolation or blended together. Whole words read without sounding out are underlined in their entirety.

There are two separate scores reported for NWF:

1. Correct Letter Sounds (CLS) is the number of letter sounds produced correctly in one minute. For example, if the student reads *dif* as /d/ /i/ /f/, the score for Correct Letter Sounds is 3. If the student reads dif as /di/ /f/ or "dif," the CLS score is also 3.

2. Whole Words Read (WWR) is the number of make-believe words read correctly as a whole word without first being sounded out. For example, if the student reads *dif* as "dif," the score is 3 points for CLS and 1 point for WWR, but if the student reads dif as "/d/ /i/ /f/ dif," the score is 3 points for CLS but 0 points for WWR.

The goal is for students to read whole words on NWF; however, an advantage of NWF is that it allows for monitoring the development of the alphabetic principle and basic phonics as early as the middle of kindergarten, when producing individual letter sounds is the more common response.

Test Construction

The word pool for Nonsense Word Fluency consists of CVC (consonant-vowel-consonant) and VC (vowel-consonant) nonsense words. The letters "q" and "x" were not used, since they typically represent more than one phoneme. The letters "h", "w", "y", and "r" were used only in the initial position, and the letters "c" and "g" were used only in the final position. Real words and words that sounded like inappropriate words were excluded, but words that sounded like real words were not excluded. The words were generated automatically in Microsoft Excel, and the excluded words were identified manually. The final word pool included a total of 1,017 items, two of which were used as example items and so do not appear as test items. The words were then divided into six difficulty categories based on the pattern (CVC and VC) and on the relative difficulty of the consonants. The consonants judged to be easier were b, c, d, f, g, h, k, l, m, n, p, r, s, and t. Letters were judged to be easier if they appear more often in words, since students will see them more often and many curricula teach higher frequency letters first.

The categories were:

Difficulty Category	Number and Percent of Items per Form	Total Items in Word Pool
VC, Easy Consonant	10%, five items per form	44
VC, Hard Consonant	4%, two items per form	11
CVC, First Consonant Easy	20%, 10 items per form	163
CVC, Last Consonant Easy	20%, 10 items per form	247
CVC, Both Consonants Easy	40%, 20 items per form	483
CVC, Both Consonants Hard	6%, three items per form	69

Each form consists of 50 items. Before creating the individual forms, a stratified sequence of the different difficulty categories was developed. For categories with 10 items on a form, one item appeared on each of the 10 rows. For the category with 20 items on a form, two items appear on each of the 10 rows. The other categories were randomly distributed across the rows. Within a row, the order of the difficulty categories was random, except the first two items on a form were selected from two of the easier categories (CVC with both consonants easy, and CVC with the first consonant easy). Once the sequence was determined, that same stratification was applied to all forms, so that the same difficulty categories appear in the same locations on every form. This stratification process ensures that every form has the same number of items from each difficulty category and that those difficulty categories will appear in the same place on every form.

In addition to the stratification of the difficulty categories, each row of five items includes one nonsense word with each of the five vowels, in random order. The order of the vowels was re-randomized for each row and each form.

Each word on a form was then randomly selected from the words that matched both the specified difficulty category and the specified vowel.

Oral Reading Fluency

    Grade: First Grade–Sixth Grade

    Indicator of: Advanced Phonics and Word Attack Skills

                         Accurate and Fluent Reading of Connected Text

                         Reading Comprehension

Oral Reading Fluency (ORF) is a measure of advanced phonics and word attack skills, accurate and fluent reading of connected text, and reading comprehension. The ORF passages and procedures are based on the program of research and development of Curriculum-Based Measurement of reading by Stan Deno and colleagues at the University of Minnesota (Deno, 1989). There are two components to ORF. The first part is oral reading fluency and the second part is passage retell. For the oral reading fluency component, students are given an unfamiliar, grade-level passage of text and asked to read for 1 minute. Errors such as substitutions, omissions, and hesitations for more than 3 seconds are marked while listening to the student read aloud. For benchmark assessment, students are asked to read three different grade-level passages for 1 minute each. The scores are the median number of words read correctly and the median number of errors across the three passages. Using the median score from three passages gives the best indicator of student performance over a range of different text and content.

The student's accuracy rate is calculated based on the number of words read correctly and the number of errors, using the following formula:

$$Accuracy = 100 \times \frac{median\ words\ correct}{median\ words\ correct + median\ errors}$$

The passage retell component follows the reading of each passage, provided that the student has read at least 40 words correct per minute on that passage, or if the assessor feels it is otherwise appropriate. Passage retell is intended to provide a comprehension check for the ORF assessment, and provides an indication that the student is reading for meaning. With a prompted passage retell, it is clear to the student that the intent is to read for meaning. Speed-reading without attending to text comprehension is undesirable and will be apparent when the student is asked to tell about what they have read.

Case studies have documented students who can read words but not comprehend what they read (Dewitz & Dewitz, 2003). There is concern that students who display similar reading behavior will not be identified without a comprehension check. Passage retell provides both a valuable indicator of reading comprehension as well as an efficient procedure to identify those students who are not able to talk about what they have just read. Inclusion of passage retell also explicitly instructs students to be reading fluently for meaning. The quality of a student's retell provides valuable information about overall reading proficiency and oral language skills.

During Retell, the student is asked to tell about what he/she has read. The assessor indicates the number of words in the Retell that are related to the story by drawing a line through a box of numbers. Following a hesitation of 3 seconds, students are prompted to tell as much as they can about the story. If the student hesitates again for 5 seconds or longer, or if the student is clearly responding for 5 seconds in a way that is not relevant to the passage, the task is discontinued. The assessor must make a judgment about the relevance of the Retell to the story while drawing the line. A quality of response rating allows the assessor to make a qualitative rating of the student's response after the completion of the Retell. The rating should be based on how well the student retold the portion of the passage that he/she read.

Test Construction

The ORF passages were designed to represent the different types of text that students will encounter, including a mix of narrative and expository, with different types of passages and content within those categories. A range of topics and themes was selected so that each student would encounter familiar topics and unfamiliar topics. The passages were designed to be authentic text, so they include irregular words and are not written entirely in decodable text. Passages were written and revised by professional authors according to the design specifications below.

General Passage Design Specifications for Authors and Editors

1. Passages should have a beginning, middle, and end.

2. In narrative passages, proper names should be simple and decodable according to basic phonics rules. Names in first through third grade passages should be no more than two syllables. Names in fourth through sixth grade passages should generally not be more than three syllables. Names should represent diverse cultural, racial, and ethnic groups. In expository passages, avoid unnecessary proper names. Proper names in expository passages should generally appear in grades where those names match the criteria for narrative passages.

3. Passages should be engaging in the first paragraph.

4. Passages should be gentle, positive, and friendly, modeling positive pro-social behaviors (without being preachy). For example, if a passage is about bike riding, the subject should wear a helmet. Characters should try to be friendly. Conflict should be minimized and reduced, not escalated, e.g., siblings should cooperate.

5. Passages should be sensitive and respectful to all groups and subgroups.

6. Diversity should frequently be incorporated incidentally into passages, including issues of diversity in terms of socio-economic status, disability, race, ethnicity, family structure, background, culture, urban and rural settings, etc.

7. Passages should be grammatically correct, with mature phrasing and conventional sentence structure. Avoid colloquialisms, slang, dialect, and creative or unusual sentence structures.

8. Avoid extensive dialogue.

9. Passages should flow rather than being abrupt and staccato. They should follow an easy and engaging sequence. Avoid lists of things, e.g., "I like strawberry, chocolate, peach, and cherry."

10. Avoid repetitive sentence structures, e.g., "He would do this. He would do that. He would do something else."

11. Avoid sad or frightening topics such as natural disasters or third-degree burns.

12. The initial passage set should have a mix of about 40% expository and 60% narrative for first through third grades, and about 60% expository and 40% narrative for fourth through sixth grades.

13. Passages must be factually correct.

14. The first word of the title should not be the same as the first word of the passage.

15. All passages must meet readability criteria for the grade level as measured by the DMG Passage Revision Utility, which is software that identifies the target word length, rare words, and sentence length for a passage and provides guidance when a passage is outside of the target ranges specified by the DMG Passage Difficulty Index.

DMG Passage Difficulty Index

The DMG Passage Difficulty Index was developed to address our concerns with other readability formulas for developing oral reading fluency passages. Readability formulas commonly use one or two indicators of passage difficulty that represent the (a) decoding difficulty of words in the passage, where longer words are more difficult to decode; (b) semantic difficulty of words in the passage, where passages with many low-frequency words or a high proportion of rare words are more difficult to read; and (c) syntactic difficulty of sentences, where longer sentences generally result in a more difficult passage to read. Selected examples of indicators in each area are provided in Table 2.1. Common readability formulas and the indicators they incorporate are summarized and compared with the DMG Passage Difficulty Index in Table 2.2. The readability formulas summarized use indicators in one or two areas of passage difficulty.

Table 2.1 Indicators of Passage Difficulty Frequently Incorporated in Readability Formulas

Indicator Number	Areas of Passage Difficulty
Decoding Difficulty—Word Length	
1	Characters per word
2	Proportion of words with seven or more characters
3	Syllables per word
4	Proportion of words with two or more syllables
5	Proportion of words with three or more syllables
Semantic Difficulty or Word Exposure—Rare Words or Word Frequency	
6	Word frequency
7	Proportion of rare words (words not found on a word list)
Syntactic Difficulty or Sentence Complexity – Sentence Length	
8	Words per sentence
9	Number of syllables per sentence

Table 2.2 Indicators of Passage Difficulty Incorporated in Selected Readability Formulas and DMG Passage Difficulty Index

	Decoding difficulty indicator					Semantic difficulty indicator		Syntactic difficulty indicator	
	1	2	3	4	5	6	7	8	9
SMOG					X				
Forcast				X					
Fry								X	X
Lexile						X		X	
Dale-Chall							X	X	
Spache							X	X	
Flesch			X					X	
FOG					X			X	
Powers			X					X	
DMG Passage Difficulty Index	X	X	X		X		X	X	

Because all three areas (word difficulty, semantic difficulty, and syntactic difficulty) are defensible in estimating passage difficulty, the DMG Passage Difficulty Index utilizes information from all three areas. Many readability formulas combine two indicators and provide a single result, which means that the individual indicators are not examined in isolation. Consequently, a passage may be at a third-grade level according to the Spache readability index (for example) by offsetting more difficult words with shorter (and thus easier) sentences, or vice versa. The DMG Passage Difficulty Index examines all three aspects of passage difficulty (word difficulty, semantic difficulty, and syntactic difficulty) in isolation, to ensure that each indicator is within a specified range for the grade level, as well as providing an overall composite of the three indicators that also must be in the specified range for the grade level.

For decoding difficulty, four measures of word length were used: (a) characters per word, (b) percent of words with three or more syllables, (c) percent of words with seven or more characters, and (d) number of syllables per word. The four measures were scaled to be equally weighted and averaged to provide a composite measure of word length for the passage. The median words per sentence provided a measure of the syntactic difficulty of the passage, and percent of unique rare words provided a measure of the semantic difficulty of the passage. The composite measure of decoding difficulty, the measure of syntactic difficulty, and the measure of semantic difficulty were equally weighted and averaged to obtain the DMG Passage Difficulty Index.

The target mean DMG Passage Difficulty Index and target mean for each component measure for each grade are reported in Table 2.3 and the specified ranges of acceptable difficulty for each component and for the overall index are reported in Table 2.4. The target mean DMG Passage Difficulty Index was specified such that the Acadience Reading ORF passages would be approximately equivalent in difficulty to the DIBELS 6th Edition passages for each grade level. The ranges of acceptable difficulty were specified to remove overlap in difficulty between adjacent grades for the overall DMG Passage Difficulty Index and to reduce overlap for each component of the index.

Table 2.3 Target Passage Length and Means for DMG Passage Difficulty Index and Component Measures by Grade

Grade Level	Passage length in words	Target component index values			
		DMG Passage Difficulty Index	Word length composite measure	Proportion of rare words measure	Median words per sentence measure
1	200 to 250	-1.25	-1.25	-1.25	-1.25
2	225 to 275	-0.79	-0.79	-0.79	-0.79
3	250 to 300	-0.20	-0.20	-0.20	-0.20
4	300 to 350	0.32	0.32	0.32	0.32
5	300 to 350	0.69	0.69	0.69	0.69
6	300 to 350	1.10	1.10	1.10	1.10

Note. Ranges are z-scores based on means and standard deviations from all 6th Edition benchmark and progress monitoring passages. Grade-level target values are based on 6th Edition benchmark passages for that grade.

Table 2.4 Design Specifications for DMG Passage Difficulty Index and Component Measures of Passage Difficulty

Grade Level	DMG Passage Difficulty Index		Word length composite measure		Proportion of rare words measure		Median words per sentence measure	
	Minimum	Maximum	Minimum	Maximum	Minimum	Maximum	Minimum	Maximum
1	-1.43	-1.07	-1.75	-0.75	-1.75	-0.75	-1.65	-0.85
2	-0.97	-0.54	-1.29	-0.29	-1.29	-0.29	-1.19	-0.39
3	-0.44	0.01	-0.70	0.30	-0.70	0.30	-0.60	0.20
4	0.11	0.45	-0.18	0.82	-0.18	0.82	-0.08	0.72
5	0.55	0.84	0.19	1.19	0.19	1.19	0.29	1.09
6	0.94	1.25	0.60	1.60	0.60	1.60	0.70	1.50

Passage Selection

The initial passage set included 40 passages for each grade that met the criteria above. While the passage design specifications and the DMG Passage Difficulty Index were designed to reduce variability of passages within a grade, they still do not measure everything about a passage that makes it more or less difficult for a student to read. A readability study was conducted to examine actual student performance on all of the passages and further control differences in passage difficulty within each grade level. See Study D in Chapter 3 for a description of the project.

For first grade, a final set of 26 passages was needed: three each for the middle- and end-of-year benchmark assessments, 20 for progress monitoring, and three for Acadience Reading Survey (published as Acadience Reading Survey). For second through sixth grades, a final set of 29 passages was needed for each grade: three each for the beginning-, middle-, and end-of-year benchmark assessments, 20 for progress monitoring, and three for Acadience Reading Survey. In Study D, each participating student was administered all 40 passages in the initial passage set in a different, randomized order. All passages were read over two or three weeks, and a linear growth line was fit to represent each student's overall rate of progress. For each student and each passage, a residual was calculated by subtracting the predicted score (based on the student's rate of progress) from the actual score. In this way, order and growth effects were removed from consideration of relative passage difficulty.

The primary purpose of the readability study was to identify the passages within each grade level that were the most reliable, most valid, and most consistent in level of difficulty. Based on the results of the readability study, the 32 best passages (29 in first grade) were selected for inclusion in Acadience Reading and Acadience Reading Survey, based on eight factors, including the smallest average residuals (differences from the predicted scores), standard deviation of the residuals, and alternate-form reliability.

Once the passages to be used in Acadience Reading were identified, they were arranged in designed triads of three passages. Each triad included a slightly easier, medium, and slightly harder passage based on the average residual of the passage from Study D. The triads were designed so that the median score (ORF Words Correct) of each triad was very close to the medians of the other triads, as well as to the overall grade-level mean. Finally, triads in each grade were assigned to benchmark assessment, Acadience Reading Survey, and progress monitoring.

The process used for the Acadience Reading Oral Reading passages differs significantly from that used with an earlier version of these measures, in which passages were initially written and developed according to the Spache readability formula and then assigned to role and sequence based on the average of nine readability formulas. A readability study was not conducted to empirically level those earlier passages. Individual passage variations sometimes were larger than desired. In contrast, Acadience Reading ORF results are substantially more consistent and stable from passage to passage and triad to triad.

For a full description of the readability study, passage selection, and passage arrangement, and to view the results of the DMG Passage Difficulty Index, see:

Powell-Smith, K. A., Good, R. H., III, & Atkins, T. (2010). *DIBELS Next Oral Reading Fluency Readability Study*[1] (Technical Report No. 7). Eugene, OR: Dynamic Measurement Group. Available: http://acadiencelearning.org/.

Maze

Grade: Third Grade–Sixth Grade
Indicator of: Reading Comprehension

Maze is the standardized, Acadience Reading version of a maze testing procedure for measuring reading comprehension. The purpose of a maze assessment is to measure the reasoning processes that constitute comprehension. Specifically, Maze assesses the student's ability to construct meaning from text using comprehension strategies, word recognition skills, background information and prior knowledge, familiarity with linguistic properties such as syntax and morphology, and reasoning skills.

Maze can be given to a whole class at the same time, to a small group of students, or individually. Students are given a passage where approximately every seventh word has been replaced by a box containing the correct word and two distractor words. Using standardized directions, students are asked to read the passage silently and circle their word choices. The student receives credit for selecting the word that best fits the omitted word in the reading passage. The scores that are recorded are the number of correct and incorrect responses. An adjusted score, which compensates for guessing, is calculated based on the number of correct and incorrect responses.

Maze Adjusted Score = number of correct responses – (number of incorrect responses ÷ 2).

The result of the formula is rounded to the nearest whole number, with half-points (0.5) rounded up. If the adjusted score is negative, a score of 0 is recorded. Since there are three possible responses for each item, the adjusted score expected for a student who guesses on every item is 0. Scores are not prorated. The maximum score a student can receive is equal to the number of items on the form.

Test Construction

Maze passages were written according to the same specifications as Acadience Reading ORF passages, except for passage length which was longer for Maze than for ORF. Maze passages were leveled using the DMG Passage Difficulty Index.

Table 2.5 Maze Passage Length

Grade Level	Maze Passage Length in Words
3	350 to 400
4	400 to 450
5	450 to 500
6	500 to 550

A maze procedure was then applied to each passage. The Acadience Reading maze procedure left the first sentence unchanged. Starting with the second sentence, approximately every seventh word was selected to be replaced by a multiple choice box containing the original, correct word and two distractor words, in randomized order. Certain words, such as articles, prepositions, abbreviations, and proper nouns, were excluded from the maze procedure. If an excluded word was selected, that word was skipped and the next nonexcluded word was selected. A word could be selected up to three times within a passage, but never twice in a row.

For each multiple choice box, two distractor words were randomly selected from the pool of words that appeared within the passage and were eligible for selection. The same rules about excluded words were applied to distractors as were applied to selected words. A word could be used as a distractor only once in a passage, regardless of whether the same word had also been selected as a maze item.

After the randomized selections were made and the maze passages were constructed, the passages were manually checked for appropriateness. Any inappropriate combinations of distractors with either the correct word or the rest of the passage were switched with other nearby distractors so they were no longer inappropriate. Any distractor that would have worked as well (made as much sense in the passage) as the original word was also switched with another nearby distractor.

Reading Composite Score

The Reading Composite Score (RCS) is a combination of multiple Acadience Reading scores and provides the best overall estimate of the student's reading proficiency. Most data management services will calculate the Reading Composite Score. To calculate the Reading Composite Score manually, see the Reading Composite Score Worksheets in Appendix 6 of the Acadience Reading Assessment Manual.

Since the scores used to calculate the Reading Composite Score vary by grade and time of year, it is important to note that the composite score generally cannot be used to directly measure growth over time or to compare results across grades or times of year. However, because the logic and procedures used to establish benchmark goals are consistent across grades and times of year, the percent of students at or above benchmark can be compared, even though the mean scores are not comparable.

As we constructed Acadience Reading, we were guided in equal parts by science (empirical evidence and research) and theory (models of reading acquisition and educational considerations). Theory and science guided our decisions about which measures to include, how to structure and compute scores, and how to guide interpretations.

The Reading Composite Score represents a range of different reading behaviors required for overall reading proficiency. Shown in Table 2.6, the four scores that are summed to obtain the Reading Composite Score in third through sixth grade include (a) ORF Words Correct, representing reading at an adequate rate, (b) ORF Accuracy, representing reading with a high degree of accuracy, (c) Retell, representing reading orally for meaning, and (d) Maze Adjusted Score, representing reading silently for meaning. Consequently, students who are at or above benchmark on the Reading Composite Score at the beginning of fourth grade are reading for meaning at an adequate rate and with a high degree of accuracy.

Table 2.6. The Measures that Comprise the Reading Composite Score for each Grade and Time of Year

Grade	Beginning of year	Middle of year	End of year
Kindergarten	LNF FSF	FSF LNF PSF NWF–CLS	LNF PSF NWF–CLS
First	LNF PSF NWF–CLS	NWF–CLS NWF–WWR ORF WC ORF Accuracy	NWF–WWR ORF WC ORF Accuracy
Second	NWF–WWR ORF WC ORF Accuracy	ORF WC ORF Accuracy Retell	ORF WC ORF Accuracy Retell
Third–Sixth	ORF WC ORF Accuracy Retell Maze	ORF WC ORF Accuracy Retell Maze	ORF WC ORF Accuracy Retell Maze

Chapter 3: **Description of Research Studies**

The *Acadience Reading K–6 Technical Manual* includes technical data and analyses from five research studies. This chapter describes the purpose of each study, the participants (including sample size and demographics), how participants were recruited, the measures that were the focus of the study and how they were administered, and a reference for further information. Demographic data from Study C is reported in Tables 3.3 and 3.4. Demographic data from the other studies is summarized in text.

Study A

Purpose. Study A was designed to examine the validity and reliability of a new Acadience Reading measure, First Sound Fluency (FSF).

Recruitment. School districts were recruited from a list of sites that had previously volunteered to participate in Acadience Reading-related research. All participating sites sent out information letters to parents of kindergarten students and used consent procedures approved by the DMG Institutional Review Board (IRB) for the protection of human research participants. All kindergarten students at the participating schools were included in the study.

Participants. Kindergarten students from 15 elementary schools across three school districts were eligible to participate. A total of 1,345 eligible kindergarten students participated during the 2006–2007 academic year.

Demographic information. The schools that participated in Study A are located in three states representing the North Central Midwest, Mountain West, and Pacific West regions of the United States, according to the US Census Bureau. Demographic characteristics were compiled from the National Center for Education Statistics website (NCES, 2007, http://nces.ed.gov/). The first school district, which had 11 participating schools, reports a predominantly white student body (90% white, 8% Hispanic) with a free/reduced lunch rate of 42%. The second school district (two participating schools) reports a predominantly white student body (80% white, 8% Hispanic, 6% African American) with a free/reduced lunch rate of 60%. The third school district reports a predominantly white student body (82% white, 11% Hispanic) with a free/reduced lunch rate of 62% across two schools.

Measures. Three measures were included in this study: Acadience Reading First Sound Fluency (FSF), DIBELS 6th Edition Initial Sound Fluency (ISF), and three subtests of the Comprehensive Test of Phonological Processing (CTOPP; Wagner, Torgesen, & Rashotte, 1999) comprising the Phonological Awareness composite. FSF and ISF were administered during benchmark testing. The CTOPP was administered in the spring (within two weeks of spring benchmark testing) as an external criterion measure. Alternate forms of FSF were given at monthly intervals between the fall and winter benchmark testing to assess the reliability of FSF.

Three schools were selected from the first district (317 students), and one school was selected from the second districted (56 students) for alternate-form administrations of FSF. A total of 82 students from the third district were randomly selected for administration of the CTOPP.

Data for all measures were collected by district personnel trained by DMG.

Descriptive Statistics. Descriptive statistics for Acadience Reading FSF, DIBELS 6th Edition ISF, and the CTOPP are shown in Table 3.1.

Table 3.1 Descriptive Statistics by Measure from Study A

Measure	*Time of year*	*N*	*Mean*	*SD*
Acadience Reading				
First Sound Fluency	Beginning	1107	13.88	11.95
First Sound Fluency	Middle	1167	34.23	13.54
First Sound Fluency	End	1194	41.01	10.71
DIBELS 6th Edition				
Initial Sound Fluency	Beginning	1258	11.62	9.48
Initial Sound Fluency	Middle	1275	24.36	12.63
CTOPP				
Phonological Awareness	End	81	101.53	11.60
Sound Matching	End	82	10.05	1.86
Elision	End	82	9.76	2.88
Blending Words	End	81	10.85	2.47

Note. Based on Study A data. CTOPP data were gathered in the spring. CTOPP results are reported as age-referenced standard scores.

For more information on this study, see:

> Cummings, K. D., Kaminski, R. A., Good, R. H., & O'Neil, M. E. (2011). Assessing phonemic awareness in preschool and kindergarten: Development and initial validation of First Sound Fluency. *Assessment for Effective Intervention, 36*(2), 94–106.

Study B

Purpose. Study B was designed to evaluate the new directions and materials that would become part of Acadience Reading, and to examine the reliability of two new Acadience Reading measures, First Sound Fluency and Maze.

Recruitment. The school district in Study B was one participating district of 13 that were involved in a larger study during the 2008–2009 school year on Acadience Reading measures. Sites that had previously volunteered to participate in Acadience Reading-related research were recruited, and all participating schools sent out information letters to parents of kindergarten students and used consent procedures approved by the DMG IRB. All students at the participating schools were included in the study.

Participants. Five schools from a single school district participated. There were 688 student participants from kindergarten through fifth grade during the 2008–2009 school year.

Demographic information. The school involved in Study B is located in the Pacific West region of the United States, according to the US Census Bureau. Demographic data at the school level were gathered from

the NCES website for the 2007–2008 school year (NCES, 2008, http://nces.ed.gov/). For the participating school, NCES reports a predominantly white student body (72% white, 17% American Indian/Alaskan, 8% Latino) with a free/reduced lunch rate of 58%. The school district in this study had two years of experience administering earlier versions of the assessment.

Measures. Developmental versions of all Acadience Reading measures were included in Study B. Acadience Reading Oral Reading Fluency results are not reported here because the passages underwent substantial revision (see Study D) after the study. All other measures were the same as or similar to Acadience Reading. Measures were administered at beginning-, middle-, and end-of-year benchmark assessment. Two weeks after middle-of-year benchmark assessment, students were given alternate forms of FSF and Maze. For all benchmark assessments, LNF, PSF, and NWF used Acadience Reading forms with DIBELS 6th Edition directions and scoring procedures. Two weeks after the middle-of-year benchmark assessment, students were given alternate forms of LNF, PSF, and NWF using the Acadience Reading directions and scoring procedures.

Benchmark assessment data were collected by school personnel trained by DMG. Data on alternate forms and the revised directions were collected by DMG personnel.

Descriptive Statistics. Descriptive statistics for Acadience Reading measures from Study B are displayed in Table 3.2.

Table 3.2 Descriptive Statistics for Middle-of-Year Acadience Reading Measures from Study B

Measure by Grade	N	Mean	SD
Kindergarten			
First Sound Fluency	97	30.10	14.74
Letter Naming Fluency	95	23.99	15.77
Phoneme-Segmentation Fluency	97	34.26	17.60
NWF Correct Letter Sounds	91	19.51	12.46
NWF Whole Words Read	91	0.95	2.89
First Grade			
NWF Correct Letter Sounds	71	45.35	19.73
NWF Whole Words Read	70	6.64	7.24
Third Grade			
Maze Adjusted Score	42	12.46	7.53
Fourth Grade			
Maze Adjusted Score	42	17.26	8.40
Fifth Grade			
Maze Adjusted Score	61	23.09	8.47

Note. N = 688. Based on middle-of-year data. All measures administered with Acadience Reading directions and scoring procedures.

For more information on Study B, see:

Dewey, E. N., Latimer, R. J., Kaminski, R. A., & Good, R. H. (2011). *DIBELS Next Development: Findings from Beta 2 Validation Study[1]* (Tech. Report No. 10). Eugene, OR: Dynamic Measurement Group. Available: https://acadiencelearning.org/.

---

[1]Acadience™ Reading K–6 is the new name for the DIBELS Next® assessment. Some historical supporting documents are referenced here with the original name. Acadience is a trademark of Dynamic Measurement Group, Inc. (DMG). The DIBELS Next copyrighted content is owned by DMG. The DIBELS Next registered trademark was sold by DMG to the University of Oregon (UO) and is now owned by the UO.

Study C

Purpose. Study C was designed to obtain the necessary information to set benchmark goals for Acadience Reading, in addition to obtaining data on the reliability and validity of all Acadience Reading measures.

Recruitment. Five school districts participated in Study C. Personnel at each of these sites had previously indicated interest in participating in Acadience Reading-related research. All students at the participating schools were included in the benchmark assessment portion of the study. In all cases of additional testing, participating sites sent out information letters and IRB-approved consent forms to the parents of selected students. Students who returned the consent forms were included in those parts of the study that required additional testing.

Participants. Thirteen schools across five districts participated. There were 3,816 student participants from kindergarten through sixth grade during the 2009–2010 school year.

Demographic information. The schools involved in Study C are located in five states in the North Central Midwest and Pacific West regions of the United States. Demographic data at the school level were gathered from NCES website for the 2008–2009 school year, and then aggregated across participating schools in each district (NCES, 2008, http://nces.ed.gov/). NCES reports a predominantly white student body (94% white, 4% Hispanic) with a free/reduced lunch rate of 16% (based on five districts). All five school districts had between four and ten years of experience administering an earlier version of these measures and using the resulting data for decision-making. NCES-reported demographic characteristics for participating districts are shown in Tables 3.3 and 3.4. Parent-reported demographic characteristics are provided in Tables 3.5 and 3.6 for those students who participated in Group Reading Assessment and Diagnostic Evaluation (GRADE) testing.

Table 3.3 United States and Research Site Demographic Comparisons

Population	Total schools	Total students	Student: Teacher ratio	Expenditure per student
District 1	2	806	18.0	$9,428
District 2	3	1682	12.9	$9,272
District 3	1	571	10.3	$16,182
District 4	5	1278	16.9	$10,562
District 5	1	255	17.2	$3,027
U.S. Primary & Secondary Schools	132,436	49,298,945	15.8	$10,041

Population	Total schools	District-wide ELL students	District-wide students with IEPs	Free/Reduced lunch eligible
District 1	2	2	135	N/A
District 2	3	9	310	300
District 3	1	34	51	20
District 4	5	45	265	302
District 5	1	15	82	96

Note. Source: U.S. Dept. of Education, National Center for Education Statistics, Common Core of Data (CCD) for the 2008–09 school year. Fiscal data available for the 2007–08 school year. Data is based on actual reported numbers and may not include students who elected to not report these data. District 4 includes data for two schools from the PSS Private School Universe Survey for the 2007–08 school year. "N/A" indicates the data are not available or not applicable. English Language Learners (ELLs), students with Individualized Education Programs (IEPs), and expenditure per student information is reported at the district level as it is unavailable at the school level, and therefore may include grades not involved in the study, such as pre-K and grades 7 through 12. Districts 1, 2, and 4 include grades not involved in the study, such as pre-K, 7, and/or 8. "U.S. Primary and Secondary" totals represent data from the 2005–06 school year. All schools were Title 1 eligible, with the exception of one school in District 2 and 3, and two schools in District 4.

Table 3.4 Demographic Information by Site Compared with Total U.S. Population

Population	Race/Ethnicity					
	American Indian or Alaska Native	Asian	Black or African American	Hispanic or Latino/a	White	Total
District 1	0	< 1% (1)	1% (6)	1% (5)	98% (784)	796
District 2	< 1% (2)	1% (11)	1% (15)	2% (32)	96% (1622)	1682
District 3	1% (3)	5% (31)	1% (5)	8% (47)	85% (478)	564
District 4	< 1% (4)	1% (7)	1% (13)	7% (89)	91% (1149)	1262
District 5	< 1% (1)	0	2% (4)	4% (10)	91% (232)	247
Total	< 1% (6)	1% (53)	< 1% (38)	4% (183)	94% (4347)	4627
US population under 18 years	1% (840 thousand)	3% (2.5 million)	14% (10.9 million)	16% (12.3 million)	65% (49.6 million)	76.14 million

Note. All data are reported from the National Center for Education Statistics (NCES) for the 2008–09 school year. District 4 includes data for two schools from the PSS Private School Universe Survey for the 2007–08 school year. Data is based on actual reported numbers, indicated in parentheses, and may not include students who elected to not report these data. Population data are the aggregate of school-level information. Districts 1, 2, and 4 include grades not involved in the study, such as pre-K, 7, and/or 8. Data for the total U.S. population under 18 years are from the 2000 Census.

Table 3.5 Parent-Reported Demographic Information for Students Receiving the GRADE

Student Demographic Category	Population					
	District 1	District 2	District 3	District 4	District 5	Total
Gender						
Male	52% (86)	49% (112)	47% (75)	53% (175)	47% (94)	50% (542)
Female	48% (80)	51% (118)	53% (84)	47% (154)	53% (107)	50% (543)
Ethnic Background						
Hispanic or Latino/a	0% (0)	5% (12)	11% (17)	7% (23)	6% (12)	6% (64)
Not Hispanic or Latino/a	97% (142)	92% (208)	88% (139)	92% (290)	94% (176)	92% (955)
Other	3% (5)	2% (5)	1% (2)	1% (2)	0% (0)	1% (14)
Racial Background						
American Indian or Native Alaskan	1% (2)	1% (2)	0% (0)	1% (2)	4% (7)	1% (13)
Asian	0% (0)	1% (2)	3% (5)	0% (0)	1% (1)	< 1% (8)
Black or African American	1% (1)	< 1% (1)	2% (3)	0% (0)	1% (2)	< 1% (7)
Multiracial	2% (4)	5% (11)	8% (13)	2% (5)	9% (17)	5% (50)
Native Hawaiian or Pacific Islander	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)
White	96% (159)	92% (212)	85% (131)	93% (302)	84% (168)	91% (972)
Other	0% (0)	1% (2)	2% (3)	4% (12)	3% (5)	2% (22)
Unknown	0% (0)	0% (0)	0% (0)	1% (2)	0% (0)	< 1% (2)

Note. Data is based on actual reported numbers, indicated in parentheses, and may not include students who elected to not report these data. Percent of students that reported any demographic information is 97% (n = 1240).

Table 3.6 Parent-Reported Demographic Information for Households of Students Receiving the GRADE

Household Demographic Category	Population					
	District 1	District 2	District 3	District 4	District 5	Total
Highest Level of Education						
Grade School	2% (3)	0% (0)	0% (0)	2% (5)	0% (0)	1% (8)
Middle School/ Junior High School	2% (4)	0% (0)	3% (5)	3% (8)	0% (0)	2% (17)
High School	35% (57)	10% (22)	1% (2)	11% (33)	40% (75)	18% (189)
2-year college	13% (21)	22% (49)	4% (7)	19% (59)	16% (29)	16% (165)
4-year college	15% (24)	33% (75)	31% (49)	28% (88)	13% (25)	25% (261)
Vocational or Technical Training	18% (29)	9% (21)	3% (5)	18% (55)	17% (31)	13% (141)
Some graduate training	4% (6)	10% (22)	9% (14)	4% (12)	5% (9)	6% (63)
Completed Masters degree	12% (19)	12% (28)	36% (56)	15% (46)	9% (16)	16% (165)
Completed Doctoral degree	1% (1)	4% (8)	12% (19)	3% (8)	1% (1)	4% (37)
Total Household Income						
$14,570 or less	20% (29)	1% (3)	0% (0)	5% (14)	8% (14)	6% (60)
$14,571–$18,310	7% (11)	0% (0)	1% (1)	1% (3)	2% (3)	2% (18)
$18,311–$22,050	5% (7)	1% (3)	1% (1)	3% (9)	2% (4)	3% (24)
$22,051–$25,790	1% (2)	2% (5)	3% (4)	1% (3)	6% (10)	3% (24)
$25,791–$29,530	3% (4)	< 1% (1)	1% (2)	3% (8)	3% (6)	2% (21)
$29,531–$33, 270	1% (2)	3% (6)	1% (1)	8% (21)	2% (3)	3% (33)
$33,271–$37,010	8% (12)	2% (5)	0% (0)	4% (10)	5% (8)	4% (35)
$37,011–$49,999	9% (13)	8% (18)	0% (0)	10% (27)	17% (29)	9% (87)
$50,000–$74,999	20% (30)	21% (45)	1% (2)	25% (68)	30% (52)	21% (197)
$75,000–$99,999	15% (22)	21% (46)	2% (3)	17% (46)	17% (30)	16% (149)
$100,000 or more	10% (15)	39% (83)	90% (132)	21% (56)	6% (10)	32% (300)

Note. Data is based on actual reported numbers, indicated in parentheses, and may not include students who elected to not report these data. Percent of students that reported any demographic information is 97% ($n = 1240$).

Measures. All Acadience Reading measures were included in this study and were administered at the beginning-, middle-, and end-of-year benchmark assessment. To assess the reliability of the measures, three types of reliability testing were conducted at various sites: 1) shadow-scoring at all sites during beginning-of-year benchmark assessment (inter-rater reliability); 2) alternate forms at one site two weeks after middle-of-year benchmark assessment (alternate-form reliability); and 3) retesting students on the same forms at a second site two weeks after middle-of-year benchmark assessment (test-retest reliability). The Group Reading Assessment and Diagnostic Evaluation (GRADE; Williams, 2001) was administered at all sites at the end of the school year as an external criterion to assess the validity of Acadience Reading measures.

During beginning- and middle-of-year benchmark assessment, 567 students participated in reliability testing. Near the end-of-year benchmark assessment, 1,306 student participants were given the GRADE measures.

All data were collected by DMG personnel or by district personnel trained by DMG.

Descriptive Statistics. The benchmark status of all the student participants in Study C is reported in Table 3.7. Descriptive statistics for the Acadience Reading measures are given in Table 3.8, and for the GRADE measures in Table 3.9.

Table 3.7 Percent of Study C Students at Each Benchmark Score Level for All Acadience Reading Measures

Measure by Grade	Beginning of Year			Middle of Year			End of Year		
	Well Below Bench-mark	Below Bench-mark	At or Above Bench-mark	Well Below Bench-mark	Below Bench-mark	At or Above Bench-mark	Well Below Bench-mark	Below Bench-mark	At or Above Bench-mark
Kindergarten									
FSF	20%	9%	71%	7%	15%	78%	--	--	--
PSF	--	--	--	9%	15%	76%	4%	14%	81%
NWF–CLS	--	--	--	12%	22%	67%	9%	29%	62%
Reading Composite Score	17%	16%	67%	11%	21%	68%	9%	23%	68%
First Grade									
PSF	12%	23%	65%	--	--	--	--	--	--
NWF–CLS	11%	22%	66%	18%	18%	64%	18%	15%	67%
NWF–WWR	–	32%	68%	18%	16%	66%	14%	17%	69%
ORF Words Correct	--	--	--	15%	17%	68%	17%	14%	69%
ORF Accuracy	--	--	--	18%	17%	65%	15%	15%	70%
Retell	--	--	--	–	–	–	–	19%	81%
Retell Quality	--	--	--	–	–	–	–	–	–
Reading Composite Score	18%	13%	68%	18%	15%	67%	18%	15%	68%
Second Grade									
NWF–CLS	15%	23%	62%	--	--	--	--	--	--
NWF–WWR	22%	19%	59%	--	--	--	--	--	--
ORF Words Correct	21%	12%	67%	19%	13%	68%	13%	18%	68%
ORF Accuracy	14%	16%	70%	17%	17%	66%	13%	19%	68%
Retell	8%	13%	79%	9%	14%	77%	7%	17%	77%
Retell Quality	–	–	–	–	24%	76%	–	13%	87%
Reading Composite Score	17%	11%	72%	17%	10%	73%	13%	14%	73%
Third Grade									
ORF Words Correct	16%	12%	71%	14%	13%	72%	15%	15%	69%
ORF Accuracy	13%	21%	66%	9%	18%	72%	15%	16%	69%
Retell	6%	19%	75%	9%	13%	77%	9%	13%	78%
Retell Quality	–	29%	71%	–	15%	85%	7%	30%	62%
Maze Adjusted Score	15%	14%	71%	14%	16%	69%	14%	17%	69%
Reading Composite Score	18%	10%	72%	16%	11%	73%	14%	13%	72%

Table 3.7 Percent of Study C Students at Each Benchmark Score Level for All Acadience Reading Measures, (*continued*)

Measure by Grade	Beginning of Year			Middle of Year			End of Year		
	Well Below Bench-mark	Below Bench-mark	At or Above Bench-mark	Well Below Bench-mark	Below Bench-mark	At or Above Bench-mark	Well Below Bench-mark	Below Bench-mark	At or Above Bench-mark
Fourth Grade									
ORF Words Correct	21%	14%	65%	15%	17%	68%	15%	18%	67%
ORF Accuracy	20%	21%	59%	15%	18%	67%	10%	26%	64%
Retell	11%	21%	68%	14%	23%	64%	13%	17%	69%
Retell Quality	–	25%	75%	–	15%	85%	8%	34%	58%
Maze Adjusted Score	16%	20%	64%	16%	27%	57%	14%	18%	68%
Reading Composite Score	21%	12%	67%	21%	12%	68%	14%	18%	68%
Fifth Grade									
ORF Words Correct	22%	14%	64%	15%	22%	63%	16%	21%	64%
ORF Accuracy	17%	28%	55%	13%	23%	63%	17%	38%	45%
Retell	14%	24%	62%	12%	23%	65%	12%	24%	63%
Retell Quality	–	17%	83%	7%	32%	61%	10%	26%	64%
Maze Adjusted Score	15%	20%	65%	14%	32%	53%	14%	21%	65%
Reading Composite Score	13%	22%	65%	14%	21%	65%	14%	22%	64%
Sixth Grade									
ORF Words Correct	10%	12%	78%	12%	10%	79%	9%	13%	78%
ORF Accuracy	10%	16%	75%	8%	17%	75%	11%	20%	70%
Retell	4%	17%	80%	4%	17%	79%	9%	11%	80%
Retell Quality	–	12%	88%	–	8%	92%	7%	26%	67%
Maze Adjusted Score	10%	12%	78%	8%	12%	80%	8%	17%	76%
Reading Composite Score	9%	13%	78%	9%	12%	79%	9%	12%	79%

Note. Sample size = 3,816; Approximate grade-level sample sizes: kindergarten ≈ 450; first grade ≈ 435 (370 for Retell); second grade ≈ 540 (480 for Retell); third grade ≈ 450 ; fourth grade ≈ 560; fifth grade ≈ 510; sixth grade ≈ 510. Based on Study C data. There is no benchmark goal for LNF. '–' indicates that the measure is administered, but there is no recommended benchmark goal or cut point for that grade or support level at that time of year. In most cases, this appears with reference to Retell or Retell Quality of Response, in which case, the benchmark goal also functions as the cut-point for strategic support. '--' indicates that the measure is not administered during that time point.

Table 3.8 Descriptive Statistics for Acadience Reading Measures from Study C

Measure	Beginning of Year			Middle of Year			End of Year		
	N	Mean	SD	N	Mean	SD	N	Mean	SD
Kindergarten									
FSF	478	18.10	12.75	476	38.19	12.08	-	-	-
LNF	478	22.35	15.72	476	41.92	16.34	479	50.54	15.64
PSF	-	-	-	476	37.71	18.55	479	50.52	14.70
NWF–CLS	-	-	-	476	23.69	14.56	477	36.77	21.27
NWF–WWR	-	-	-	474	1.93	4.08	477	6.02	8.42
Reading Composite Score	476	40.45	25.18	476	141.50	48.81	477	138.21	41.99
First Grade									
LNF	461	49.56	15.97	-	-	-	-	-	-
PSF	461	43.06	15.23	-	-	-	-	-	-
NWF–CLS	461	38.82	25.31	458	59.86	32.52	461	77.08	33.86
NWF–WWR	461	6.93	9.81	457	15.52	13.50	461	21.63	14.50
ORF Words Correct	-	-	-	458	45.08	35.14	459	66.73	35.00
ORF Accuracy	-	-	-	451	81%	16	459	90%	12
Retell	-	-	-	271	12.14	15.14	373	27.39	15.09
Reading Composite Score	460	131.72	44.78	450	186.00	99.96	459	189.48	83.80
Second Grade									
NWF–CLS	560	68.53	33.44	-	-	-	-	-	-
NWF–WWR	560	18.22	13.93	-	-	-	-	-	-
ORF Words Correct	560	70.44	37.20	565	88.38	37.69	566	103.21	38.96
ORF Accuracy	558	91%	12	564	95%	9	566	96%	7
Retell	520	26.88	15.95	548	31.26	17.35	541	38.61	18.48
Reading Composite Score	558	186.86	83.82	564	234.11	99.80	565	270.16	97.23
Third Grade									
ORF Words Correct	502	89.56	37.64	504	103.55	37.44	498	116.84	37.74
ORF Accuracy	502	94%	9	502	96%	6	498	97%	6
Retell	479	30.98	16.86	492	38.00	19.77	486	42.98	19.01
Maze Adjusted Score	501	11.40	6.58	503	15.39	8.85	497	22.02	8.54
Reading Composite Score	501	273.29	116.37	501	332.65	119.43	495	385.24	116.76
Fourth Grade									
ORF Words Correct	589	101.15	37.71	596	114.91	38.16	589	129.62	37.41
ORF Accuracy	589	95%	7	592	96%	6	589	97%	5
Retell	577	35.61	18.94	592	37.91	19.40	582	44.05	21.04
Maze Adjusted Score	585	16.91	7.49	589	18.64	7.85	586	26.65	8.80
Reading Composite Score	584	321.88	118.42	585	361.48	112.55	585	425.02	113.19

Table 3.8 Descriptive Statistics for Acadience Reading Measures from Study C, *(continued)*

Measure	Beginning of Year			Middle of Year			End of Year		
	N	Mean	SD	N	Mean	SD	N	Mean	SD
Fifth Grade									
ORF Words Correct	546	121.47	38.32	543	132.02	36.09	539	139.50	38.71
ORF Accuracy	546	97%	6	542	98%	4	539	98%	3
Retell	540	38.99	17.53	541	43.53	19.01	530	45.15	19.73
Maze Adjusted Score	543	20.39	8.83	537	20.01	8.03	536	27.11	10.06
Reading Composite Score	541	377.72	112.45	535	402.92	105.24	532	442.08	114.00
Sixth Grade									
ORF Words Correct	535	131.33	35.26	528	136.03	38.10	531	143.93	37.16
ORF Accuracy	535	97%	4	528	97%	4	531	98%	3
Retell	529	41.52	17.58	526	47.80	21.05	527	51.63	21.33
Maze Adjusted Score	535	22.41	7.65	520	26.52	9.89	529	27.00	9.89
Reading Composite Score	534	403.36	103.46	519	437.78	120.89	526	458.74	114.94

Note. Based on Study C data. N = 3,816.

Table 3.9 Descriptive Statistics for the Group Reading Assessment and Diagnostic Evaluation (GRADE) Total Test Raw Scores

Grade	N	Mean	SD
Kindergarten	168	74.79	8.51
First	196	70.53	15.71
Second	219	87.68	12.74
Third	187	88.39	14.56
Fourth	187	54.39	17.00
Fifth	195	55.02	17.14
Sixth	105	59.29	13.67

Note. Based on Study C end-of-year data.

For more information on Study C, see:

> Powell-Smith, K. A., Good, R. H., Latimer, R. J., Dewey, E. N., & Kaminski, R. A. (2011). *DIBELS Next Benchmark Goals Study*[1] (Tech. Report No. 11). Eugene, OR: Dynamic Measurement Group. Available: https://acadiencelearning.org/.

Study D

Purpose. The goal of Study D was to evaluate Acadience Reading Oral Reading Fluency (ORF) passages for reliability, validity, and passage difficulty.

Recruitment. Student participants were from one elementary and one middle school. Students whose teachers volunteered to participate were recruited for participation in the study. Students receiving English-language reading instruction in first- through sixth-grade general education classrooms were eligible for participation.

---

[1]Acadience™ Reading K–6 is the new name for the DIBELS Next® assessment. Some historical supporting documents are referenced here with the original name. Acadience is a trademark of Dynamic Measurement Group, Inc. (DMG). The DIBELS Next copyrighted content is owned by DMG. The DIBELS Next registered trademark was sold by DMG to the University of Oregon (UO) and is now owned by the UO.

Participants. All data were collected during the spring of 2009. Twenty-one teachers elected to participate in the study. Between 28 and 30 IRB-approved consent letters per grade were distributed. The final sample included 140 students.

Demographic information. The schools involved in Study D are located in one state in the Mountain West region of the United States. Demographic data at the school level were gathered from NCES website for the 2006–2007 school year (NCES, 2007, http://nces.ed.gov/). The elementary school reports a predominantly white student body (81% white, 13% American Indian) and a free/reduced lunch rate of 39%. The middle school also reports a predominantly white student body (89% white, 6% American Indian) and a free/reduced lunch rate of 56%.

Measures. Three measures were included in this study: Acadience Reading Oral Reading Fluency (ORF), DIBELS 6th Edition Oral Reading Fluency, and the Standard 4th Grade Reading Passage used in the National Assessment of Education Progress (NAEP) 2002 Special Study of Oral Reading (Daane, Campbell, Grigg, Goodman, & Oranje, 2005). Acadience Reading Oral Reading Fluency directions were used for all passages. Over approximately a two-week period, students were administered 40 Acadience Reading passages at their grade level, one DIBELS 6th Edition passage at their grade level, plus the fourth-grade NAEP Oral Reading Study passage, "The Box in the Barn". Acadience Reading passages were administered in a random order specific to each participating student. The NAEP passage was administered as the second passage in the second session, and the 6th Edition ORF passage was administered as the second passage in the third session. Each testing session was approximately 8 to 10 minutes in length. Testing was discontinued and no further passages were administered if students met their grade-level discontinue criteria. If more than five students per grade met the discontinue criterion, another student at that grade level was selected from the pool of eligible students so that the sample did not drop below 20 per grade.

All data were collected by the onsite coordinator and 13 university students trained by DMG.

Descriptive Statistics. Descriptive statistics for Acadience Reading ORF passages are given in Table 3.10.

Table 3.10 Descriptives for all Acadience Reading ORF Benchmark Passages from Study D

Grade	Number of Students	Number of Passages	Median Passage-Level Mean Score	Median Passage-Level SD
First	23	29	81.52	43.11
Second	25	32	115.12	36.53
Third	22	32	109.89	39.13
Fourth	23	32	131.87	31.99
Fifth	23	32	136.24	36.07
Sixth	24	32	150.99	28.63

Note. Data gathered from Study D. All passages administered at end of year.

For more information on Study D, see:

Powell-Smith, K. A., Good, R. H., & Atkins, T. (2010). *DIBELS Next Oral Reading Fluency Readability Study*[1] (Tech. Report No. 7). Eugene, OR: Dynamic Measurement Group. Available: https://acadiencelearning.org/.

---

[1]Acadience™ Reading K–6 is the new name for the DIBELS Next® assessment. Some historical supporting documents are referenced here with the original name. Acadience is a trademark of Dynamic Measurement Group, Inc. (DMG). The DIBELS Next copyrighted content is owned by DMG. The DIBELS Next registered trademark was sold by DMG to the University of Oregon (UO) and is now owned by the UO.

Study E

Purpose. Study E was designed to obtain alternate-form reliability information on Acadience Reading Phoneme Segmentation Fluency (PSF) in first grade and all sixth grade measures.

Recruitment. Personnel at each of these sites had previously indicated interest in participating in Acadience Reading-related research. All students at the participating schools were included in the benchmark assessment portion of the study. In the cases of additional testing, participating sites sent out information letters and IRB-approved opt-out forms to all parents with students in the appropriate grade levels. Students who returned the opt-out forms were not included in those parts of the study that required additional testing.

Participants. Three schools across two districts participated. There were 345 student participants from first and sixth grade during the fall of the 2012–2013 school year.

Demographic information. The schools involved in Study E are located in one state in the East North Central region of the United States. Demographic data at the school level were gathered from NCES website for the 2010–2011 school year (NCES, 2012, http://nces.ed.gov/). NCES reports a predominantly white student body (90% white, 8% American Indian / Alaska Native) with a free/reduced lunch rate of 28% (based on both districts).

Measures. Students in all participating grades were given their Acadience Reading benchmark assessment in the fall. Approximately two weeks later, students were assessed using progress monitoring forms to evaluate the alternate-form reliability. During this second round of testing, students in grade 1 were administered a single assessment of PSF. In sixth grade, students were given three ORF passages, each followed by an administration of Retell. Sixth-grade students were also given one administration of Maze.

Descriptive Statistics. Descriptive Statistics for Acadience Reading measures from Study E are reported in Table 3.11.

Table 3.11 Descriptive Statistics for Beginning-of-Year Acadience ReadingMeasures from Study E

Grade and Measure	N	M	SD
First Grade			
Phoneme Segmentation Fluency	164	49.19	13.84
Sixth Grade			
ORF Words Correct	61	127.46	28.59
ORF Accuracy	61	.98	.02
Retell	61	32.57	16.50
Maze	60	27.03	8.89
Reading Composite Score	60	405.23	87.68

Note. N = 225. Based on Beginning-of-year data.

For more information on Study E, please contact Dynamic Measurement Group at https://acadiencelearning.org/.

Chapter 4: **Benchmark Goals**

In this chapter, we describe the Acadience Reading benchmark goals and cut points for risk, as well as the procedures for establishing those goals. A complete list of the goals and cut-points for risk as well as the odds of achieving later important reading outcomes are given in Appendix A.

The Acadience Reading benchmark goals, cut points for risk, and composite score were developed based upon data collected in Study C. The goals represent a series of conditional probabilities of meeting later important reading outcomes. The external criterion was the Group Reading Assessment and Diagnostic Evaluation (GRADE; Williams, 2001). The 40th percentile on the GRADE assessment was used as an indicator that the student was making adequate progress in acquisition of important early reading and/or reading skills. Data for the study were collected in 13 elementary and middle schools in five states. Data collection included administering the Acadience Reading measures to participating students in grades K–6 in addition to the GRADE. Participants in the study were 3,816 students across grades K–6 from general education classrooms who were receiving English language reading instruction, including students with disabilities and students who were English language learners provided they had the response capabilities to participate. The study included both students who were struggling in reading and those who were typically achieving. A subset of the total sample participated in the GRADE assessment (*n* = 1,306 across grades K–6). See Chapter 3 for more information on Study C.

Benchmark Goals

The Acadience Reading benchmark goals are empirically derived, criterion-referenced target scores that represent adequate reading progress. A benchmark goal indicates a level of skill where the student is likely to achieve the next Acadience Reading benchmark goal or reading outcome. Benchmark goals for Acadience Reading are based on research that examines the predictive validity of a score on a measure at a particular point in time, compared to later Acadience Reading measures and external outcome assessments. If a student achieves a benchmark goal, then the odds are in favor of that student achieving later reading outcomes if he/she receives generally effective, research-based instruction from a core classroom curriculum.

Cut Points for Risk

The cut points for risk indicate a level of skill below which the student is unlikely to achieve subsequent reading goals without receiving additional, targeted instructional support. Students with scores below the cut point for risk are identified as likely to need intensive support. Intensive support refers to interventions that incorporate something more or something different from the core curriculum or supplemental support. Intensive support might entail:

- delivering instruction in a smaller group,
- providing more instructional time or more practice,
- presenting smaller skill steps in the instructional hierarchy,
- providing more explicit modeling and instruction, and/or
- providing greater scaffolding and practice

Because students needing intensive support are likely to have individual and sometimes unique needs, we recommend that their progress be monitored frequently and their intervention modified dynamically to ensure adequate progress.

Between a benchmark goal and a cut point for risk is a range of scores where the student's future performance is harder to predict. To ensure that the greatest number of students achieve later reading success, it is best for students with scores in this range to receive carefully targeted additional support in the skill areas where they are having difficulty, to be monitored regularly to ensure that they are making adequate progress, and to receive increased or modified support if necessary to achieve subsequent reading goals. This type of instructional support is referred to as strategic support.

Table 4.1 (on page 49) provides the specified target odds of achieving later reading outcomes and labels for "likely need for support" for each of the score levels. Benchmark goals and cut points for risk are provided for the Reading Composite Score as well as for individual Acadience Reading measures.

Reading Composite Score Benchmark Goals

Benchmark goals and cut points for risk for the Reading Composite Score are based on the same logic and procedures as the individual Acadience Reading measures; however, since the Reading Composite Score provides the best overall estimate of a student's skills, the Reading Composite Score should usually be interpreted first. If a student is at or above the benchmark goal on the Reading Composite Score, the odds are in the student's favor of reaching later important reading outcomes. Some students who score at or above the Reading Composite Score benchmark goal may still need additional support in one or more of the basic early literacy skills, as indicated by a below-benchmark score on an individual Acadience Reading measure (FSF, PSF, NWF, ORF, or Maze), especially those students whose composite score is close to the benchmark goal.

Determining the Acadience Reading K–6 Benchmark Goals and Cut Points for Risk

Adequate Reading Skills

The Acadience Reading benchmark goals provide targeted levels of skill that students need to achieve by specific times to be considered to be making adequate progress. In developing benchmark goals, our focus is on general adequate reading skills, and is not specific to a particular state assessment, published reading test, or national assessment. A student with adequate reading skills should read adequately regardless of the specific assessment that is used.

In the 2007 National Assessment of Educational Progress, 34% of students scored below the level of reading skills judged to be Basic, and 68% of students scored below the level judged to be Proficient. According to the NAEP, "Basic denotes partial mastery of prerequisite knowledge and skills that are fundamental for proficient work at a given grade (Daane et al., 2005, p. 18)." Thus, students who score at the 40th percentile or above on a high-quality, nationally norm-referenced test are likely to be rated Basic or above on the NAEP and can be considered to have adequate reading skills. In our benchmark goal study, we used the 40th percentile or

above on the GRADE as *one approximation* of adequate reading skills. Our intent is to develop generalizable benchmark goals and cut points that will be relevant and appropriate for a wide variety of reading outcomes, across a wide variety of states and regions, and for diverse groups of students. No single study can provide all the information necessary to evaluate generalizability. Multiple studies will evaluate the reliability, validity, and utility of Acadience Reading. We are ultimately most interested in the convergence of evidence from many research studies that utilize many different sites, samples of students, and reading outcome measures.

GRADE as Initial External Criterion

We used the Group Reading Assessment and Diagnostic Evaluation (GRADE; Williams, 2001), a high-quality, nationally norm-referenced assessment, as an external criterion in our Benchmark Goal Study. We emphasized the GRADE Total Test Raw Score as the primary score to examine. In our analyses we found that the total score worked better as a criterion than the individual scores, and that the individual scores were related to other measures much the same as the total score was related to other measures. The lowest raw score on the GRADE that was at or above the 40th percentile compared to the GRADE normative sample was used as an approximation of the external criterion of adequate reading skills. The lowest raw score on the GRADE that was at or above the 20th percentile compared to the GRADE normative sample was used as an approximation for the external cut point for risk. Subsequent research will be essential to verify and replicate these findings with a range of other external criterion measures.

Reading Composite Score as Primary Internal Criterion

We used the Reading Composite Score as a primary internal (i.e., within the Acadience Reading assessment system) criterion because it is the best indicator of the student's overall reading proficiency. This represents a change from our earlier work where ORF was used as the primary indicator of a student's reading proficiency. In our research with Acadience Reading, we find that, although the Acadience Reading ORF Words Correct score is very good in isolation, the Reading Composite Score is substantially better. For example, the end-of-year third-grade ORF Words Correct correlates .66 with the end-of-year GRADE Total Test Raw Score, which is a very strong validity coefficient. However, the end-of-year, third-grade Reading Composite Score correlates .75 with the end-of-year GRADE Total Test Raw Score, explaining 13% more variance than ORF alone. In general, we find that the Reading Composite Score provides a better overall measure of reading proficiency than the best single Acadience Reading measure at almost every grade and time of year. In addition to correlating more highly with external outcomes, the Reading Composite Score also provides a larger and more complete sample of reading behavior than any single measure in isolation. Thus, the Reading Composite Score serves as a very important internal criterion in developing and validating the Acadience Reading benchmark goals and cut points for risk.

Step-by-Step Procedures

The principle vision for Acadience Reading is a step-by-step vision. Student skills at or above benchmark at the beginning of the year put the odds in favor of the student achieving the middle-of-year benchmark goal. In turn, students with skills at or above benchmark in the middle of the year have the odds in favor of achieving the end-of-year benchmark goal. Finally, students with skills at or above benchmark at the end of the year have the odds in favor of adequate reading skills on a wide, general variety of external measures of reading proficiency.

Our fundamental logic for developing the benchmark goals and cut points for risk was to begin with the external outcome goal and work backward in that step-by-step system. We first obtained an external criterion measure (the GRADE Total Test Raw Score) at the end of the year with a level of performance that would represent adequate reading skills. Next we specified the benchmark goal and cut point for risk on the end-of-year Reading

Composite Score with respect to the end-of-year external criterion. Then, using the Reading Composite end-of-year goal as an internal criterion, we established the benchmark goals and cut points for risk on the middle-of-year Reading Composite Score. Finally, we established the benchmark goals and cut points for risk on the beginning-of-year Reading Composite Score using the middle-of-year Reading Composite Score as an internal criterion.

Once the benchmark goals and cut points for risk were established for the Reading Composite Score, they were used to establish the specific goals and cut points for risk for each individual Acadience Reading measure. The same step-by-step procedures were used for the individual measures.

Primary Design Specifications for Benchmark Goals and Cut Points for Risk

The primary specification for the Acadience Reading benchmark goals was to establish a level of skill where students scoring at or above benchmark have favorable odds (80%–90%) of achieving subsequent reading outcomes. In other words, students scoring at or above the benchmark goal are in a zone where we are reasonably confident they will make adequate progress. The primary specification for a Acadience Reading cut point for risk is a level of skill where students scoring below that level have low odds (10%–20%) of achieving subsequent reading outcomes. In other words, students scoring below the cut point for risk are in a zone where we are reasonably confident the student will not make adequate progress unless provided with additional, intensive support.

In between the benchmark goal and the cut point for risk is a level of skill where the odds are about even (40%–60%) of achieving subsequent reading outcomes. We are not confident that students with skills in this range will make adequate progress; we are also not confident that they will not. In other words, between the benchmark goal and the cut point for risk is a zone of uncertainty where we cannot make a good prediction of outcomes. By providing additional, strategic support to students with skills in this range along with progress monitoring, we can increase the likelihood that the student will make adequate progress.

Secondary Design Specifications for Benchmark Goals and Cut Points for Risk

A secondary consideration in establishing benchmark goals and cut points for risk was based on an examination of marginal percents. We tried to keep the marginal percent of students in each score level consistent from predictor to criterion. For example, 73% of students in our third-grade sample scored at or above the 40th percentile on the GRADE external criterion measure, indicating a fairly high performing sample. We set the third-grade end-of-year benchmark goal so that 73% of the sample also scored at or above benchmark on the Reading Composite Score. Thus, the sample appears equally high performing on both the Acadience Reading predictor and the GRADE criterion.

Another important secondary consideration in establishing benchmark goals and cut points for risk was based on the logistic regression predicting the odds of scoring at or above benchmark on the criterion, based on their score on the predictor. For all students in the "At or Above Benchmark" range, the odds of achieving subsequent goals may be 80% to 90%; however, for students at the high end of that range the odds are somewhat higher, and for students at the low end of that range the odds are somewhat lower. The logistic regression analysis was used to estimate the odds of achieving subsequent early literacy goals for students who obtain the exact benchmark goal or the exact cut point for risk score. We tried to keep the predicted odds for students obtaining the exact benchmark goal at 60% or higher of achieving subsequent goals. We also tried to keep the predicted odds of achieving subsequent goals at 40% or less for students obtaining the exact score corresponding to the cut point for risk. For example, on the third-grade end-of-year Acadience Reading assessment, the predicted odds of scoring at or above the 40th percentile on the GRADE were 67% for students scoring exactly the Reading Composite Score benchmark goal; the odds were 32% for students scoring exactly the cut point for risk.

Other Design Specifications for Benchmark Goals and Cut Points for Risk

In addition to the primary and secondary considerations in establishing benchmark goals and cut points for risk, we also considered a number of issues including:

- The pattern of student performance in the scatterplot. We tried to establish goals where students scoring at or above benchmark on the predictor were mostly also at or above benchmark on the criterion; where students who scored below benchmark on the predictor were equally split by the benchmark goal on the criterion; and where students who were below the cut point for risk were mostly below the benchmark goal on the criterion.
- The receiver operator characteristic (ROC) curve analysis. A large area under curve (AUC) is desirable in ROC analysis and indicates a good trade-off of sensitivity and specificity. Benchmark goals in the upper left corner of the curve represent a balance of sensitivity and sensitivity.
- We also examined and considered other metrics for decision utility including sensitivity, specificity, negative predictive power, positive predictive power, percent accurate classification, and Kappa.
- Finally, we considered the overall pattern of benchmark goals and cut points for risk across measures and grades, and the historical benchmark goals and cut points for risk from DIBELS 6th Edition. In addition, we considered the theoretical relations between core components of early literacy in our model.

Overall Evaluative Judgment

We specified the benchmark goals and cut points for risk as an overall evaluative judgment of primary, secondary, and other design specifications. No single concern was used in isolation from other concerns. Frequently we had to balance disparate concerns to obtain a satisfactory compromise. For example, increasing the benchmark goal might result in a better match of marginal percents, but might compromise the predicted odds in the logistic regression analysis. Alternatively, a lower benchmark goal might work better for the beginning-of-year to middle-of-year analysis, but perform more poorly in the middle-of-year to end-of-year analysis. In other cases, the logistic regression analysis did not fit the data well, and consequently the role of the logistic regression analysis was discounted in establishing the benchmark goals and cut points for risk. The benchmark goals and cut points for risk represent our best balance of all the considerations identified here.

Linking Acadience Reading Score Levels to Likely Need for Support

A key point in this discussion of odds is that the student's outcome is unknown and not fixed at the time of the initial screening. Instead, the outcome is the result of both the student's initial skills and the targeted, differentiated instruction and intervention that are provided as a direct result of the screening information. Our instructional goal is to ruin initial screening predictions of less than adequate progress. For example, if a student screens as being at high risk on a measure of early literacy skills on the beginning-of-year kindergarten assessment (i.e., low odds of achieving kindergarten goals), then he/she is likely to need additional instructional support to be successful. The student's later outcomes, such as reading skills in first grade, are a direct result of the targeted, differentiated instruction and early intervention that are provided. The linkage between the odds of achieving subsequent early literacy goals, Acadience Reading score levels, and likely need for support is summarized in Table 4.1. For all students, those who are at or above benchmark, below benchmark, and well below benchmark, our charge is to provide adequate support so they all achieve subsequent early literacy goals.

Table 4.1 Odds of Achieving Subsequent Early Literacy Goals, Score Levels, and Likely Need for Support

Odds of achieving subsequent early literacy goals	Score level	Likely need for support to achieve subsequent early literacy goals
80% to 90%	At or Above Benchmark *scores at or above the benchmark goal*	Likely to Need Core Support
40% to 60%	Below Benchmark *scores below the benchmark goal and at or above the cut point for risk*	Likely to Need Strategic Support
10% to 20%	Well Below Benchmark *scores below the cut point for risk*	Likely to Need Intensive Support

Benchmark Goals and Cut Points for Risk Analysis Detail

The benchmark goals and cut points for risk are summarized in Table 4.2. Each benchmark goal and cut point for risk is supported by one or more detailed analyses. The analysis details for the Reading Composite Scores are included in this technical manual in pages 58 to 78. Each analysis detail page reports how a predictor (or screening decision) variable is related to a criterion (or outcome) variable. For each grade level, an analysis detail page is provided for: (a) beginning of year to middle of year, (b) middle of year to end of year, and (c) end of year to the end-of-year external criterion assessment. In this way, we provide information on how earlier Acadience Reading measures relate to later Acadience Reading measures, and also on how Acadience Reading measures relate to the external criterion. Each analysis detail consists of: (a) heading, (b) scatterplot, (c) contingency table, (d) logistic regression analysis, (e) receiver operator characteristic (ROC) curve analysis, and (e) summary of other decision-utility metrics.

Table 4.2 Acadience Reading Summary of Benchmark Goals and Cut Points for Risk

Acadience Reading: Summary of Benchmark Goals and Cut Points for Risk

Each cell shows **Benchmark Goal / Cut Point for Risk**.

Measure	K Beg	K Mid	K End	G1 Beg	G1 Mid	G1 End	G2 Beg	G2 Mid	G2 End	G3 Beg	G3 Mid	G3 End	G4 Beg	G4 Mid	G4 End	G5 Beg	G5 Mid	G5 End	G6 Beg	G6 Mid	G6 End
Reading Composite Score	26 / 13	122 / 85	119 / 89	113 / 97	130 / 100	155 / 111	141 / 109	190 / 145	238 / 180	220 / 180	285 / 235	330 / 280	290 / 245	330 / 290	391 / 330	357 / 258	372 / 310	415 / 340	344 / 280	358 / 285	380 / 324
First Sound Fluency (FSF)	10 / 5	30 / 20																			
Letter Naming Fluency (LNF)	No benchmark set for LNF																				
Phoneme Segmentation Fluency (PSF)		20 / 10	40 / 25	40 / 25																	
Nonsense Word Fluency (NWF) – Correct Letter Sounds			17 / 8	28 / 15	43 / 33	58 / 47	54 / 35														
Nonsense Word Fluency (NWF) – Whole Words Read				1 / 0	8 / 3	13 / 6	13 / 6														
Oral Reading Fluency (ORF) – Words Correct					23 / 16	47 / 32	52 / 37	72 / 55	87 / 65	70 / 55	86 / 68	100 / 80	90 / 70	103 / 79	115 / 95	111 / 96	120 / 101	130 / 105	107 / 90	109 / 92	120 / 95
Oral Reading Fluency (ORF) – Accuracy					78% / 68%	90% / 82%	90% / 81%	96% / 91%	97% / 93%	95% / 89%	96% / 92%	97% / 94%	96% / 93%	97% / 94%	98% / 95%	98% / 95%	98% / 96%	99% / 97%	97% / 94%	97% / 94%	98% / 96%
Oral Reading Fluency (ORF) – Retell						15 / 0	16 / 8	21 / 13	27 / 18	20 / 10	26 / 18	30 / 20	27 / 14	30 / 20	33 / 24	33 / 22	36 / 25	36 / 25	27 / 16	29 / 18	32 / 24
Oral Reading Fluency (ORF) – Retell Quality of Response								2 / 1	2 / 1	2 / 1	2 / 1	3 / 2	2 / 1	2 / 1	3 / 2	2 / 1	3 / 2	3 / 2	2 / 1	2 / 1	3 / 2
Maze										8 / 5	11 / 7	19 / 14	15 / 10	17 / 12	24 / 20	18 / 12	20 / 13	24 / 18	18 / 14	19 / 14	21 / 15

Reading Composite Score: A combination of multiple Acadience Reading scores, which provides the best overall estimate of the student's reading proficiency. For information on how to calculate the composite score, see the *Acadience Reading Composite Score* worksheets available in Appendix 6 of this manual or from https://acadience.org/.

BENCHMARK GOAL (bold number at top of each box): Students scoring at or above the benchmark goal have the odds in their favor (approximately 80%–90%) of achieving later important reading outcomes. These scores are identified as "At or Above Benchmark" and the students are likely to need *Core Support*.

CUT POINT FOR RISK (smaller number in each box): Students scoring below the cut point for risk are unlikely (approximately 10%–20%) to achieve subsequent goals without receiving additional, targeted instructional support. These scores are identified as "Well Below Benchmark" and the students are likely to need *Intensive Support*.

Scores below the benchmark goal and at or above the cut point for risk are identified as "Below Benchmark." In this range, a student's future performance is harder to predict, and these students are likely to need *Strategic Support*.

This page is adapted from a chart developed by Cache County School District.

Heading

The heading at the top of each analysis detail page provides information about which two variables are examined. The predictor is provided first, along with the corresponding variable name, benchmark goal and cut point for risk, and description of the variable. Variable names are used in the benchmark goal analysis detail pages to summarize a large amount of information on one page per analysis. The criterion or outcome is provided next, also with the corresponding variable name, benchmark goal and cut point for risk, and a text description of the variable. In the analysis detail pages included in this technical manual, the predictor and criterion are specified so that (a) the predictor variable is the beginning-of-year Reading Composite Score and the criterion is the middle-of-year Reading Composite Score, (b) the predictor variable is the middle-of-year Reading Composite Score and the criterion is the end-of-year Reading Composite Score, or (c) the predictor variable is the end-of-year Reading Composite Score and the criterion measure is the Group Reading Assessment and Diagnostic Evaluation (GRADE) Total Test Raw Score, also administered at the end of the year.

In Figure 4.1, the predictor or basis for a screening decision is the Reading Composite Score (identified by RCS in the variable name) from the third-grade (identified by the number 3 in the variable name), end-of-year (identified by the letter *e* in the variable name) assessment with a benchmark goal of 330 and a cut point for risk of 280. The criterion is the GRADE Total Test Raw Score (*gtotr*) from the third-grade (3), end-of-year (e) assessment with a benchmark goal of 83 and a cut point for risk of 71. The benchmark goal for the GRADE Total Test Raw Score (83) was specified to be the lowest raw score that was at or above the 40th percentile according to the GRADE norms. The corresponding cut point for risk (71) was the lowest raw score that was at or above the 20th percentile.

Figure 4.1 Benchmark goal analysis detail header

	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	RCS3e	330	280	Reading Composite Score, Grade 3, End of Year
Outcome	Criterion	gtotr3e	83	71	GRADE Total Test, Grade 3, End of Year

Scatterplot

The scatterplot provides a visual representation of the relation between student scores on the predictor (along the horizontal axis) and the criterion (along the vertical axis). Each dot represents an individual student's scores on the predictor and criterion. The vertical solid line (on the right) represents the benchmark goal for the predictor where students performing at or above the benchmark goal are likely to need core support to achieve the benchmark goal on the criterion. The horizontal solid line (on the top) represents the benchmark goal on the criterion. The vertical dashed line (to the left) represents the cut point for risk on the predictor where students scoring below the cut point for risk are likely to need intensive support to achieve the benchmark goal on the criterion. The horizontal dashed line (on the bottom) represents the cut point for well below benchmark performance on the criterion. The correlation between the predictor and criterion is also reported in the lower right corner of the scatterplot.

In Figure 4.2, the scatterplot for Reading Composite Score for third grade end of year (RCS3e) and GRADE Total Test Raw Score for third grade end of year (gtotr3e) is portrayed. Two extreme outliers are evident in the lower right corner of the scatterplot who received very high scores on the Reading Composite, and very low scores on the GRADE. An educator would want to follow up with those students to make sure an accurate estimate of their skills and progress is obtained. Even with those outliers, the two measures are correlated .75. Visually we see

that most students who scored at or above the benchmark goal on RCS3e also scored at or above benchmark on the GRADE. Most students who scored below the cut point for risk on the RCS3e also scored below the benchmark goal on the GRADE. Students who scored between the cut point for risk and the benchmark goal on the RCS3e were about evenly split, with about half above the benchmark goal on the GRADE and half below the goal.

Figure 4.2 Scatterplot illustrating the relation between Reading Composite Score for third grade end of year (RCS3e) and GRADE Total Test Raw Score for third grade end of year (gtotr3e)

Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).
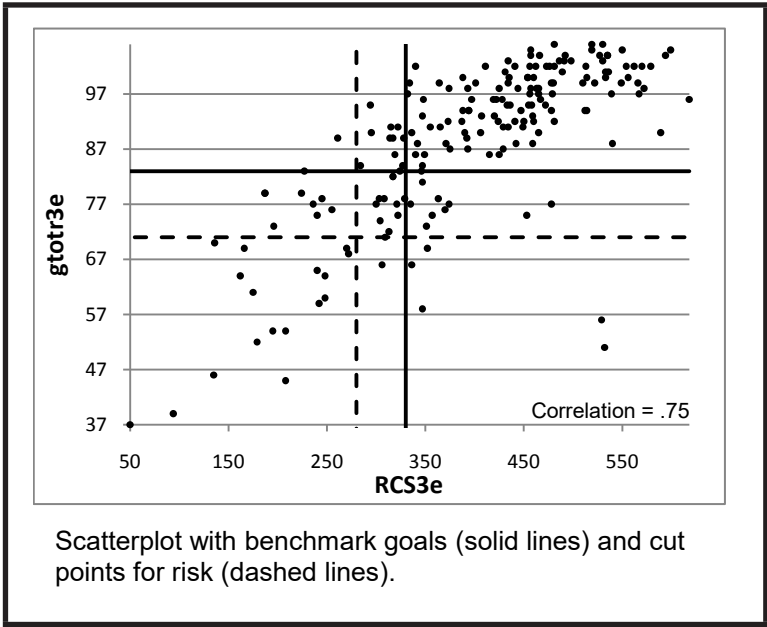
Contingency Table

The contingency table is presented directly below the scatterplot. The contingency table summarizes the number of students scoring in each zone of the scatterplot. The predictor variable forms the columns of the table and the criterion variable forms the rows. In this way, the contingency table corresponds directly to the zones of the scatterplot. The likely need for support is used to label levels of performance on the predictor; score level is used to label levels of performance on the criterion. The marginal total number of students in each level of the predictor is provided in the first column margin; the marginal total number of students scoring in each level of the criterion is provided in the first row margin. Marginal percents of the total number of students are provided in the second margin. Below the table, the odds of students with that screening decision achieving the benchmark goal on the criterion are provided. The odds were obtained as a conditional percent of students achieving the criterion goal given that level of performance on the predictor.

For example, the contingency table for the third-grade, end-of-year Reading Composite Score as the predictor and the GRADE Total Test Raw Score also administered at end of year as the criterion is illustrated in Figure 4.3. The upper-left cell of the table, which corresponds to the upper-left zone of the scatterplot, indicates that there were two students who scored in the *Likely to Need Intensive Support* level of the predictor who also achieved the goal (*At or Above Benchmark*) on the GRADE external criterion measure. Overall, there were 27 students who scored in the *Likely to Need Intensive Support* range on the predictor, which was 14% of the total sample. That marginal percent is similar to the 12% of the sample that scored in the *Well Below Benchmark* range on the GRADE external criterion. Of the 27 students who were identified as *Likely to Need Intensive Support* on

the predictor, only two (7%) achieved the goal. Thus, the odds were about 7% of achieving the benchmark goal for students with a screening decision of *Likely to Need Intensive Support*. For students who were identified as *Likely to Need Core Support* on the predictor, the odds of achieving the goal were 90%. For students who were identified as *Likely to Need Strategic Support*, the odds were 48%.

Figure 4.3 The contingency table summarizes the number of students in each zone of the scatterplot, marginal totals, marginal percents, and the odds of students with a specific screening decision (e.g., *Likely to Need Intensive Support*) achieving the goal on the criterion.

| | RCS3e Screening Decision: | | | | |
gtotr3e Outcome:	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
At or Above Benchmark	2	11	123	136	73%
Below Benchmark	8	11	9	28	15%
Well Below Benchmark	17	1	5	23	12%
Marginal Total	27	23	137	187	
Marginal Percent	14%	12%	73%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	7%	48%	90%		

Logistic Regression Analysis

The logistic regression analysis is provided directly to the right of the scatterplot. The small dots represent the moving percent (or likelihood) of students with that particular score on the predictor to achieve the benchmark goal on the outcome. To calculate those points, the students were rank-ordered from lowest to highest on the predictor score, and then overlapping intervals of students were created based on their scores on the predictor. The first interval started with the lowest score on the predictor and went up to the score necessary to include at least 20 students. All students with a score in that range were included in the interval. For each subsequent interval, the lowest numeric score on the predictor was dropped, along with all students with that score, and the upper limit was raised until again at least 20 students were included in the interval. For example, the first interval might include all students with scores from 0 to 4 (if that was the minimum range necessary to include at least 20 students), and the second interval might include all students with scores from 1 to 7. Some intervals included more than 20 students because multiple students obtained the maximum or minimum score for the interval. For each interval, the midpoint score on the predictor was identified and plotted on the horizontal axis. In the examples above, the midpoint of 0 to 4 is 2, and the midpoint of 1 to 7 is 4. The percent of students within the interval who achieved the benchmark goal on the outcome was calculated, and plotted on the vertical axis. The solid line connects the data points across intervals.

A logistic regression was performed predicting the odds of scoring at or above benchmark on the criterion based on the score on the predictor. The dashed line represents the logistic line of best fit to the data points. The large solid dot on the logistic regression line represents the predicted odds of achieving the goal for students who obtain the exact benchmark goal. The large open dot on the logistic regression line represents the predicted odds of achieving the goal for students who obtain the exact cut point for risk. Sometimes the logistic regression

line provided a very good fit to the data and assisted in establishing benchmark goals and cut points for risk, but sometimes the model provided a poor fit to the data and was interpreted with caution.

For example, the third-grade, end-of-year Reading Composite Score and GRADE Total Test Raw Score logistic regression analysis is represented in Figure 4.4. The model fits fairly well and contributed to establishing benchmark goals and cut points for risk. Using the logistic regression model, the predicted odds of achieving the goal for a student exactly at benchmark on the predictor is 67%. The predicted odds of achieving the goal for a student exactly at the cut point for risk on the predictor is 32%.

> Figure 4.4 The logistic regression analysis summarizes the moving percent of students achieving the goal (solid line connecting small dots) and the logistic regression line fit to the moving percents (dashed line) with benchmark goal (large solid dot) and cut point for risk (large open dot).

Logistic regression with goal (solid dot) and cut point (open dot).
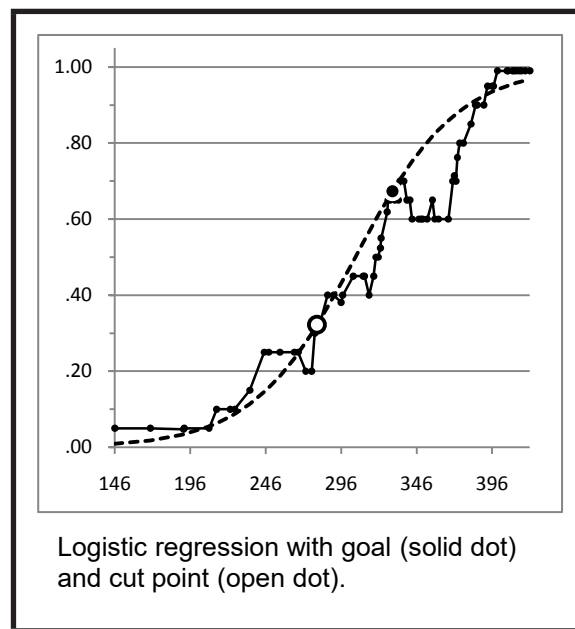
Receiver Operator Characteristic Curve Analysis

The receiver operator characteristic (ROC) curve analysis is summarized directly to the right of the logistic regression analysis. The ROC curve is plotted by considering each possible score of the predictor as a potential decision point (either benchmark goal or cut point for risk). For each potential decision point, the number of True Positives (TP), False Positives (FP), True Negatives (TN), and False Negatives (FN) are computed. True Positives refer to the number of students who were below the predictor score and who do not reach the goal (i.e., the screener indicated they would not reach the goal and was correct because they did not). False Positives refer to the number of students who were below the predictor score who did reach the goal (i.e., the screener or predictor indicated they would not reach the goal and the screener was in error because they did achieve the goal). Similarly, True Negatives are the number of students who were above the predictor score who did achieve the goal, and False Negatives are the number of students who were above the predictor score but did not achieve the goal. The horizontal axis of the ROC curve is the False Positive Rate or 1 – Specificity calculated by FP / (FP + TN). The vertical axis is the True Positive Rate or Sensitivity calculated by TP / (TP + FN). In general there is a trade-off of sensitivity and specificity: as higher scores are considered for the decision rule, the sensitivity of the decision increases but the specificity declines. When the curve extends higher into the upper-left corner of the graph and the area under the curve (AUC) increases, there is a more favorable trade-off of sensitivity and specificity.

We should note here that we are troubled by the terminology of True Positives and False Positives. The premise of Acadience Reading is that there are no True Positives. In biomedical research, True Positive refers to subjects who were identified as likely to have the condition on the screening test (for example, they screened positive for tuberculosis) and who were then determined to actually have the condition (i.e., they were true positives). However, in an educational context, students who are at risk for later difficulty are provided with additional support for the explicit purpose of changing the outcome. Thus, in an educational context, students who are labeled as True Positives are actually students for whom an effective intervention was not provided. When calculating sensitivity and specificity in biomedical research, a False Positive represents an error in the screening decision. For example, a subject might be screened as positive for tuberculosis, but then determined not to have the condition (i.e., a false screening decision of positive). However, in an educational context, a student who is classified as a False Positive may not represent a screening error. The student may, in fact, have been accurately identified as at risk for a poor outcome and been provided with an effective intervention that actively changed the outcome. As educators, it is our charge to ruin predictions of difficulty. Similarly, it is not possible to distinguish between a False Negative (a student who is identified as needing core support, but who does not reach future benchmark goals) and an ineffective core curriculum or instruction.

In the context of Acadience Reading, we are interested in two levels of performance on the outcome or criterion measure: *At or Above Benchmark* where we are reasonably confident the student is making adequate progress, and *Well Below Benchmark* where we are reasonably confident the student is not making adequate progress. Thus, we compute two ROC curves for each analysis: one representing the benchmark goal on the criterion measure (solid line) and a second representing the cut point for risk on the criterion (dashed line). We also are interested in the sensitivity and specificity of two particular predictor scores: the benchmark goal on the predictor (large solid dot), and the cut point for risk on the predictor (large open dot).

For example, the ROC curves for the third-grade, end-of-year Reading Composite Score as predictor with respect to the GRADE Total Test Raw Score at end-of-year third grade is illustrated in Figure 4.5. The curves are both similar, and toward the upper-left corner of the graph with AUCs of .90 and .87, indicating an excellent trade-off of sensitivity and specificity. The benchmark goal is in the upper-left corner of the curve, indicating a good balance of sensitivity and specificity for that decision point. The cut point for risk on the benchmark goal ROC is lower on the curve, indicating a less than optimum balance of sensitivity and specificity for that decision point.

Figure 4.5 The receiver operator characteristic (ROC) curve analysis summarizes the trade-off of sensitivity (vertical axis) and specificity (1 – specificity on the horizontal axis).

Receiver Operator Characteristic (ROC) curves.
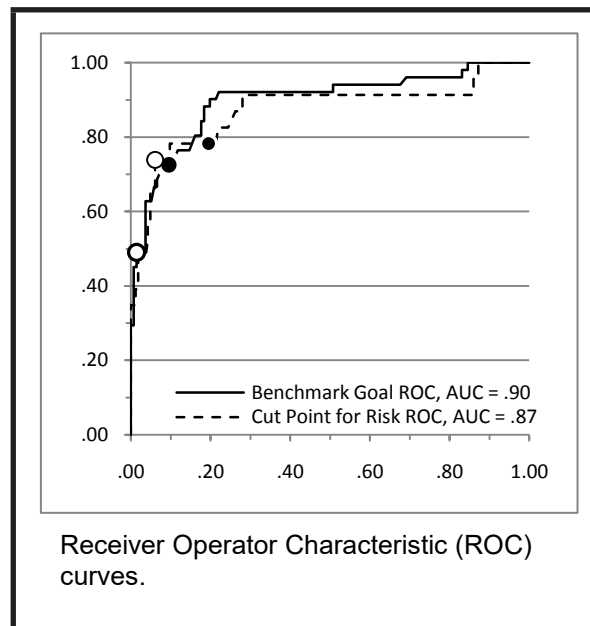
Other Decision-Utility Indices

The final section of the benchmark goal analysis detail is a table summarizing other decision-utility indices that were considered. The table of other indices is located in the lower right corner of the analysis detail page. These decision-utility indices were developed primarily for 2-by-2 decisions: 2 levels on the predictor (at risk/positive or not at risk/negative) and 2 levels on the criterion (above goal/negative or below goal/positive). For Acadience Reading, we are interested in 3 levels on the predictor and 3 levels on the criterion. We handled this distinction by constructing and evaluating four 2-by-2 decision tables: (a) above or below benchmark on the criterion and above or below benchmark on the predictor, (b) above or below benchmark on the criterion and above or below the cut point for risk on the predictor, (c) above or below cut point for risk on the criterion and above or below benchmark on the predictor, (b) above or below cut point for risk on the criterion and above or below the cut point for risk on the predictor.

For example, the other decision-utility indices for the third-grade, end-of-year Reading Composite Score as predictor with respect to the GRADE Total Test Raw Score at end-of-year third grade are summarized in Figure 4.6. The greatest sensitivity is obtained when the predictor decision is whether the student is above or below the benchmark goal and the criterion outcome is whether the student is above or below the cut point for risk (.78). The greatest specificity is obtained when the predictor decision is whether the student is above or below the cut point for risk and the criterion outcome is whether the student is above or below the benchmark goal (.99). The percent accurate classification ranged from .80 to .91, and Kappa coefficients ranged from .39 to .63.

Figure 4.6 Other decision-utility indices

	At or Above Benchmark outcome		Well Below Benchmark outcome	
	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	123	134	132	154
False Negative	14	26	5	6
True Positive	37	25	18	17
False Positive	13	2	32	10
Sensitivity	.73	.49	.78	.74
Specificity	.90	.99	.80	.94
Negative Predictive Power	.90	.84	.96	.96
Positive Predictive Power	.74	.93	.36	.63
Accurate Classification	.86	.85	.80	.91
Kappa	.63	.56	.39	.63

Reading Composite Score Details Pages

Figures 4.7 through 4.27 are the detail pages for each grade and time of year of the Reading Composite Score. The detail pages for all Acadience Reading measures are available in the *Benchmark Goals Study Technical Report* (Tech Report No. 11) at Dynamic Measurement Group's website, https://acadiencelearning.org/.

Figure 4.7 Benchmark Goal Detail

Screening Decision	Role	Variable	Goal	Cut Point	Description
Outcome	Predictor	RCS0b	26	13	Reading Composite Score, Grade K, Beginning of Year
	Criterion	RCS0m	122	85	Reading Composite Score, Grade K, Middle of Year

Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).

Logistic regression with goal (solid dot) and cut point (open dot).

Receiver Operator Characteristic (ROC) curves.

RCS0b Screening Decision:

RCS0m Outcome:	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
At or Above Benchmark	17	36	265	318	68%
Below Benchmark	24	25	46	95	20%
Well Below Benchmark	37	11	6	54	12%
Marginal Total	78	72	317	467	
Marginal Percent	17%	15%	68%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	22%	50%	84%		

	At or Above Benchmark outcome		Well Below Benchmark outcome	
	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	265	301	311	372
False Negative	52	88	6	17
True Positive	97	61	48	37
False Positive	53	17	102	41
Sensitivity	.65	.41	.89	.69
Specificity	.83	.95	.75	.90
Negative Predictive Power	.84	.77	.98	.96
Positive Predictive Power	.65	.78	.32	.47
Accurate Classification	.78	.78	.77	.88
Kappa	.48	.41	.36	.49

Figure 4.8 Benchmark Goal Detail

Screening Decision	Role	Variable	Goal	Cut Point	Description
Outcome	Predictor	RCS0m	122	85	Reading Composite Score, Grade K, Middle of Year
	Criterion	RCS0e	119	89	Reading Composite Score, Grade K, End of Year

Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).

Logistic regression with goal (solid dot) and cut point (open dot).

Receiver Operator Characteristic (ROC) curves.

RCS0m Screening Decision:

RCS0e Outcome:	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
At or Above Benchmark	18	36	263	317	68%
Below Benchmark	17	37	52	106	23%
Well Below Benchmark	17	22	2	41	9%
Marginal Total	52	95	317	464	
Marginal Percent	11%	20%	68%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	35%	38%	83%		

	At or Above Benchmark outcome		Well Below Benchmark outcome	
	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	263	299	315	388
False Negative	54	113	2	24
True Positive	93	34	39	17
False Positive	54	18	108	35
Sensitivity	.63	.23	.95	.41
Specificity	.83	.94	.74	.92
Negative Predictive Power	.83	.73	.99	.94
Positive Predictive Power	.63	.65	.27	.33
Accurate Classification	.77	.72	.76	.87
Kappa	.46	.21	.32	.30

Figure 4.9 Benchmark Goal Detail

Screening Decision	Role	Variable	Goal	Cut Point	Description
Outcome	Predictor	RCS0e	119	89	Reading Composite Score, Grade K, End of Year
	Criterion	gtotr0e	74	66	GRADE Total Test, Grade K, End of Year

Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).

Logistic regression with goal (solid dot) and cut point (open dot).

Receiver Operator Characteristic (ROC) curves.

gtotr0e Outcome:

RCS0e Screening Decision:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
At or Above Benchmark	4	13	96	113	68%
Below Benchmark	5	5	22	32	19%
Well Below Benchmark	2	8	11	21	13%
Marginal Total	11	26	129	166	
Marginal Percent	7%	16%	78%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	36%	50%	74%		

	At or Above Benchmark outcome		Well Below Benchmark outcome	
	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	96	109	118	136
False Negative	33	46	11	19
True Positive	20	7	10	2
False Positive	17	4	27	9
Sensitivity	.38	.13	.48	.10
Specificity	.85	.96	.81	.94
Negative Predictive Power	.74	.70	.91	.88
Positive Predictive Power	.54	.64	.27	.18
Accurate Classification	.70	.70	.77	.83
Kappa	.25	.12	.22	.04

Figure 4.10 Benchmark Goal Detail

Screening Decision Outcome

	Role	Variable	Goal	Cut Point	Description
	Predictor	RCS1b	113	97	Reading Composite Score, Grade 1, Beginning of Year
	Criterion	RCS1m	130	100	Reading Composite Score, Grade 1, Middle of Year

Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).

Correlation = .79

Logistic regression with goal (solid dot) and cut point (open dot).

Receiver Operator Characteristic (ROC) curves.

Benchmark Goal ROC, AUC = .86
Cut Point for Risk ROC, AUC = .85

RCS1b Screening Decision:

RCS1m Outcome:	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal Total	Marginal percent
At or Above Benchmark	21	22	254	297	67%
Below Benchmark	19	19	27	65	15%
Well Below Benchmark	42	21	20	83	19%
Marginal Total	82	62	301	445	
Marginal Percent	18%	14%	68%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	26%	35%	84%		

	At or Above Benchmark outcome		Well Below Benchmark outcome	
	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	254	276	281	322
False Negative	47	87	20	41
True Positive	101	61	63	42
False Positive	43	21	81	40
Sensitivity	.68	.41	.76	.51
Specificity	.86	.93	.78	.89
Negative Predictive Power	.84	.76	.93	.89
Positive Predictive Power	.70	.74	.44	.51
Accurate Classification	.80	.76	.77	.82
Kappa	.54	.38	.42	.40

Figure 4.11 Benchmark Goal Detail

Screening Decision Outcome	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	RCS1m	130	100	Reading Composite Score, Grade 1, Middle of Year
Outcome	Criterion	RCS1e	155	111	Reading Composite Score, Grade 1, End of Year

Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines). Correlation = .89

Logistic regression with goal (solid dot) and cut point (open dot).

Receiver Operator Characteristic (ROC) curves.

Benchmark Goal ROC, AUC = .94
Cut Point for Risk ROC, AUC = .94

RCS1m Screening Decision:

RCS1e Outcome:	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
At or Above Benchmark	10	22	263	295	67%
Below Benchmark	19	24	23	66	15%
Well Below Benchmark	54	18	7	79	18%
Marginal Total	83	64	293	440	
Marginal Percent	19%	15%	67%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	12%	34%	90%		

	At or Above Benchmark outcome		Well Below Benchmark outcome	
	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	263	285	286	332
False Negative	30	72	7	25
True Positive	115	73	72	54
False Positive	32	10	75	29
Sensitivity	.79	.50	.91	.68
Specificity	.89	.97	.79	.92
Negative Predictive Power	.90	.80	.98	.93
Positive Predictive Power	.78	.88	.49	.65
Accurate Classification	.86	.81	.81	.88
Kappa	.68	.53	.53	.59

Figure 4.12 Benchmark Goal Detail

Screening Decision	Role	Variable	Goal	Cut Point	Description
Outcome	Predictor	RCS1e	155	111	Reading Composite Score, Grade 1, End of Year
	Criterion	gtotr1e	62	50	GRADE Total Test, Grade 1, End of Year

Correlation = .77

Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).

Logistic regression with goal (solid dot) and cut point (open dot).

Receiver Operator Characteristic (ROC) curves.

Benchmark Goal ROC, AUC = .90
Cut Point for Risk ROC, AUC = .90

RCS1e Screening Decision:

gtotr1e Outcome:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
At or Above Benchmark	3	13	123	139	72%
Below Benchmark	7	11	9	27	14%
Well Below Benchmark	19	3	5	27	14%
Marginal Total	29	27	137	193	
Marginal Percent	15%	14%	71%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	10%	48%	90%		

	At or Above Benchmark outcome		Well Below Benchmark outcome	
	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	123	136	132	156
False Negative	14	28	5	8
True Positive	40	26	22	19
False Positive	16	3	34	10
Sensitivity	.74	.48	.81	.70
Specificity	.88	.98	.80	.94
Negative Predictive Power	.90	.83	.96	.95
Positive Predictive Power	.71	.90	.39	.66
Accurate Classification	.84	.84	.80	.91
Kappa	.62	.54	.42	.62

Figure 4.13 Benchmark Goal Detail

Screening Decision	Role	Variable	Goal	Cut Point	Description
Outcome	Predictor	RCS2b	141	109	Reading Composite Score, Grade 2, Beginning of Year
	Criterion	RCS2m	190	145	Reading Composite Score, Grade 2, Middle of Year

Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).

Logistic regression with goal (solid dot) and cut point (open dot).

Receiver Operator Characteristic (ROC) curves.

Correlation = .86

Benchmark Goal ROC, AUC = .94
Cut Point for Risk ROC, AUC = .94

RCS2b Screening Decision:

RCS2m Outcome:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
At or Above Benchmark	10	25	356	391	72%
Below Benchmark	17	19	22	58	11%
Well Below Benchmark	70	17	10	97	18%
Marginal Total	97	61	388	546	
Marginal Percent	18%	11%	71%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	10%	41%	92%		

	At or Above Benchmark outcome		Well Below Benchmark outcome	
	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	356	381	378	422
False Negative	32	68	10	27
True Positive	123	87	87	70
False Positive	35	10	71	27
Sensitivity	.79	.56	.90	.72
Specificity	.91	.97	.84	.94
Negative Predictive Power	.92	.85	.97	.94
Positive Predictive Power	.78	.90	.55	.72
Accurate Classification	.88	.86	.85	.90
Kappa	.70	.60	.59	.66

Figure 4.14 Benchmark Goal Detail

Screening Decision	Role	Variable	Goal	Cut Point	Description
Outcome	Predictor	RCS2m	190	145	Reading Composite Score, Grade 2, Middle of Year
	Criterion	RCS2e	238	180	Reading Composite Score, Grade 2, End of Year

Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).

Logistic regression with goal (solid dot) and cut point (open dot).

Receiver Operator Characteristic (ROC) curves.

Correlation = .89

Benchmark Goal ROC, AUC = .96
Cut Point for Risk ROC, AUC = .96

RCS2m Screening Decision:

RCS2e Outcome:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
At or Above Benchmark	9	22	363	394	71%
Below Benchmark	26	31	24	81	15%
Well Below Benchmark	66	7	6	79	14%
Marginal Total	101	60	393	554	
Marginal Percent	18%	11%	71%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	9%	37%	92%		

	At or Above Benchmark outcome		Well Below Benchmark outcome	
	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	363	385	387	440
False Negative	30	68	6	13
True Positive	130	92	73	66
False Positive	31	9	88	35
Sensitivity	.81	.58	.92	.84
Specificity	.92	.98	.81	.93
Negative Predictive Power	.92	.85	.98	.97
Positive Predictive Power	.81	.91	.45	.65
Accurate Classification	.89	.86	.83	.91
Kappa	.73	.62	.52	.68

Figure 4.15 Benchmark Goal Detail

Screening Decision Outcome	Role	Variable	Goal	Cut Point	Description
	Predictor	RCS2e	238	180	Reading Composite Score, Grade 2, End of Year
	Criterion	gtotr2e	84	74	GRADE Total Test, Grade 2, End of Year

Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).

Correlation = .75

Logistic regression with goal (solid dot) and cut point (open dot).

Receiver Operator Characteristic (ROC) curves.

Benchmark Goal ROC, AUC = .88
Cut Point for Risk ROC, AUC = .98

RCS2e Screening Decision:

gtotr2e Outcome:	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
At or Above Benchmark	4	14	141	159	73%
Below Benchmark	3	10	17	30	14%
Well Below Benchmark	21	7	0	28	13%
Marginal Total	28	31	158	217	
Marginal Percent	13%	14%	73%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	14%	45%	89%		

	At or Above Benchmark outcome		Well Below Benchmark outcome	
	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	141	155	158	182
False Negative	17	34	0	7
True Positive	41	24	28	21
False Positive	18	4	31	7
Sensitivity	.71	.41	1.00	.75
Specificity	.89	.97	.84	.96
Negative Predictive Power	.89	.82	1.00	.96
Positive Predictive Power	.69	.86	.47	.75
Accurate Classification	.84	.82	.86	.94
Kappa	.59	.47	.57	.71

Figure 4.16 Benchmark Goal Detail

Screening Decision Outcome	Role	Variable	Goal	Cut Point	Description
	Predictor	RCS3b	220	180	Reading Composite Score, Grade 3, Beginning of Year
	Criterion	RCS3m	285	235	Reading Composite Score, Grade 3, Middle of Year

Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).

Logistic regression with goal (solid dot) and cut point (open dot).

Receiver Operator Characteristic (ROC) curves.

Correlation = .91

Benchmark Goal ROC, AUC = .95
Cut Point for Risk ROC, AUC = .97

RCS3b Screening Decision:

RCS3m Outcome:	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
At or Above Benchmark	4	22	324	350	71%
Below Benchmark	20	16	21	57	12%
Well Below Benchmark	70	9	4	83	17%
Marginal Total	94	47	349	490	
Marginal Percent	19%	10%	71%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	4%	47%	93%		

	At or Above Benchmark outcome		Well Below Benchmark outcome	
	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	324	346	345	383
False Negative	25	50	4	13
True Positive	115	90	79	70
False Positive	26	4	62	24
Sensitivity	.82	.64	.95	.84
Specificity	.93	.99	.85	.94
Negative Predictive Power	.93	.87	.99	.97
Positive Predictive Power	.82	.96	.56	.74
Accurate Classification	.90	.89	.87	.92
Kappa	.75	.70	.63	.75

Figure 4.17 Benchmark Goal Detail

	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	RCS3m	285	235	Reading Composite Score, Grade 3, Middle of Year
Outcome	Criterion	RCS3e	330	280	Reading Composite Score, Grade 3, End of Year

Correlation = .90
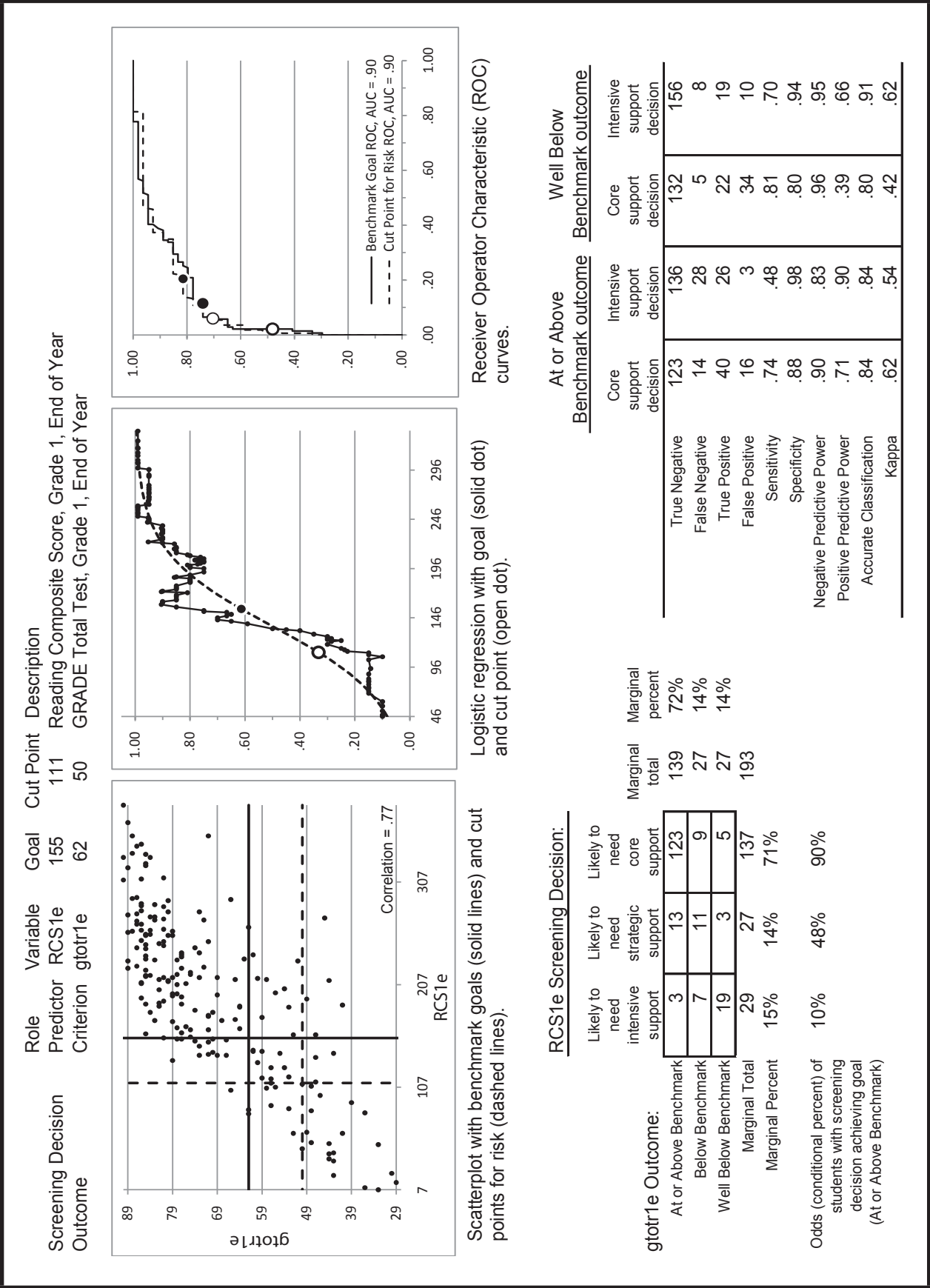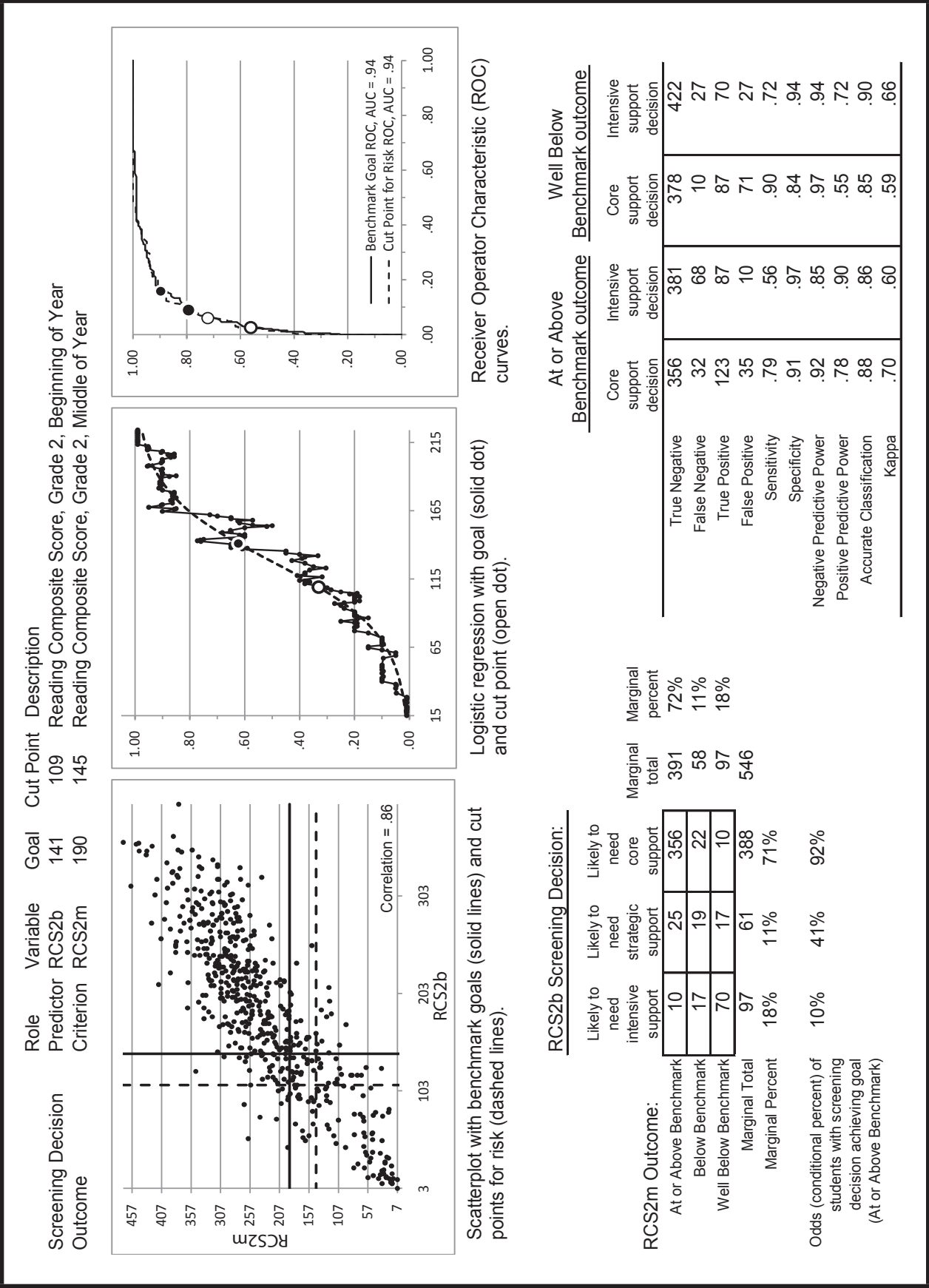
Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).

Logistic regression with goal (solid dot) and cut point (open dot).

Receiver Operator Characteristic (ROC) curves.

Benchmark Goal ROC, AUC = .94
Cut Point for Risk ROC, AUC = .95

RCS3m Screening Decision:

RCS3e Outcome:	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
At or Above Benchmark	6	25	318	349	72%
Below Benchmark	19	18	26	63	13%
Well Below Benchmark	55	15	4	74	15%
Marginal Total	80	58	348	486	
Marginal Percent	16%	12%	72%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	8%	43%	91%		

	At or Above Benchmark outcome		Well Below Benchmark outcome	
	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	318	343	344	387
False Negative	30	63	4	19
True Positive	107	74	70	55
False Positive	31	6	68	25
Sensitivity	.78	.54	.95	.74
Specificity	.91	.98	.83	.94
Negative Predictive Power	.91	.84	.99	.95
Positive Predictive Power	.78	.93	.51	.69
Accurate Classification	.87	.86	.85	.91
Kappa	.69	.60	.58	.66

Figure 4.18 Benchmark Goal Detail

Screening Decision Outcome	Role	Variable	Goal	Cut Point	Description
	Predictor	RCS3e	330	280	Reading Composite Score, Grade 3, End of Year
	Criterion	gtotr3e	83	71	GRADE Total Test, Grade 3, End of Year

Correlation = .75

Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).

Logistic regression with goal (solid dot) and cut point (open dot).

Receiver Operator Characteristic (ROC) curves.

Benchmark Goal ROC, AUC = .90
Cut Point for Risk ROC, AUC = .87

RCS3e Screening Decision:

gtotr3e Outcome:	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
At or Above Benchmark	2	11	123	136	73%
Below Benchmark	8	11	9	28	15%
Well Below Benchmark	17	1	5	23	12%
Marginal Total	27	23	137	187	
Marginal Percent	14%	12%	73%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	7%	48%	90%		

	At or Above Benchmark outcome		Well Below Benchmark outcome	
	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	123	134	132	154
False Negative	14	26	5	6
True Positive	37	25	18	17
False Positive	13	2	32	10
Sensitivity	.73	.49	.78	.74
Specificity	.90	.99	.80	.94
Negative Predictive Power	.90	.84	.96	.96
Positive Predictive Power	.74	.93	.36	.63
Accurate Classification	.86	.85	.80	.91
Kappa	.63	.56	.39	.63

Figure 4.19 Benchmark Goal Detail

Screening Decision Outcome	Role	Variable	Goal	Cut Point	Description
	Predictor	RCS4b	290	245	Reading Composite Score, Grade 4, Beginning of Year
	Criterion	RCS4m	330	290	Reading Composite Score, Grade 4, Middle of Year

Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).

Correlation = .92

Logistic regression with goal (solid dot) and cut point (open dot).

Receiver Operator Characteristic (ROC) curves.

Benchmark Goal ROC, AUC = .95
Cut Point for Risk ROC, AUC = .96

	At or Above Benchmark outcome		Well Below Benchmark outcome	
	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	351	377	372	424
False Negative	29	72	8	25
True Positive	158	115	112	95
False Positive	33	7	79	27
Sensitivity	.84	.61	.93	.79
Specificity	.91	.98	.82	.94
Negative Predictive Power	.92	.84	.98	.94
Positive Predictive Power	.83	.94	.59	.78
Accurate Classification	.89	.86	.85	.91
Kappa	.75	.66	.62	.73

RCS4b Screening Decision:

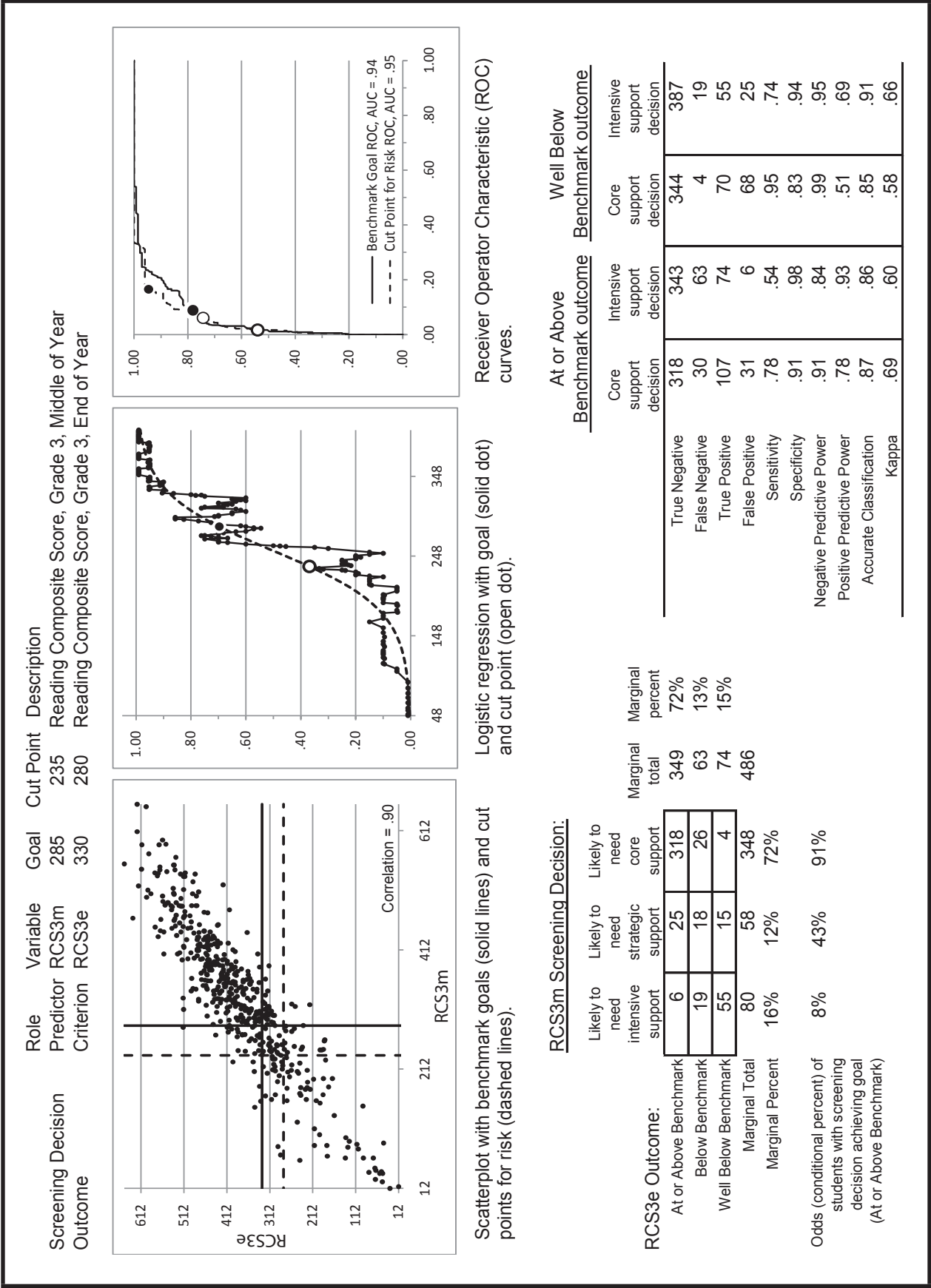| RCS4m Outcome:	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
At or Above Benchmark	7	26	351	384	67%
Below Benchmark	20	26	21	67	12%
Well Below Benchmark	95	17	8	120	21%
Marginal Total	122	69	380	571	
Marginal Percent	21%	12%	67%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	6%	38%	92%		

Figure 4.20 Benchmark Goal Detail

	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	RCS4m	330	290	Reading Composite Score, Grade 4, Middle of Year
Outcome	Criterion	RCS4e	391	330	Reading Composite Score, Grade 4, End of Year

Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).

Correlation = .90

(axes: RCS4m, RCS4e)

Logistic regression with goal (solid dot) and cut point (open dot).

Receiver Operator Characteristic (ROC) curves.

Benchmark Goal ROC, AUC = .94
Cut Point for Risk ROC, AUC = .96

RCS4m Screening Decision:

RCS4e Outcome:	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
At or Above Benchmark	12	28	347	387	67%
Below Benchmark	34	34	34	102	18%
Well Below Benchmark	76	6	3	85	15%
Marginal Total	122	68	384	574	
Marginal Percent	21%	12%	67%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	10%	41%	90%		

	At or Above Benchmark outcome		Well Below Benchmark outcome	
	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	347	375	381	443
False Negative	37	77	3	9
True Positive	150	110	82	76
False Positive	40	12	108	46
Sensitivity	.80	.59	.96	.89
Specificity	.90	.97	.78	.91
Negative Predictive Power	.90	.83	.99	.98
Positive Predictive Power	.79	.90	.43	.62
Accurate Classification	.87	.84	.81	.90
Kappa	.70	.61	.49	.68

Figure 4.21 Benchmark Goal Detail

	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	RCS4e	391	330	Reading Composite Score, Grade 4, End of Year
Outcome	Criterion	gtotr4e	51	39	GRADE Total Test, Grade 4, End of Year

Receiver Operator Characteristic (ROC) curves.

Benchmark Goal ROC, AUC = .88
Cut Point for Risk ROC, AUC = .94

	At or Above Benchmark outcome		Well Below Benchmark outcome	
	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	101	119	116	143
False Negative	19	32	4	8
True Positive	47	34	31	27
False Positive	19	1	35	8
Sensitivity	.71	.52	.89	.77
Specificity	.84	.99	.77	.95
Negative Predictive Power	.84	.79	.97	.95
Positive Predictive Power	.71	.97	.47	.77
Accurate Classification	.80	.82	.79	.91
Kappa	.55	.57	.49	.72

Logistic regression with goal (solid dot) and cut point (open dot).

RCS4e Screening Decision:

gtotr4e Outcome:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
At or Above Benchmark	1	18	101	120	65%
Below Benchmark	7	9	15	31	17%
Well Below Benchmark	27	4	4	35	19%
Marginal Total	35	31	120	186	
Marginal Percent	19%	17%	65%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	3%	58%	84%		

Correlation = .80

RCS4e

gtotr4e

Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).

Figure 4.22 Benchmark Goal Detail

Screening Decision	Role	Variable	Goal	Cut Point	Description
Outcome	Predictor	RCS5b	357	258	Reading Composite Score, Grade 5, Beginning of Year
	Criterion	RCS5m	372	310	Reading Composite Score, Grade 5, Middle of Year

Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).

Correlation = .89

RCS5b

Logistic regression with goal (solid dot) and cut point (open dot).

Receiver Operator Characteristic (ROC) curves.

Benchmark Goal ROC, AUC = .93
Cut Point for Risk ROC, AUC = .96

RCS5b Screening Decision:

RCS5m Outcome:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
At or Above Benchmark	3	29	304	336	64%
Below Benchmark	14	65	32	111	21%
Well Below Benchmark	58	21	1	80	15%
Marginal Total	75	115	337	527	
Marginal Percent	14%	22%	64%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	4%	25%	90%		

	At or Above Benchmark outcome		Well Below Benchmark outcome	
	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	304	333	336	430
False Negative	33	119	1	22
True Positive	158	72	79	58
False Positive	32	3	111	17
Sensitivity	.83	.38	.99	.73
Specificity	.90	.99	.75	.96
Negative Predictive Power	.90	.74	1.00	.95
Positive Predictive Power	.83	.96	.42	.77
Accurate Classification	.88	.77	.79	.93
Kappa	.73	.42	.47	.71

Figure 4.23 Benchmark Goal Detail

Screening Decision	Role	Variable	Goal	Cut Point	Description
Outcome	Predictor	RCS5m	372	310	Reading Composite Score, Grade 5, Middle of Year
	Criterion	RCS5e	415	340	Reading Composite Score, Grade 5, End of Year

Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).

Logistic regression with goal (solid dot) and cut point (open dot).

Receiver Operator Characteristic (ROC) curves.

RCS5m Screening Decision:

RCS5e Outcome:	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
At or Above Benchmark	2	34	291	327	64%
Below Benchmark	13	58	39	110	21%
Well Below Benchmark	61	15	1	77	15%
Marginal Total	76	107	331	514	
Marginal Percent	15%	21%	64%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	3%	32%	88%		

	At or Above Benchmark outcome		Well Below Benchmark outcome	
	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	291	325	330	422
False Negative	40	113	1	16
True Positive	147	74	76	61
False Positive	36	2	107	15
Sensitivity	.79	.40	.99	.79
Specificity	.89	.99	.76	.97
Negative Predictive Power	.88	.74	1.00	.96
Positive Predictive Power	.80	.97	.42	.80
Accurate Classification	.85	.78	.79	.94
Kappa	.68	.45	.47	.76

Benchmark Goal ROC, AUC = .94
Cut Point for Risk ROC, AUC = .98

Correlation = .90

Figure 4.24 Benchmark Goal Detail

Screening Decision	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	RCS5e	415	340	Reading Composite Score, Grade 5, End of Year
Outcome	Criterion	gtotr5e	47	35	GRADE Total Test, Grade 5, End of Year

Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).

Correlation = .77

Logistic regression with goal (solid dot) and cut point (open dot).

Receiver Operator Characteristic (ROC) curves.

Benchmark Goal ROC, AUC = .88
Cut Point for Risk ROC, AUC = .91

RCS5e Screening Decision:

gtotr5e Outcome:	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
At or Above Benchmark	2	14	115	131	68%
Below Benchmark	7	12	12	31	16%
Well Below Benchmark	21	5	5	31	16%
Marginal Total	30	31	132	193	
Marginal Percent	16%	16%	68%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	7%	45%	87%		

	At or Above Benchmark outcome		Well Below Benchmark outcome	
	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	115	129	127	153
False Negative	17	34	5	10
True Positive	45	28	26	21
False Positive	16	2	35	9
Sensitivity	.73	.45	.84	.68
Specificity	.88	.98	.78	.94
Negative Predictive Power	.87	.79	.96	.94
Positive Predictive Power	.74	.93	.43	.70
Accurate Classification	.83	.81	.79	.90
Kappa	.61	.50	.45	.63

Figure 4.25 Benchmark Goal Detail

Screening Decision	Role	Variable	Goal	Cut Point	Description
Outcome	Predictor	RCS6b	344	280	Reading Composite Score, Grade 6, Beginning of Year
	Criterion	RCS6m	358	285	Reading Composite Score, Grade 6, Middle of Year

Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines). Correlation = .91

Logistic regression with goal (solid dot) and cut point (open dot).

Receiver Operator Characteristic (ROC) curves. Benchmark Goal ROC, AUC = .95; Cut Point for Risk ROC, AUC = .98

RCS6b Screening Decision:

RCS6m Outcome:	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
At or Above Benchmark	5	22	372	399	78%
Below Benchmark	9	30	25	64	13%
Well Below Benchmark	34	12	0	46	9%
Marginal Total	48	64	397	509	
Marginal Percent	9%	13%	78%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	10%	34%	94%		

	At or Above Benchmark outcome		Well Below Benchmark outcome	
	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	372	394	397	449
False Negative	25	67	0	12
True Positive	85	43	46	34
False Positive	27	5	66	14
Sensitivity	.77	.39	1.00	.74
Specificity	.93	.99	.86	.97
Negative Predictive Power	.94	.85	1.00	.97
Positive Predictive Power	.76	.90	.41	.71
Accurate Classification	.90	.86	.87	.95
Kappa	.70	.48	.52	.70

Figure 4.26 Benchmark Goal Detail

Screening Decision	Role	Variable	Goal	Cut Point	Description
Outcome	Predictor	RCS6m	358	285	Reading Composite Score, Grade 6, Middle of Year
	Criterion	RCS6e	380	324	Reading Composite Score, Grade 6, End of Year

Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).

Logistic regression with goal (solid dot) and cut point (open dot).

Receiver Operator Characteristic (ROC) curves.

RCS6m Screening Decision:

RCS6e Outcome:	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
At or Above Benchmark	4	23	374	401	79%
Below Benchmark	8	30	21	59	12%
Well Below Benchmark	34	9	5	48	9%
Marginal Total	46	62	400	508	
Marginal Percent	9%	12%	79%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	9%	37%	94%		

	At or Above Benchmark outcome		Well Below Benchmark outcome	
	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	374	397	395	448
False Negative	26	65	5	14
True Positive	81	42	43	34
False Positive	27	4	65	12
Sensitivity	.76	.39	.90	.71
Specificity	.93	.99	.86	.97
Negative Predictive Power	.94	.86	.99	.97
Positive Predictive Power	.75	.91	.40	.74
Accurate Classification	.90	.86	.86	.95
Kappa	.69	.48	.48	.70

Figure 4.27 Benchmark Goal Detail

Screening Decision Outcome	Role	Variable	Goal	Cut Point	Description
	Predictor	RCS6e	380	324	Reading Composite Score, Grade 6, End of Year
	Criterion	gtotr6e	48	37	GRADE Total Test, Grade 6, End of Year

Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).

Correlation = .73

Logistic regression with goal (solid dot) and cut point (open dot).

Receiver Operator Characteristic (ROC) curves.

Benchmark Goal ROC, AUC = .92
Cut Point for Risk ROC, AUC = .87

RCS6e Screening Decision:

gtotr6e Outcome:	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
At or Above Benchmark	1	5	80	86	82%
Below Benchmark	3	5	3	11	10%
Well Below Benchmark	4	1	3	8	8%
Marginal Total	8	11	86	105	
Marginal Percent	8%	10%	82%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	13%	45%	93%		

	At or Above Benchmark outcome		Well Below Benchmark outcome	
	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	80	85	83	93
False Negative	6	12	3	4
True Positive	13	7	5	4
False Positive	6	1	14	4
Sensitivity	.68	.37	.63	.50
Specificity	.93	.99	.86	.96
Negative Predictive Power	.93	.88	.97	.96
Positive Predictive Power	.68	.88	.26	.50
Accurate Classification	.89	.88	.84	.92
Kappa	.61	.46	.29	.46

Chapter 5: **Reliability**

This chapter reports the reliability of Acadience Reading. An overall summary of reliability estimates can be found in Table 5.18 on page 90.

Reliability refers to the relative stability with which a test measures the same skills across minor differences in conditions. Information about a test's reliability can be obtained in three different ways:

1) Alternate forms of the test with different items should give approximately the same scores.

2) The same test given at two different points in time should give approximately the same scores.

3) Two different examiners who give the test should obtain approximately the same scores.

Nunnally and Bernstein (1994) offer a hierarchy for estimating the reliability of a test. The most desirable choice is alternate-form reliability with a two-week interval, which addresses the first and second approach described above. The second choice is test-retest reliability with a two-week interval, which addresses the second approach described above. Both of these methods were used to assess the reliability of Acadience Reading. In addition, to address the third approach, inter-rater reliability was evaluated by examining how different assessors score the same test.

In this section we present a discussion of the types of reliability included and the sources of error for each. "Error" refers to unintended factors that contribute to changes in scores. For a discussion of error associated with the testing process, see the Types of Measurement Error section later in this chapter.

Alternate-Form Reliability. Alternate-form reliability indicates the extent to which test results generalize to different item samples. Students are tested with two different (i.e., alternate) but equivalent forms of the test and scores from these two forms are correlated. Student learning can be interpreted as error, and there may be a practice effect due to the similarity of the test, although reliability estimates from alternate forms are less subject to practice effects than test-retest reliability where the same form is administered twice.

Test-Retest Reliability. Test-retest reliability is an index of score stability or the degree to which results from student testing are replicated when the same test form is administered twice within a short interval. Scores from the two test administrations are then correlated. Sources of error are similar to those from alternate-form reliability. Student learning and practice effects can affect the reliability estimate; students may improve their performance on the second test after having become familiar with the test items on the first test.

Inter-Rater Reliability. Inter-rater reliability indicates the extent to which results generalize across assessors. The inter-rater reliability estimates reported here are based on two independent assessors simultaneously scoring student performance during a single test administration ("shadow-scoring"). The two raters' scores are then correlated. Score fluctuations are attributable to the raters' level of agreement. The

reliability coefficient presented in this chapter represents the reliability of the directions and scoring procedures of the measures themselves as interpreted by the assessors administering the measure. The sources of error associated with inter-rater reliability lie in the assessor, and include factors such as presenting an incorrect item, timing, or simple mistakes.

Salvia, Ysseldyke, & Bolt's (2007) standards for reliability were used to evaluate the reliability data for Acadience Reading. According to these standards, a minimum of .60 is required for administrative purposes and scores that are reported for groups of individuals, a minimum of .80 is required for screening decisions, and a minimum of .90 is required for important educational decisions concerning an individual student.

Reliability estimates are reported for individual test administrations of each measure as well as for the aggregate (mean or median) of three alternate forms. Three-form reliability estimates are provided to correspond to the recommended Acadience Reading practice of examining a pattern of performance on repeated assessments for increased confidence in decisions. For ORF benchmark assessments, three passages are administered and the median score is used. For progress monitoring and validating need for support, repeated assessment using alternate forms of the same measure is recommended (see Chapter 1). For ORF, the reliability of three-form triads is reported. For other measures, the reliability of three-form aggregates is estimated using the Spearman-Brown Prophecy Formula. Reliability estimates for the Reading Composite Score represent the reliability of an aggregate of multiple different measures administered at one time. Data gathered from five studies is reported; for more information about the studies referenced in this chapter, see Chapter 3.

Alternate-Form Reliability

Information about the alternate-form reliability of all Acadience Reading measures was gathered from five studies.

In Study A, alternate-form reliability information was collected on First Sound Fluency (FSF). The FSF measure was administered at key benchmark time periods. Alternate forms of FSF were given at monthly intervals between the fall and winter benchmark testing to assess the reliability of FSF. A total of 317 randomly selected kindergarten students from one district and all 56 kindergarten students from a second district participated in the alternate-form testing. The alternate-form reliability results from Study A are reported in Table 5.1. Overall, a single form has sufficient reliability for screening decisions, and a three-form aggregate has sufficient reliability for important individual decisions.

In Study B, alternate forms of the Acadience Reading measures were given approximately two weeks after the middle-of-year benchmark assessment. A total of 687 students in kindergarten through fifth grade from five schools participated. Reliability estimates from Study B are reported for FSF and Maze in Table 5.2. The coefficients indicate sufficient reliability for screening decisions. With repeated assessment across multiple forms, reliability increases substantially as noted in the far-right column of Table 5.2 where the estimated three-form reliability is reported. The three-form aggregates are sufficiently reliable for important individual decisions. For LNF, PSF, and NWF, the effects of changing the directions and scoring procedures are discussed in Chapter 7.

In Study C, alternate forms of all Acadience Reading measures were given approximately two weeks after the middle-of-year benchmark assessment. A stratified random sample was selected based on student performance from the beginning-of-year Acadience Reading benchmark assessment. Of the 322 students randomly selected across kindergarten through sixth grade, 166 students returned IRB-approved consent forms and were given an alternate form of each Acadience Reading measure. Data are not reported where sample sizes were not adequate to provide a stable reliability estimate. Results for Study C are reported in Tables 5.3, 5.4, 5.5, and 5.6. For individual scores, most coefficients are above .80, indicating sufficient reliability for screening decisions.

Several coefficients are above .90, indicating sufficient reliability for important individual educational decisions. For the Reading Composite Score, reliability is consistently high across first through fifth grade.

In Study D, information was collected on the alternate-form reliability of individual ORF passages. The final sample included 140 students across first through sixth grade from two schools. Alternate-form reliability results from Study D are reported in Table 5.7. All coefficients are above .90, indicating excellent reliability for important individual decisions.

In Study E, information was collected on first-grade Phoneme Segmentation Fluency and sixth-grade Acadience Reading Oral Reading Fluency, Retell, and Maze. Alternate-form reliability results are reported in Table 5.8.

Overall, the alternate-form reliability of a single form of most Acadience Reading measures is sufficient for screening decisions and in many instances sufficient for important individual decisions. Alternate-form reliability for individual ORF passages is particularly strong, indicating high consistency between passages. Reliability estimates increase substantially to be sufficient for important individual decisions for most measures and grade levels when three-form aggregates are examined. Test results from multiple administrations of the same measure are highly reliable as indicated in the estimated three-form reliability coefficients. Even greater confidence in educational decisions can be attained by examining the student's pattern of performance on four or more alternate forms.

In addition to repeated assessments with the same measure, the aggregate of multiple different measures using the Reading Composite Score also provides highly reliable information for educational decisions. The Reading Composite Score provides the best estimate of the student's overall reading proficiency, and reliability for this score is above .90 for first through sixth grades, indicating sufficient reliability for important individual educational decisions. In general, the results presented here indicate that the Acadience Reading measures and the Reading Composite Score possess stability across forms for all grades.

Table 5.1 One-Month Alternate-Form Reliability for Kindergarten First Sound Fluency from Study A

	Descriptive Statistics			Reliability	
FSF by Administration	N	Mean	SD	1	2
1. First Administration	383	20.40	13.35	-	-
2. Second Administration	385	26.78	13.88	.82 (373)	-
3. Third Administration	363	32.21	13.48	.74 (355)	.82 (356)

Note. Based on Study A data. Pair-wise sample sizes for reliability coefficients are reported in parentheses. All correlations significant, $p < .001$.

Table 5.2 Two-Week Alternate-Form Reliability for First Sound Fluency and Maze from Study B

		First Form		Second Form		Reliability	
Study	N	Mean	SD	Mean	SD	Single-Form	Estimated Three-Form
First Sound Fluency	97	30.10	14.74	28.66	14.32	.83	.94
Maze Adjusted Score							
Third Grade	40	13.00	7.30	16.35	6.90	.75	.90
Fourth Grade	40	17.69	8.24	15.46	6.15	.81	.93
Fifth Grade	61	23.09	8.47	22.73	9.22	.83	.94

Note. Based on Study B middle-of-year data. Estimated three-form reliability is based on the Spearman-Brown Prophecy Formula. All correlations are significant, $p < .001$.

Table 5.3 Two-Week Alternate-Form Reliability for Kindergarten and First Grade Acadience Reading Measures from Study C

Acadience Reading Measures by Grade	N	First Form		Second Form		Reliability	
		Mean	SD	Mean	SD	Single-Form	Estimated Three-Form
Kindergarten							
First Sound Fluency	29	32.34	10.67	32.79	6.65	.52**	.76
Letter Naming Fluency	29	39.76	15.90	45.48	15.64	.86	.95
Phoneme Segmentation Fluency	29	25.45	14.46	29.97	11.43	.44	.70
NWF Correct Letter Sounds	27	17.37	10.78	21.89	14.82	.71	.88
NWF Whole Words Read	27	0.74	1.81	2.04	3.78	.92	.97
First Grade							
NWF Correct Letter Sounds	28	53.25	23.91	54.18	25.96	.85	.94
NWF Whole Words Read	28	9.50	12.00	10.29	12.52	.90	.96

Note. Based on Study C from middle-of-year data. The estimated three-form reliability is based on the Spearman-Brown Prophecy Formula. Unless marked, correlations significant, $p < .001$; ** $p < .01$.

Table 5.4 Two-Week Alternate-Form Reliability for Three-Passage Groups (Triads) of Acadience Reading Oral Reading Fluency Passages from Study C

ORF Scores by Grade	N	First Triad		Second Triad		Triad Reliability
		Mean	SD	Mean	SD	
ORF Words Correct						
First Grade	28	37.39	40.06	46.00	41.57	.98
Second Grade	24	75.08	42.06	82.46	38.01	.97
Third Grade	30	91.87	39.93	95.80	35.21	.96
Fourth Grade	30	104.47	39.48	110.43	37.86	.96
Fifth Grade	25	113.56	27.96	120.48	27.98	.95
ORF Accuracy						
First Grade	28	77%	15	84%	11	.88
Second Grade	24	91%	9	93%	8	.83
Third Grade	30	96%	4	95%	5	.80
Fourth Grade	30	96%	5	97%	4	.85
Fifth Grade	25	97%	2	98%	2	.76
ORF Retell						
Second Grade	20	26.6	13.32	29.73	17.22	.68
Third Grade	27	32.11	20.00	27.80	16.33	.81
Fourth Grade	30	34.17	18.16	38.50	18.74	.80
Fifth Grade	25	37.24	15.86	36.04	18.55	.65

Note. Based on Study C from middle-of-year data. 'Triad' refers to a group of three ORF passages, and the mean scores reported in this table represent the mean of the student-level median scores based upon a standardized benchmark administration of the triad. Data is unavailable for first-grade ORF Retell and all sixth-grade measures due to insufficient sample sizes. ORF passages are administered in triads, thus the alternate-form reliability is reported for triads. All correlations significant, $p < .001$.

Table 5.5 Two-Week Alternate-Form Reliability for Maze from Study C

Grade	N	First Form Mean	First Form SD	Second Form Mean	Second Form SD	Reliability Single-Form	Reliability Estimated Three-Form
Third Grade	24	11.13	7.83	13.75	7.93	.86	.95
Fourth Grade	29	16.34	5.92	20.93	7.28	.67	.86
Fifth Grade	20	13.15	5.96	23.35	8.41	.49*	.74

Note. Based on Study C from middle-of-year data. The estimated three-form reliability is based on the Spearman-Brown Prophecy Formula. Unless marked, correlations significant, $p < .001$; * $p < .05$.

Table 5.6 Two-Week Alternate-Form Reliability for Reading Composite Score from Study C

Grade	N	First Composite Mean	First Composite SD	Second Composite Mean	Second Composite SD	Reliability
Kindergarten	27	119.04	36.47	132.63	36.21	.66
First Grade	28	156.07	92.35	177.18	95.37	.95
Second Grade	24	183.08	108.63	209.08	99.35	.92
Third Grade	20	271.40	137.57	273.95	121.86	.97
Fourth Grade	25	317.80	118.60	359.56	123.74	.95
Fifth Grade	20	327.60	87.21	376.50	95.65	.91

Note. Based on middle-of-year Study C data. The first composite was calculated from middle-of-year benchmark assessment data. The second composite was calculated from alternate forms that were administered two weeks after middle-of-year benchmark assessment. All correlations significant, $p < .001$.

Table 5.7 Single-Passage and Three-Passage (Triad) Alternate-Form Reliability for Acadience Reading ORF Benchmark Passages from Study D

Grade	Number of: Students	Number of: Passages	Median of the: Passage Means	Median of the: Passage SDs	Median of the: Triad Means	Median of the: Triad SDs	Median Reliability Single-Passage	Median Reliability Triad
First Grade	23	29	81.52	43.11	81.63	43.91	.95	.97
Second Grade	25	32	115.12	36.53	114.68	35.18	.91	.94
Third Grade	22	32	109.89	39.13	110.44	38.01	.93	.97
Fourth Grade	23	32	131.87	31.99	132.47	31.01	.90	.94
Fifth Grade	23	32	136.24	36.07	137.33	34.62	.92	.96
Sixth Grade	24	32	150.99	28.63	148.02	27.63	.84	.90

Note. Based on Study D data. Every student read every passage for their grade level. Single-Passage reliability is the median alternate-form reliability between each Acadience Reading assessment passage and all other grade-level passages. Triad reliability is the median alternate-form reliability between groups of three Acadience Reading passages and all other grade-level three-passage groups. All correlations significant, $p < .001$

Table 5.8 Two-Week Alternate Form Reliability for Acadience Reading First Grade Phoneme Segmentation Fluency and Sixth Grade Oral Reading Fluency, Retell, Maze, and the Composite Score from Study E

Measure	N	Benchmark Administration		Alternate-Form Administration		Reliability	
		M	SD	M	SD	Single-Form	Three-Form
First Grade							
Phoneme Segmentation Fluency	164	49.19	13.84	53.57	12.81	.54	.78
Sixth Grade							
ORF Words Correct	61	127.46	28.59	127.10	28.83	--	.94
ORF Accuracy	61	.98	.02	.98	.02	--	.48
Retell	61	32.57	16.50	35.02	14.63	--	.62
Maze	60	27.03	8.89	23.68	10.50	.79	.89
Reading Composite Score	60	405.23	87.68	398.87	88.62	.91	--

Note. Based on Study E data. Based on beginning-of-year data. Three-form alternate-form reliability estimates for first grade PSF, sixth grade Maze, and sixth grade composite score were calculated using the Spearman-Brown Prophecy Formula. All correlations significant, $p < .001$.

Test-Retest Reliability

Test-retest data were collected by conducting testing in the two weeks following the middle-of-year benchmark assessment in Study C for all measures except Maze. Maze was not administered due to time constraints.

A random stratified sample from a single school district was selected based upon student Acadience Reading performance from the beginning-of-year benchmark assessment. Of the 318 students selected, 152 students returned consent forms and were retested. Data are not reported for kindergarten and sixth-grade measures, or for first-grade Retell, due to insufficient sample sizes. The final sample included 120 students.

Test-retest reliability is reported for Nonsense Word Fluency (NWF) in Table 5.9, Oral Reading Fluency (ORF) in Table 5.10, and for the Reading Composite Score in Table 5.11. Test-retest reliability coefficients appear to be conservative estimates in light of the alternate-form reliability coefficients presented previously. For NWF, reliability coefficients are sufficient for screening decisions. In general, for ORF Words Correct and the Reading Composite Score, reliability coefficients are sufficient for making important individual educational decisions.

Table 5.9 Test-Retest Reliability for First Grade Nonsense Word Fluency

NWF Scores	N	First Administration		Second Administration		Reliability	
		Mean	SD	Mean	SD	Single-Form	Estimated Three-Form
NWF Correct Letter Sounds	27	58.63	22.27	69.00	22.83	.76	.90
NWF Whole Words Read	27	12.63	10.58	17.11	11.54	.70	.88

Note. Based on Study C middle-of-year data. Data not available for kindergarten due to insufficient sample size. The estimated three-form test-retest reliability is based on the Spearman-Brown Prophecy Formula. All correlations are significant, $p < .001$.

Table 5.10 Test-Retest Reliability for Oral Reading Fluency (ORF)

ORF Scores by Grade	N	First Form		Second Form		Reliability
		Mean	SD	Mean	SD	Triad
ORF Words Correct						
First Grade	28	35.86	26.22	44.29	28.66	.95
Second Grade	21	102.38	27.74	113.76	28.37	.91
Third Grade	27	104.93	35.03	123.37	38.51	.93
Fourth Grade	21	121.14	38.49	140.14	37.09	.97
Fifth Grade	23	124.43	42.71	134.13	43.56	.97
ORF Accuracy						
First Grade	28	77%	22	83%	4	.84
Second Grade	21	97%	3	99%	1	.57**
Third Grade	27	97%	2	99%	2	.68
Fourth Grade	21	97%	3	99%	2	.91
Fifth Grade	23	96%	5	97%	8	.94
ORF Retell						
Second Grade	21	48.33	15.21	49.86	17.81	.27†
Third Grade	27	57.07	20.22	58.89	19.78	.69
Fourth Grade	21	57.57	22.11	52.90	15.18	.36†
Fifth Grade	22	52.32	19.15	60.27	15.75	.58**

Note. Based on Study C middle-of-year data. Data not available for first-grade ORF Retell and sixth-grade measures due to insufficient sample size. ORF passages are administered in triads, thus the test-retest reliability is reported as three-form. Unless marked, all correlations significant, $p < .001$; ** $p < .01$; † Not significant.

Table 5.11 Test-Retest Reliability for Reading Composite Score from Study C

Grade	N	First Form		Second Form		Reliability
		Mean	SD	Mean	SD	
First Grade	27	163.63	78.57	194.44	82.56	.94
Second Grade	21	298.86	60.79	321.67	64.48	.81

Note. Based on middle-of-year Study C data. Test-retest reliability for Reading Composite Score for third through sixth grade is unavailable, because information about Maze was not available. All correlations significant, $p < .001$.

Inter-Rater Reliability

Information about inter-rater reliability of all Acadience Reading measures was gathered from Study C during beginning-of-year benchmark administration. A total of 3,676 students from ten schools were eligible to participate. Of these, 264 students across all grades were randomly selected in five schools for shadow-scoring practices. All Acadience Reading measures were included in this portion of the study. In third through sixth grade, students were divided into two groups; one group had shadow-scoring for ORF and the other for Maze. Thus, for these grades, Reading Composite Scores are not available.

Inter-rater reliability coefficients for kindergarten, first, and second grade measures are presented in Table 5.12, for ORF in Table 5.13, for Maze in Table 5.14, and Reading Composite Score in Table 5.15. Mean scores across all grades are different by approximately 1 point or less (see Tables 5.12, 5.13, and 5.14). Correlations for most measures are above .90. The exception is ORF accuracy in third grade, which is .85. Examination of ORF Words Correct scores and errors revealed a 10-point error discrepancy for one student; all other scores were within 1 point. Inter-rater reliability is high for all measures, indicating that scoring directions were applied in a consistent manner across assessors in this study.

Table 5.12 Inter-Rater Reliability for Kindergarten, First and Second Grade Acadience Reading Measures

Acadience Reading Measures by Grade	N	First Rater		Second Rater		Inter-Rater Reliability
		Mean	SD	Mean	SD	
Kindergarten						
First Sound Fluency	25	12.36	11.98	11.56	12.17	.94
Letter Naming Fluency	25	20.52	14.31	20.12	14.50	.99
First Grade						
Letter Naming Fluency	25	48.52	19.79	48.68	19.90	.99
Phoneme Segmentation Fluency	25	38.76	17.16	37.20	16.29	.95
NWF Correct Letter Sounds	25	41.32	32.18	40.80	32.41	.99
NWF Whole Words Read	25	8.00	12.19	7.60	12.14	.99
Second Grade						
NWF Correct Letter Sounds	25	64.08	32.63	64.00	33.39	.90
NWF Whole Words Read	25	16.72	14.69	16.56	14.36	.99

Note. Based on Study C beginning-of-year data. The estimated three-form reliability based on the Spearman-Brown Prophecy Formula for all measures was above .98. All correlations were significant, *p* < .001.

Table 5.13 Inter-Rater Reliability for Oral Reading Fluency (ORF)

| ORF Scores by Grade | N | First Rater | | Second Rater | | Triad Reliability |
		Mean	SD	Mean	SD	
ORF Words Correct						
Second Grade	25	58.72	28.67	58.32	29.37	.99
Third Grade	25	95.24	37.97	94.68	37.79	.99
Fourth Grade	24	98.71	32.44	98.38	31.92	.99
Fifth Grade	28	110.04	40.24	110.25	40.51	.99
Sixth Grade	20	140.80	32.30	141.25	32.34	.99
ORF Accuracy						
Second Grade	25	90%	10	90%	10	.99
Third Grade	25	95%	6	95%	6	.85
Fourth Grade	24	96%	4	96%	4	.93
Fifth Grade	28	96%	4	95%	4	.95
Sixth Grade	20	98%	2	98%	2	.91
ORF Retell						
Second Grade	20	26.60	12.65	26.75	13.35	.98
Third Grade	24	36.96	14.95	37.29	15.80	.92
Fourth Grade	24	39.17	18.13	39.75	19.25	.98
Fifth Grade	28	35.79	16.96	35.07	18.26	.96
Sixth Grade	20	41.10	19.60	42.50	19.31	.99

Note. Based on Study C beginning-of-year data. ORF passages are administered in triads, thus the inter-rater reliability is reported as three-form. All correlations were significant, $p < .001$.

Table 5.14 Inter-Rater Reliability for Maze

| Grade | N | First Rater | | Second Rater | | Single-Form Reliability |
		Mean	SD	Mean	SD	
Third Grade	25	10.60	6.64	10.56	6.60	.99
Fourth Grade	25	15.92	6.20	15.96	6.33	.98
Fifth Grade	26	20.81	9.87	21.23	9.95	.99
Sixth Grade	20	22.55	8.61	22.40	8.75	.99

Note. Based on Study C beginning-of-year data. The estimated three-form reliability of Maze based on the Spearman-Brown Prophecy Formula was above .99. All correlations were significant, $p < .001$.

Table 5.15 Inter-Rater Reliability for Reading Composite Score

| Grade | N | First Rater | | Second Rater | | Reliability |
		Mean	SD	Mean	SD	
Kindergarten	25	32.88	21.47	31.68	22.25	.97
First Grade	25	128.60	55.93	126.68	55.37	.99
Second Grade	25	169.32	80.13	150.32	99.69	.98

Note. Based on beginning-of-year Study C data. Reliability for third through sixth grade is unavailable because students in this portion of the study received only ORF or Maze, and not both. All correlations significant, $p < .001$.

Types of Measurement Error

Measurement error is unavoidable, and there is always uncertainty in the student's score. A student's obtained score represents the student's true score and error. Error refers to unintended factors that contribute to changes in scores. The error is essentially random but can be inflated by different sources. There are many ways to reduce error to get the best estimate of the student's true score. In this section, the different types of measurement error are addressed, and the steps taken to reduce these types of error are discussed.

There are five factors that can affect test reliability: test length, testing interval, range of student ability in the sample, testing situation, and guessing (Salvia, Ysseldyke, & Bolt, 2007).

Test Length. Most Acadience Reading measures are 1-minute, timed assessments. Generally, students do not complete the form or passage within the allotted time. Ceiling effects are usually not a concern with Acadience Reading assessments, but floor effects can be observed in the earlier grades.

Testing Interval. Generally, the closer together the administrations, the higher the reliability. The inter-rater reliability estimates provided are not affected by the testing interval, because shadow-scoring was conducted during the original administration of the measure. Test-retest and alternate-form reliability testing were conducted approximately two weeks following middle-of-year benchmark assessment, the preferred amount of time between administrations.

Range of Student Ability in the Sample. When too much or too little variability exists in the sample to provide information on a range of student abilities, the resulting reliability estimates can be inaccurate. The sample for which Acadience Reading reliability is estimated was drawn from a fairly high-performing population of students (from Study C), so to obtain appropriate variability in student skills, the sample was selected using stratification techniques based on student performance.

Guessing. When a student guesses, even if the guesses are correct, the responses introduce random error into the score. To minimize random guessing, students should be encouraged to do their best and the assessor should monitor the student's effort and level of engagement.

Testing Situation. The student may react to the test (e.g., become frustrated, bored, or lose his/her place). The environment may not be suitable to the student (e.g., the furniture might be uncomfortable or the room might be cold). These circumstances may introduce an indeterminate amount of error and could lower the reliability of the test. In each study, care was taken to ensure that students were comfortable in their testing environment and that a rapport was developed between the student and the assessor.

Standard Error of Measurement

The standard error of measurement (SEM) is an index of measurement error. The SEM is the standard deviation of the errors attributable to sampling. The SEMs presented in this chapter are attributed to the Study C sample. They provide information about the confidence with which the score can be interpreted; with a small SEM, there is greater confidence that the student's current score reflects the student's true performance and skill level.

The single-form SEMs shown in Table 5.16 can be used to provide a confidence interval for a particular test score. To compute a confidence interval, a multiplier of 1 is used for a 68% confidence interval, a multiplier of 1.96 is used for a 95% confidence interval, and a multiplier of 2.58 is used for a 99% confidence interval. For example, the SEM of FSF is 5.13. If a student has a score of 35 on FSF, then there is 68% confidence that the student's true score lies within the range of 30 to 40, 95% confidence that the true score lies within the range

of 25 to 45 (1.96 x 5), and 99% confidence that the true score lies within the range of 22 to 48 (2.58 x 5). SEMs are low in second through fifth grade and are variable in kindergarten and first grade. Estimated three-form SEMs are reported in Table 5.17. These results support a high degree of test reliability associated with repeated assessments and corresponding increases in confidence for the resulting decisions.

Table 5.16 Estimated Standard Error of Measurement of Acadience Reading Measures

Acadience Reading Measure	Grade						
	K	1	2	3	4	5	6
First Sound Fluency	5.13	--	--	--	--	--	--
Letter Naming Fluency	6.11	--	--	--	--	--	--
Phoneme Segmentation Fluency	13.88	9.39	--	--	--	--	--
NWF Correct Letter Sounds	7.84	12.59	--	--	--	--	--
NWF Whole Words Read	1.15	4.27	--	--	--	--	--
ORF Single-Passage Words Correct	--	10.33	11.29	11.12	10.50	10.39	10.96
ORF Triad: Words Correct	--	5.56	8.00	7.00	8.53	7.66	7.00
ORF Triad: Accuracy	--	0.06	0.04	0.03	0.02	0.02	0.01
ORF Triad: Retell	--	6.42	9.81	8.62	8.68	15.32	10.17
Maze Adjusted Score	--	--	--	3.91	4.00	4.68	2.95
Reading Composite Score	28.46	22.35	28.23	20.69	25.17	31.57	15.19

Note. The ORF Single-Passage Words Correct SEM calculated from Study D data. First-grade Phoneme Segmentation Fluency and sixth-grade ORF Triad SEM calculated from Study E.

Table 5.17 Estimated Three-Form Standard Error of Measurement of Acadience Reading Measures Based on Alternate-Form Reliability Coefficients

Acadience Reading Measure	Grade						
	K	1	2	3	4	5	6
First Sound Fluency	3.15	--	--	--	--	--	--
Letter Naming Fluency	3.71	--	--	--	--	--	--
Phoneme Segmentation Fluency	10.12	6.51	--	--	--	--	--
NWF Correct Letter Sounds	5.04	7.67	--	--	--	--	--
NWF Whole Words Read	0.68	2.55	--	--	--	--	--
Maze Adjusted Score	--	--	--	2.42	2.54	3.08	1.77

Note. Based on middle-of-year Study C data, and beginning-of-year Study E data. Estimated three-form *SEMs* are calculated using the estimated three-form alternate-form reliability. The estimated three-form alternate-form reliability is calculated using the Spearman-Brown Prophecy Formula based on the single-form alternate-form reliability.

Summary

The overall reliability of Acadience Reading is summarized in Table 5.18. The coefficients reported in this table are the same as those reported in previous sections in this chapter; they are summarized here to provide an efficient quick reference for Acadience Reading users. Alternate-form reliability reported is the median reliability (where available) from Studies A, B, C, and D.

Reliability coefficients are consistently high across all three forms of reliability. The magnitude of the coefficients suggests that Acadience Reading possesses little test error and that users can have confidence in test results. With repeated assessment across multiple forms, reliability increases substantially, as noted where the estimated three-form reliability is reported.

Table 5.18 Summary Table of Reliability Estimates for Acadience Reading Measures

| Acadience Reading Measure | Type of Reliability | | | | | |
| | Alternate Form | | Test-Retest | | Inter-Rater | |
	Single-Form	Three-Form	Single-Form	Three-Form	Single-Form	Three-Form
Kindergarten						
First Sound Fluency	.82	.93	--	--	.94	.98
Letter Naming Fluency	.86	.95	--	--	.99	1.00
Phoneme Segmentation Fluency	.44	.70	--	--	.96	.99
NWF Correct Letter Sounds	.71	.88	--	--	.99	1.00
NWF Whole Words Read	.92	.97	--	--	.99	1.00
Reading Composite Score	.66	--	--	--	.97	--
First Grade						
Letter Naming Fluency	--	--	--	--	.99	1.00
Phoneme Segmentation Fluency	.54	.78	--	--	.95	.98
NWF Correct Letter Sounds	.85	.94	.76	.90	.99	1.00
NWF Whole Words Read	.90	.96	.70	.88	.99	1.00
ORF Words Correct	.95	.98[a]	--	.95[a]	--	--
ORF Accuracy	--	.88[a]	--	.84[a]	--	--
Reading Composite Score	.95	--	.94	--	.99	--
Second Grade						
ORF Words Correct	.89	.96[a]	--	.91[a]	--	.99[a]
ORF Accuracy	--	.83[a]	--	.57**,[a]	--	.99[a]
Retell	--	.68**,[a]	--	.27[a]	--	.98[a]
Reading Composite Score	.92	--	.81	--	.98	--
Third Grade						
ORF Words Correct	.89	.97[a]	--	.93[a]	--	.99[a]
ORF Accuracy	--	.80[a]	--	.68[a]	--	.85[a]
Retell	--	.81[a]	--	.69[a]	--	.92[a]
Maze	.81	.93	--	--	.99	1.00
Reading Composite Score	.97	--	--	--	--	--
Fourth Grade						
ORF Words Correct	.88	.95[a]	--	.97[a]	--	.99[a]
ORF Accuracy	--	.85[a]	--	.91[a]	--	.93[a]
Retell	--	.80[a]	--	--	--	.98[a]
Maze	.74	.90	--	--	.98	.99
Reading Composite Score	.95	--	--	--	--	--

Table 5.18 Summary Table of Reliability Estimates for Acadience Reading Measures, *(continued)*

Acadience Reading Measure	Alternate Form		Test-Retest		Inter-Rater	
	Single-Form	Three-Form	Single-Form	Three-Form	Single-Form	Three-Form
Fifth Grade						
ORF Words Correct	.92	.96[a]	--	.97[a]	--	.99[a]
ORF Accuracy	--	.76[a]	--	.94[a]	--	.95[a]
Retell	--	.65[a]	--	.58**, [a]	--	.96[a]
Maze	.66	.85	--	--	.99	1.00
Reading Composite Score	.91	--	--	--	--	--
Sixth Grade						
ORF Words Correct	.83	.94	--	--	--	.99[a]
ORF Accuracy	--	.48	--	--	--	.91[a]
Retell	--	.62	--	--	--	.99[a]
Maze	.79	.89	--	--	.99	1.00
Reading Composite Score	.91	--	--	--	--	--

Note. Inter-rater and test-retest reliability calculated from Study C. Alternate-form reliability reported is the median reliability from studies A, B, C, and D, for measures where multiple-reliability coefficients were available. ORF single-form reliability is based on Study D, and ORF three-form reliability is based on Study C. Inter-rater is based on beginning-of-year data. Alternate-form and test-retest based on middle-of-year data.

[a] Reliability coefficients are calculated from the median score of three benchmark passages and are thus reported as three-form or triad reliability. Three-form reliabilities that are not marked are estimated using the Spearman-Brown Prophecy Formula.

Unless marked, all correlations significant, $p < .001$; significance codes: ** $p < .01$

Chapter 6: **Validity**

In this chapter we discuss the validity evidence gathered on Acadience Reading. Validity data were obtained from four separate research studies. Types of validity included in this chapter are content, criterion-related, and discriminant validity. When describing validity correlation coefficients, the following descriptors from Hopkins (2002) are used:

Table 6.1 Validity Correlation Coefficient Descriptors

Validity Correlation Range	Descriptor
.70 and above	Strong
.50 – .69	Moderate-Strong
.30 – .49	Moderate
.10 – .29	Small
.09 or less	Very Small

Content Validity

According to Salvia, Ysseldyke, & Bolt (2007), "content validity is the extent to which a test's items actually represent the domain or universe to be measured" (p. 145). Elements of content validity typically evaluated include content appropriateness (i.e., do items measure the domain or construct), completeness (i.e., does test content assess the domain or construct completely), and how content is measured (e.g., selection-response items where a student may guess or production-response items requiring actual demonstration of the skill assessed). Importantly, content validity reflects how much the assessment is representative of and relevant to the target construct as it relates to the purpose for testing (Haynes, Richards, & Kubany, 1995). In fact, Messick (1993) suggests that content validity is a state, not a trait of scores obtained from an assessment instrument—content validity varies with the inferences that are to be drawn from the assessment data. As such, the evaluation of content validity must take into consideration the purpose for testing.

Acadience Reading measures were designed to be general outcome measures (GOMs). As such, Acadience Reading measures are indicators of overall performance in a particular skill domain, meaning, "they measure key skills that are representative of and related to important global outcomes, such as reading competence" (Kaminski & Cummings, 2007, p. 27). The Acadience Reading measures were designed specifically to be linked to foundational early literacy skills and sensitive to growth and change in response to instruction or intervention in those areas.

Acadience Reading measures serve as key indicators of foundational skills in beginning reading (National Reading Panel, 2000; National Research Council, 1998). Table 6.2 illustrates the linkage between the foundational early literacy skills and each Acadience Reading measure.

Table 6.2 Alignment of Acadience Reading Measures with Basic Early Literacy Skills

Basic Early Literacy Skills	Acadience Reading Measures
Phonemic Awareness	First Sound Fluency (FSF) Phoneme Segmentation Fluency (PSF)

Alphabetic Principle and Basic Phonics	Nonsense Word Fluency (NWF)[1]
	    –Correct Letter Sounds
	    –Whole Words Read

Advanced Phonics and Word Attack Skills	Oral Reading Fluency (ORF)[2] –Accuracy

Accurate and Fluent Reading of Connected Text	Oral Reading Fluency (ORF)[2]
	    –Correct Words Per Minute
	    –Accuracy

Reading Comprehension	Maze
	Oral Reading Fluency (ORF)[2]
	    –Correct Words Per Minute
	    –Retell Total/Quality of Response

Vocabulary and Language Skills	Word Use Fluency-Revised[3]

[1]Nonsense Word Fluency is an indicator of basic phonics skills, specifically a student's knowledge of the most common letter-sound correspondences and ability to apply that knowledge to decode simple vowel-consonant and consonant-vowel-consonant words.

[2]Oral Reading Fluency is a more advanced indicator of word reading decoding skills and the student's application of those skills to reading connected text.

[3]Word Use Fluency-Revised is available as an experimental measure. Email info@acadiencelearning.org for more information.

For additional information on the foundation for the Acadience Reading measures, please see Chapter 1 of this *Technical Manua*l as well as Good, Simmons, & Smith (1998); Kaminski (1992; pp. 23–32); Kaminski, Cummings, Powell-Smith & Good (2008); and Kaminski & Good (1996).

Content Validity for Individual Measures

The design specifications for Acadience Reading measures relate directly to their content validity. Each measure was designed according to specific criteria to maximize their utility and sensitivity. For information on design specifications for Acadience Reading measures, see Chapter 2.

Criterion-Related Validity

Criterion-related validity is the extent to which a person's performance on a criterion measure can be estimated from that person's performance on the assessment procedure being validated (Salvia, Ysseldyke, & Bolt, 2007). A test is valid if it accurately measures what it is supposed to measure. Evidence of validity is presented as a correlation between the assessment and the criterion. Concurrent validity estimates how well student performance on the assessment is related to student performance on the criterion when both are given at about the same time. Predictive validity estimates how well student performance on the assessment predicts student performance on the criterion at a later time.

Validity of the Acadience Reading measures was examined using a variety of criterion measures including the Group Reading Assessment and Diagnostic Evaluation (GRADE), the Standard 4th Grade Reading Passage used in the National Assessment of Education Progress (NAEP) 2002 Special Study of Oral Reading (Daane et al., 2005), and the Comprehensive Test of Phonological Processing (CTOPP; Wagner, Torgesen, & Rashotte,

1999), as well as comparisons to other Acadience Reading measures. The criterion measure varied depending upon which Acadience Reading measure was being examined. Evidence for the validity of Acadience Reading is first presented based on an external criterion measure, the GRADE Total Test composite score, followed by results for each Acadience Reading measure. Finally, evidence for the validity of the Reading Composite Score is presented.

Summary of Criterion-Related Validity of All Acadience Reading Measures with the Group Reading Assessment and Diagnostic Evaluation (GRADE)

The Group Reading Assessment and Diagnostic Evaluation (GRADE) was administered in the spring for Study C, concurrent with the end-of-year Acadience Reading benchmark assessment. The GRADE is an untimed, group-administered, norm-referenced reading achievement test appropriate for children in preschool through grade 12. The GRADE is comprised of 16 subtests within five components. Not all 16 subtests are used at each testing level. Various subtest scores are combined to form the Total Test composite score. The GRADE Total Test raw score was compared to all Acadience Reading measures given during the year, providing both predictive criterion-related validity correlations for beginning- and middle-of-year Acadience Reading measures and concurrent criterion-related validity data for end-of-year Acadience Reading measures. The GRADE Total Test score is comprised of scores across subtests of the GRADE that vary by grade level. In kindergarten, the GRADE Total Test score is comprised of measures that assess phonics and phonemic and phonological awareness. In first and second grade, GRADE Total Test includes word meaning, passage (or sentence) reading, and comprehension measures. In third grade, GRADE Total Test is comprised of measures assessing word reading, vocabulary, and comprehension. In fourth, fifth, and sixth grade, GRADE Total Test includes scores from measures of vocabulary and comprehension.

Correlation coefficients indicating the strength of the relation between the Acadience Reading measures and GRADE Total Test are reported in Table 6.3. Overall, the validity of all Acadience Reading measures is well supported by GRADE Total Test. The Reading Composite Score in kindergarten and first grade is moderately to strongly correlated with the GRADE Total Test. For second through sixth grade, predictive validity coefficients for the Reading Composite Score indicate moderate-strong to strong relations with the GRADE Total Test. When examining individual measures, predictive and concurrent validity coefficients are moderate to strong for second- through sixth-grade measures with the GRADE Total Test.

Table 6.3 Criterion-Related Validity for Acadience Reading Measures with GRADE Total Test

Acadience Reading Measure by Time of Year	Grade Level						
	K	1	2	3	4	5	6
Predictive Validity Coefficients							
Beginning of year							
First Sound Fluency	.52	--	--	--	--	--	--
Letter Naming Fluency	.39	.54	--	--	--	--	--
Phoneme Segmentation Fluency	--	.33	--	--	--	--	--
NWF Correct Letter Sounds	--	.43	.51	--	--	--	--
NWF Whole Words Read	--	.39	.51	--	--	--	--
ORF Words Correct	--	--	.69	.66	.77	.69	.64
ORF Accuracy	--	--	.75	.68	.62	.53	.55
Retell	-	--	.53	.48	.56	.61	.55
Maze Adjusted Score	--	--	--	.65	.67	.56	.60
Reading Composite Score	.50	.55	.75	.73	.80	.76	.71
Middle of year							
First Sound Fluency	.40	--	--	--	--	--	--
Letter Naming Fluency	.35	--	--	--	--	--	--
Phoneme Segmentation Fluency	.34	--	--	--	--	--	--
NWF Correct Letter Sounds	.47	.51	--	--	--	--	--
NWF Whole Words Read	.19*	.52	--	--	--	--	--
ORF Words Correct	--	.64	.76	.67	.77	.65	.59
ORF Accuracy	--	.80	.78	.71	.62	.49	.47
Retell	--	.55	.52	.56	.63	.63	.59
Maze Adjusted Score	--	--	--	.61	.61	.59	.56
Reading Composite Score	.48	.71	.80	.78	.80	.76	.68
Concurrent Validity Coefficients							
End of year							
Letter Naming Fluency	.35	--	--	--	--	--	--
Phoneme Segmentation Fluency	.24	--	--	--	--	--	--
NWF Correct Letter Sounds	.40	.56	--	--	--	--	--
NWF Whole Words Read	.35	.56	--	--	--	--	--
ORF Words Correct	--	.75	.73	.66	.74	.65	.61
ORF Accuracy	--	.73	.67	.59	.54	.49	.55
Retell	--	.40	.48	.53	.62	.65	.56
Maze Adjusted Score	--	--	--	.67	.68	.66	.64
Reading **Composite Score**	.37	.77	.75	.75	.80	.77	.73

Note. Based on Study C data. GRADE Total Test = Group Reading Assessment and Diagnostic Evaluation Total Test raw composite scores. Total sample size = 1,306. GRADE administered at end of year. Unless marked, all correlations significant, $p < .001$; * $p < .05$.

Summary of Predictive Validity of All Acadience Reading Measures with Later Reading Composite Scores

Correlation coefficients indicating the strength of the relation between the Acadience Reading measures and the Reading Composite Score at a later time are reported in Table 6.4. Overall, the predictive validity of all Acadience Reading measures is well supported by correlations with the Reading Composite Score at a later time. With the exception of PSF, the Acadience Reading measures in kindergarten and first grade are moderately to strongly correlated with the later Reading Composite Scores. For second through sixth grade, predictive validity coefficients of all measures with later Reading Composite Scores are moderate-strong to strong.

Table 6.4 Predictive Criterion-Related Validity for all Acadience Reading Measures with the Reading Composite Score

Acadience Reading Measure	Reading Composite Score by Grade and Time of Year													
	Middle of Year							End of Year						
	K	1	2	3	4	5	6	K	1	2	3	4	5	6
Predictive Validity Coefficients–Beginning of Year														
FSF	.57	--	--	--	--	--	--	.43	--	--	--	--	--	--
LNF	.60	.65	--	--	--	--	--	.49	.65	--	--	--	--	--
PSF	--	.25	--	--	--	--	--	--	.26	--	--	--	--	--
NWF–CLS	--	.82	.69	--	--	--	--	--	.71	.65	--	--	--	--
NWF–WWR	--	.79	.65	--	--	--	--	--	.66	.62	--	--	--	--
ORF Words Correct	--	--	.85	.88	.90	.89	.87	--	--	.81	.86	.86	.85	.86
ORF Accuracy	--	--	.75	.71	.72	.69	.66	--	--	.71	.70	.71	.66	.65
Retell	--	--	.63	.64	.62	.58	.61	--	--	.62	.64	.62	.58	.62
Maze	--	--	--	.79	.76	.74	.78	--	--	--	.74	.76	.69	.77
Predictive Validity Coefficients–Middle of Year														
FSF	--	--	--	--	--	--	--	.47	--	--	--	--	--	--
LNF	--	--	--	--	--	--	--	.60	--	--	--	--	--	--
PSF	--	--	--	--	--	--	--	.47	--	--	--	--	--	--
NWF–CLS	--	--	--	--	--	--	--	.65	.78	--	--	--	--	--
NWF–WWR	--	--	--	--	--	--	--	.52	.78	--	--	--	--	--
ORF Words Correct	--	--	--	--	--	--	--	--	.83	.87	.86	.87	.87	.87
ORF Accuracy	--	--	--	--	--	--	--	--	.81	.75	.69	.68	.64	.62
Retell	--	--	--	--	--	--	--	--	.67	.70	.65	.68	.67	.72
Maze	--	--	--	--	--	--	--	--	--	--	.72	.75	.77	.77

Note. Based on Study C data. Approximate pair-wise sample sizes: kindergarten ≈ 465; first grade ≈ 440; second grade ≈ 540; third grade ≈ 480; fourth grade ≈ 570; fifth grade ≈ 520; sixth grade ≈ 510. All correlations significant, $p < .001$.

Criterion-Related Validity Evidence for Each Acadience Reading Measure

First Sound Fluency

The validity of First Sound Fluency (FSF) is moderately supported. Predictive criterion-related validity correlations with external criterion measures GRADE and CTOPP are presented in Table 6.5. Concurrent and predictive validity coefficients with Phoneme Segmentation Fluency (PSF) and both Nonsense Word Fluency scores, Correct Letter Sounds (NWF–CLS) and Whole Words Read (NWF–WWR), are reported in Table 6.6. FSF predicts future literacy outcomes well. Coefficients fall in the moderate to moderate-strong range with PSF, NWF–CLS, and GRADE TT. Correlations with the CTOPP are small to moderate.

Table 6.5 Predictive Criterion-Related Validity for First Sound Fluency

| | GRADE[c] | Selected CTOPP Composite and Subtests[a] | | | |
Time of Year	Total Test	PA	SM	EL	BW
Beginning	.52	--	--	--	--
Middle	.40	.45	.29**	.49	.19[†]

Note. Sample size for GRADE = 166. Sample size for CTOPP = 81. GRADE Total Test = Group Reading Assessment and Diagnostic Evaluation Total Test raw score composite; CTOPP = Chronological Test of Phonological Processing age-standardized scores; PA = Phonemic Awareness composite; SM = Sound Matching; EL = Elision; and BW = Blending Words. Both GRADE and CTOPP administered at end of year. Unless marked, correlations significant, $p < .001$; ** $p < .01$; [†] Not significant.
[a] from Study A.
[c] from Study C.

Table 6.6 Concurrent and Predictive Criterion-Related Validity for First Sound Fluency Based on Acadience Reading Measures

| | Acadience Reading Criterion Measures by Time of Year | | | | | |
| | PSF | | NWF–CLS | | NWF–WWR | |
Time of Year	Middle	End	Middle	End	Middle	End
Beginning	.47	.29	.47	.41	.36	.35
Middle	.74[b], .56	.42	.45[b], ·46	.35	.21[b], .26	.25

Note. Unless marked, data gathered from Study C. Approximate pair-wise sample size ≈ 455. Correlations between measures administered at middle of year represent concurrent criterion-related validity; all other correlations presented in this table represent predictive criterion-related validity. All correlations are significant, $p < .001$.
[b] from Study B, sample size ≈ 90.

Letter Naming Fluency

The predictive validity of Letter Naming Fluency (LNF) is moderate-strong. Correlations with First Sound Fluency (FSF), Phoneme Segmentation Fluency (PSF), both Nonsense Word Fluency scores, Correct Letter Sounds (NWF–CLS) and Whole Words Read (NWF–WWR), and the external criterion measure GRADE Total Test are presented in Table 6.7. Coefficients are small to moderate-strong in kindergarten and moderate-strong in first grade. LNF predicts future reading outcomes well. Predictive validity coefficients with the GRADE Total Test fall in the moderate to moderate-strong range.

Table 6.7 Concurrent and Predictive Criterion-Related Validity for Letter Naming Fluency

Grade by Time of Year	Acadience Reading Criterion Measures by Time of Year							GRADE Total Test
	FSF	PSF		NWF–CLS		NWF–WWR		
	Middle	Middle	End	Middle	End	Middle	End	
Kindergarten								
Beginning	.37	.34	.18	.56	.49	.39	.39	.39
Middle	--	.64[b], .44	.27	.66[b], .62	.49	.38[b], .41	.43	.35
First Grade								
Beginning	--	--	--	.57	.58	.55	.54	.54

Note. Unless marked, data gathered from Study C. Approximate pair-wise sample sizes for measures: kindergarten ≈ 460; first grade ≈ 445. Sample size with GRADE: kindergarten = 166; first grade = 193. GRADE TT = Group Reading Assessment and Diagnostic Evaluation Total Test raw composite scores. GRADE measures administered at end of year. Correlations between measures administered at middle of year represent concurrent criterion-related validity; all other correlations presented in this table represent predictive criterion-related validity. All correlations were significant, $p < .001$.
[b] Study B, sample size for kindergarten ≈ 90.

Phoneme Segmentation Fluency

The validity of Phoneme Segmentation Fluency (PSF) is moderately supported. Concurrent validity coefficients with both Nonsense Word Fluency scores, Correct Letter Sounds (NWF–CLS) and Whole Words Read (NWF–WWR), and GRADE Total Test are presented in Table 6.8. Predictive validity coefficients with NWF–CLS, NWF–WWR, Oral Reading Fluency (ORF) Words Correct and Accuracy, and the external criterion measure GRADE Total Test are presented in Table 6.9. Discussion focuses on kindergarten correlations, because PSF in first grade is used primarily to identify students who have not reached the end-of-year kindergarten goal. Additionally, GRADE subtests in first grade are based on vocabulary and comprehension measures, thus we would not expect PSF to be a strong indicator for those outcomes. Concurrent and predictive validity coefficients with NWF–CLS and NWF–WWR, ORF Words Correct and Accuracy, and GRADE Total Test are in the small-to-moderate range in kindergarten. The highest predictive and concurrent validity coefficients are found with NWF–CLS.

Table 6.8 Concurrent Criterion-Related Validity for Phoneme Segmentation Fluency

Grade by Time of Year	Acadience Reading Criterion Measures		GRADE Total Test
	NWF–CLS	NWF–WWR	
Kindergarten			
Middle	.51[b], .45	.26[b], .24	--
End	.43	.35	.24**
First Grade			
Beginning	.30	.18	--

Note. Unless noted, all data is from Study C. Approximate pair-wise sample sizes for Acadience Reading measures: kindergarten ≈ 473; first grade = 461. Approximate sample sizes for GRADE: kindergarten ≈ 170. GRADE Total Test = Group Reading Assessment and Diagnostic Evaluation Total Test raw composite score.
[b] Study B, sample sizes: kindergarten middle of year ≈ 91, first grade beginning of year = 71.
Unless marked, all correlations are significant, $p < .001$; ** $p < .01$.

Table 6.9 Concurrent and Predictive Criterion-Related Validity for Phoneme Segmentation Fluency

Grade by Time of Year	Acadience Reading Criterion Measures by Time of Year								GRADE Total Test
	NWF–CLS		NWF–WWR		ORF Words Correct		ORF Accuracy		
	Middle	End	Middle	End	Middle	End	Middle	End	
Kindergarten									
Middle	.24**	.37	--	.31	--	--	--	--	.34
First Grade									
Beginning	.24	.24	.19	.20	.24	.21	.29	.30	.33

Note. Based on Study C data. Approximate pair-wise sample sizes with Acadience Reading measures: kindergarten = 454; first grade ≈ 440. Approximate sample sizes for GRADE: kindergarten ≈ 170; first grade = 193. GRADE Total Test = Group Reading Assessment and Diagnostic Evaluation Total Test raw composite score. Correlations between measures administered at middle of year represent concurrent criterion-related validity; all other correlations presented in this table represent predictive criterion-related validity. Unless marked, all correlations are significant, $p < .001$; ** $p < .01$.

Nonsense Word Fluency

The validity of Nonsense Word Fluency (NWF) is moderate to strong with respect to ORF, and predicts middle- and end-of-year outcomes very well. Validity coefficients are given for both NWF scores, Correct Letter Sounds (NWF–CLS) and Whole Words Read (NWF–WWR). Predictive validity coefficients with the external criterion GRADE Total Test are presented in Table 6.10. Concurrent validity coefficients with ORF Words Correct, Accuracy, and Retell are presented in Table 6.11. Predictive validity coefficients with these same measures are presented in Table 6.12. Correlations with Retell reflect relationships between measures where students scored higher than 40 on ORF Words Correct, as per the standardized directions for administering ORF.

Concurrent and predictive validity coefficients fall in the moderate to strong range, with slightly higher correlations with ORF Words Correct than ORF Accuracy or Retell. Correlations with GRADE Total Test are moderate to moderate-strong.

Table 6.10 Predictive Criterion-Related Validity for Nonsense Word Fluency with GRADE Total Test

NWF Score	Grade		
	K	1	2
Predictive Validity Coefficients–Beginning of Year			
NWF–CLS	--	.43	.51
NWF–WWR	--	.39	.51
Predictive Validity Coefficients–Middle of Year			
NWF–CLS	.47	.51	--
NWF–WWR	.19	.52	--
Concurrent Validity Coefficients–End of Year			
NWF–CLS	.40	.56	--
NWF–WWR	.35	.56	--

Note. Based on Study C data. Approximate pair-wise sample sizes: kindergarten ≈ 170; first grade ≈ 195; second grade ≈ 214. GRADE Total Test = Group Reading Assessment and Diagnostic Evaluation Total Test raw composite scores. GRADE administered at end of year. All correlations are significant, $p < .001$.

Table 6.11 Concurrent Criterion-Related Validity for Nonsense Word Fluency

Grade by Time of Year	Acadience Reading Criterion Measures		
	ORF Words Correct	ORF Accuracy	Retell
NWF Correct Letter Sounds			
First Grade			
Middle	.64[b], .80	.64[b], .64	.65
End	.77	.57	.28
Second Grade			
Beginning	.79	.56	.50
NWF Whole Words Read			
First Grade			
Middle	.49[b], .78	.60[b], .63	.58
End	.74	.58	.25
Second Grade			
Beginning	.74	.53	.44

Note. Unless marked, based on Study C data. Approximate pair-wise sample sizes: first grade ≈ 458 (ORF) and 300 (Retell); second grade beginning of year ≈ 520.
[b]Study B sample sizes: first grade, middle of year ≈ 68.
All correlations are significant, $p < .001$.

Table 6.12 Predictive Criterion-Related Validity for Nonsense Word Fluency

Grade by Time of Year	Acadience Reading Criterion Measures					
	ORF Words Correct		ORF Accuracy		Retell	
	Middle	End	Middle	End	Middle	End
NWF Correct Letter Sounds						
First Grade						
Middle	.80	.70	.59	.45	.63	.35
End	--	.74	--	.51	--	.34
Second Grade						
Beginning	.73	.47	.52	.44	.47	.42
NWF Whole Words Read						
First Grade						
Middle	.77	.66	.53	.39	.56	.27
End	--	.72	--	.50	--	.30
Second Grade						
Beginning	.69	.43	.48	.41	.42	.39

Note. Based on Study C data. Approximate pair-wise sample sizes: first grade ≈ 450 (ORF) and 261 (Retell); second grade ≈ 530. All correlations are significant, $p < .001$.

Oral Reading Fluency

The criterion-related validity of Oral Reading Fluency is well supported. Predictive validity coefficients with the external criterion GRADE Total Test are presented in Table 6.13. Concurrent validity coefficients with Retell and Maze, and the Standard 4th Grade Reading Passage used in the NAEP 2002 Special Study of Oral Reading (Daane et al., 2005), are presented in Table 6.14. Predictive validity coefficients with these same measures are presented in Table 6.15.

Validity coefficients fall in the moderate to strong range with other Acadience Reading scores, GRADE Total Test, and the Standard 4th Grade Reading Passage used in the NAEP 2002 Special Study of Oral Reading. In addition, ORF Words Correct predicts future reading outcomes strongly and consistently across grades.

Student accuracy scores increase across grades, with first- through third-grade students averaging 90–96% accuracy, and fourth- through sixth-grade students averaging 97–98% accuracy. Correlations between accuracy and GRADE Total Test are generally strong in the early grades and decrease to moderate-strong or moderate as grade level and accuracy rate increase.

Table 6.13 Criterion-Related Validity for Oral Reading Fluency with GRADE Total Test

ORF Score	GRADE Total Test by Grade					
	1	2	3	4	5	6
Predictive Validity Coefficients–Beginning of Year						
ORF Words Correct	--	.69	.66	.77	.69	.64
ORF Accuracy	--	.75	.68	.62	.53	.55
Retell	--	.53	.48	.56	.61	.55
Predictive Validity Coefficients–Middle of Year						
ORF Words Correct	.64	.76	.67	.77	.65	.59
ORF Accuracy	.80	.78	.71	.62	.49	.47
Retell	.55	.52	.56	.63	.63	.59
Concurrent Validity Coefficients–End of Year						
ORF Words Correct	.75	.73	.66	.74	.65	.61
ORF Accuracy	.73	.67	.59	.54	.49	.55
Retell	.40	.48	.53	.62	.65	.56

Note. Based on Study C data. Approximate pair-wise sample sizes: first grade ≈ 196 (125 with Retell); second grade ≈ 215; third grade ≈ 190; fourth grade ≈ 190; fifth grade ≈ 194; sixth grade ≈ 103. GRADE Total Test = Group Reading Assessment and Diagnostic Evaluation Total Test raw composite score. All correlations are significant, *p* < .001.

Table 6.14 Concurrent Criterion-Related Validity for Oral Reading Fluency with Select Criterion Measures

Grade	Acadience Reading Criterion Measures[c]						NAEP Oral Reading Study[d]
	Retell			Maze			4th Grade Passage
	Beginning	Middle	End	Beginning	Middle	End	End
ORF Words Correct							
First	--	.76	.44	--	--	--	.97
Second	.61	.64	.61	--	--	--	.91
Third	.61	.56	.52	.73	.70	.78	.96
Fourth	.57	.58	.52	.78	.78	.78	.89
Fifth	.51	.60	.45	.74	.78	.77	.96
Sixth	.49	.57	.53	.77	.76	.78	.83
ORF Accuracy							
First	--	.56	.29	--	--	--	--
Second	.47	.47	.33	--	--	--	--
Third	.40	.38	.33	.51	.45	.55	--
Fourth	.38	.39	.28	.56	.49	.49	--
Fifth	.38	.40	.33	.52	.53	.53	--
Sixth	.35	.40	.33	.54	.55	.53	--

Note. [c]from Study C. Approximate pair-wise sample sizes: first grade ≈ 450; second grade ≈ 550; third grade ≈ 500; fourth grade ≈ 580; fifth grade ≈ 525; sixth grade ≈ 530.
[d]from Study D. Approximate sample sizes ≈ 23 for all grades.
All correlations are significant, $p < .001$.

Table 6.15 Predictive Criterion-Related Validity for Oral Reading Fluency with Acadience Reading Retell and Maze

ORF Score by Time of Year	Acadience Reading Measure by Grade									
	Retell						Maze Adjusted Score			
	1	2	3	4	5	6	3	4	5	6
Middle of Year										
Beginning of year										
ORF Words Correct	--	.57	.56	.58	.60	.57	.70	.78	.78	.76
ORF Accuracy	--	.43	.37	.35	.36	.32	.51	.52	.49	.52
Retell	--	--	--	--	--	--	.51	.47	.33	.40
End of Year										
Beginning of year										
ORF Words Correct	--	.53	.50	.52	.44	.48	.78	.77	.78	.75
ORF Accuracy	--	.36	.33	.35	.30	.31	.58	.61	.53	.53
Retell	--	--	--	--	--	--	.51	.52	.49	.52
Middle of year										
ORF Words Correct	.39	.57	.47	.49	.44	.51	.78	.79	.79	.78
ORF Accuracy	.31	.41	.33	.32	.30	.30	.56	.58	.52	.51
Retell	--	--	--	--	--	--	.48	.45	.43	.46

Note. Based on Study C data. Approximate pair-wise sample sizes: first grade ≈ 361; second grade ≈ 520; third grade ≈ 475; fourth grade ≈ 575; fifth grade ≈ 525; sixth grade ≈ 515. All correlations are significant, *p* < .001.

Maze

The validity of Maze is moderate-strong. Evidence for the validity of Maze based on the external criterion GRADE Total Test is presented in Table 6.16. Predictive and concurrent validity coefficients for Maze adjusted score with GRADE Total Test fall in the moderate-strong range, suggesting that Maze measures reading comprehension well.

Table 6.16 Criterion-Related Validity for Maze with GRADE Total Test

Time of Year	GRADE Total Test by Grade			
	3	4	5	6
Predictive Validity Coefficients				
Beginning	.67	.68	.61	.61
Middle	.65	.63	.65	.58
Concurrent Validity Coefficients				
End	.67	.68	.66	.64

Note. Based on Study C data. Approximate pair-wise sample sizes: third grade ≈ 184; fourth grade ≈ 184; fifth grade ≈ 194; sixth grade ≈ 103. Reported score for Maze is the Maze Adjusted Score. GRADE Total Test ≈ Group Reading Assessment and Diagnostic Evaluation Total Test raw composite scores. All correlations are significant, *p* < .001.

Reading Composite Score

Overall, the validity of the Reading Composite Score is strong. Evidence for the predictive validity of the Reading Composite Score with the GRADE Total Test is presented in Table 6.17. Evidence to support the composite score as a predictor of later Reading Composite Scores is presented in Table 6.18. Predictive validity coefficients for first- through sixth-grade Reading Composite Scores are in the moderate-strong to strong range. Coefficients for kindergarten Reading Composite Scores are moderate to moderate-strong. As expected, the Reading Composite Score strongly predicts future Reading Composite Scores. In turn, the predictive validity coefficients for the Reading Composite Score are moderate-strong to strong with GRADE Total Test.

Table 6.17 Criterion-Related Validity for Reading Composite Score Based on GRADE Total Test

Reading Composite Score by Time of Year	Grade Level						
	K	1	2	3	4	5	6
Beginning (Predictive)	.50	.55	.75	.73	.80	.76	.71
Middle (Predictive)	.48	.71	.80	.78	.80	.76	.68
End (Concurrent)	.40	.77	.75	.75	.80	.77	.73

Note. Based on Study C data. Sample size = 1,306. GRADE Total Test = Group Reading Assessment and Diagnostic Evaluation Total Test raw composite score. All correlations significant, $p < .001$.

Table 6.18 Predictive Criterion-Related Validity for Reading Composite Score with Later Reading Composite Scores

Reading Composite Score by Time of Year	Grade and Time of Year													
	Middle of Year							End of Year						
	K	1	2	3	4	5	6	K	1	2	3	4	5	6
Reading Composite Score Beginning of Year	.67	.79	.86	.91	.92	.89	.91	.52	.73	.81	.88	.88	.86	.90
Reading Composite Score Middle of Year	--	--	--	--	--	--	--	.70	.89	.89	.90	.90	.90	.91

Note. Based on Study C data. Approximate pair-wise sample sizes: kindergarten ≈ 465; first grade ≈ 440; second grade ≈ 540; third grade ≈ 480; fourth grade ≈ 570; fifth grade ≈ 520; sixth grade ≈ 510. All correlations significant, p < .001.

The Reading Composite Score explains more variance in reading outcomes than does ORF Words Correct alone. Across first through sixth grade, the median additional variance explained is 9%, ranging from 3% to 17%, generally with greater additional variance explained in the upper grades. In other words, although ORF Words Correct alone is very good, the Reading Composite Score is even better in meaningful and important ways.

Table 6.19 The Additional Variance Explained in the GRADE by the Reading Composite Score over Oral Reading Fluency Alone for First Through Sixth Grade

Grade by Time of Year	ORF Words Correct Predicting GRADE Total	Reading Composite Score Predicting GRADE Total	Additional Variance Explained by Reading Composite Score
First Grade			
Middle of Year	.64	.70	8%
End of Year	.75	.77	4%
Second Grade			
Beginning of Year	.69	.75	8%
Middle of Year	.76	.80	5%
End of Year	.73	.75	3%
Third Grade			
Beginning of Year	.66	.73	10%
Middle of Year	.67	.78	15%
End of Year	.66	.75	13%
Fourth Grade			
Beginning of Year	.76	.80	5%
Middle of Year	.76	.80	6%
End of Year	.75	.80	8%
Fifth Grade			
Beginning of Year	.69	.76	11%
Middle of Year	.64	.76	17%
End of Year	.66	.77	17%
Sixth Grade			
Beginning of Year	.64	.71	9%
Middle of Year	.59	.68	12%
End of Year	.61	.73	16%

Note. N = 3,816. GRADE Total refers to the Total Test raw score of the Group Reading Assessment and Diagnostic Evaluation, administered at the end of the school year as an external criterion to assess the validity of the Acadience Reading measures. For more information, see the Benchmark Goals Technical Report (Tech Report No. 11).

Discriminant Validity

Discriminant validity is a way of evaluating how well an assessment distinguishes between two groups of students at different skill levels (Shuttleworth, 2009). In other words, a test has discriminant validity if groups believed to be different are, in fact, different on the test. There should be significant differences in student performance on an assessment when these same students display significant differences in performance on a criterion—and the differences should be comparable.

Discriminant validity was examined for the Reading Composite Score relative to two levels of performance on the GRADE Total Test, below the 40th percentile on the GRADE's national norms and at or above the 40th percentile. Reading Composite Score descriptive statistics were calculated for each group and compared. Results are reported in Table 6.19. Differences in means were examined using a between-groups *t*-test for each grade; all yielded significant results. The *t*-statistics are reported to illustrate the magnitude of the differences in

means. The effect size of the Reading Composite Score based on Cohen's *d* is large across all grades. Overall, the Reading Composite Score adequately discriminates between these two distinct levels of reading skill at kindergarten through sixth grade levels.

Table 6.20 Discriminant Validity of the Reading Composite Score Based on the 40th Percentile Rank on GRADE Total Test Raw Score

| Grade by Time of Year | Reading Composite Score Descriptive Statistics by GRADE Total Test Percentile Rank | | | | | | Difference Statistics | |
| | Below 40th Percentile | | | Above 40th Percentile | | | | |
	N	Mean	SD	N	Mean	SD	t-stat	Cohen's d
Kindergarten								
Beginning	54	22.31	19.65	112	45.42	23.76	6.41	1.03
Middle	55	111.90	54.96	113	156.10	43.16	5.45	0.94
End	53	132.10	40.78	113	156.50	39.09	3.67	0.62
First Grade								
Beginning	54	105.00	29.68	139	145.90	39.54	7.33	1.11
Middle	55	96.51	48.69	140	220.50	88.12	11.17	1.58
End	54	115.10	65.34	139	228.00	59.81	11.26	1.85
Second Grade								
Beginning	61	111.20	61.32	153	219.80	60.88	11.74	1.79
Middle	61	136.70	83.90	158	282.10	60.87	13.26	2.15
End	60	194.00	82.49	157	309.90	67.27	10.19	1.62
Third Grade								
Beginning	49	168.80	96.65	135	327.60	85.88	10.43	1.80
Middle	51	221.50	94.03	136	390.30	83.82	11.56	1.96
End	51	279.80	99.64	136	442.00	79.58	11.00	1.91
Fourth Grade								
Beginning	64	200.00	110.10	119	360.60	82.62	10.68	1.73
Middle	65	250.20	102.00	120	400.90	73.24	11.06	1.79
End	66	316.30	106.30	120	467.60	76.42	10.70	1.73
Fifth Grade								
Beginning	93	311.70	95.38	101	454.30	77.61	11.41	1.66
Middle	92	346.70	82.12	102	477.50	73.50	11.67	1.69
End	92	377.80	92.00	101	527.80	80.28	12.06	1.75
Sixth Grade								
Beginning	19	292.30	98.61	84	442.30	77.95	6.70	1.85
Middle	17	330.90	112.70	85	483.80	93.18	5.61	1.60
End	19	334.90	101.40	86	502.40	84.85	7.11	1.92

Note. Based on Study C data. All *t*-tests were performed under both equal and unequal variance assumptions, both of which yielded highly significant results; the reported *t*-statistic is the average between the two tests under different assumptions. A pooled standard deviation was calculated for Cohen's *d*.

Appendix A: **Acadience Reading K–6 Benchmark Goals and Composite Score**

Acadience Reading provides two types of scores at each benchmark assessment period: a) a raw score for each individual measure and b) a composite score (the Reading Composite Score). Each of the scores is interpreted relative to benchmark goals and cut points for risk to determine if a student's score is at or above the benchmark, below the benchmark, or below the cut point for risk (well below the benchmark).

Benchmark Goals and Cut Points for Risk

Acadience Reading *benchmark goals* are empirically derived, criterion-referenced target scores that represent adequate reading skill for a particular grade and time of year. Benchmark goals and cut points for risk are provided for the Reading Composite Score as well as for individual Acadience Reading measures.

A *benchmark goal* indicates a level of skill at which students are likely to achieve the next Acadience Reading benchmark goal or reading outcome. Thus, for students who achieve a benchmark goal, the odds are in their favor of achieving later reading outcomes if they receive effective core reading instruction.

Conversely, the *cut points for risk* indicate a level of skill below which students are unlikely to achieve subsequent reading goals without receiving additional, targeted instructional support. For students who have scores below the cut point for risk, the probability of achieving later reading goals is low unless intensive support is provided.

The Acadience Reading benchmark goals and cut points for risk provide three primary benchmark status levels that describe students' performance: a) At or Above Benchmark, b) Below Benchmark, and c) Well Below Benchmark. These levels are based on the overall likelihood of achieving specified goals on subsequent Acadience Reading assessments or external measures of reading achievement.

> **At or Above Benchmark.** For students who score at or above the benchmark goal, the overall likelihood of achieving subsequent reading goals is approximately 80% to 90%. These students are likely to need effective core instruction to meet subsequent early literacy and/or reading goals. Within this range, the likelihood of achieving subsequent goals is lower for students whose scores are right at the benchmark goal and increases as scores increase above the benchmark (see Table 1).

> To assist in setting ambitious goals for students, the At or Above Benchmark level is subdivided into *At Benchmark* and *Above Benchmark* levels.

>> **At Benchmark.** In the At Benchmark range, the overall likelihood of achieving subsequent early literacy or reading goals is 70% to 85%. Some of these students, especially those with scores near the benchmark, may require monitoring and/or strategic support on specific component skills.

>> **Above Benchmark.** In the Above Benchmark range, the overall likelihood of achieving subsequent early literacy and/or reading goals is 90% to 99%. While all students with scores in this range will likely benefit from core support, some students with scores in this range may benefit from instruction on more advanced skills.

Below Benchmark. Between the benchmark goal and cut point for risk is a range of scores where students' future performance is more difficult to predict. For students with scores in this range, the overall likelihood of achieving subsequent early literacy/reading goals is approximately 40% to 60%. These students are likely to need strategic support to ensure their achievement of future goals. Strategic support generally consists of carefully targeted supplemental support in specific skill areas in which students are having difficulty. To ensure that the greatest number of students achieve later reading success, it is best for students with scores in this range to be monitored regularly to ensure that they are making adequate progress and to receive increased or modified support if necessary to achieve subsequent reading goals.

Well Below Benchmark. For students who score below the cut point for risk, the overall likelihood of achieving subsequent early literacy/reading goals is low, approximately 10% to 20%. These students are identified as likely to need intensive support. Intensive support refers to interventions that incorporate something more or something different from the core curriculum or supplemental support.

Intensive support might entail:

- delivering instruction in a smaller group or individually,
- providing more instructional time or more practice,
- presenting smaller skill steps in the instructional hierarchy,
- providing more explicit modeling and instruction, and/or
- providing greater scaffolding and practice.

Because students who need intensive support are likely to have individual needs, we recommend that their progress be monitored frequently and their intervention modified dynamically to ensure adequate progress.

Table 1 summarizes the design specifications for achieving later reading outcomes and provides descriptions for the likely need for support for each of the benchmark status levels. It is important to note that while there is an overall likelihood for each benchmark status level, within each level the likelihood of achieving later reading outcomes increases as students' scores increase. This is illustrated in the first column of Table 1.

Benchmark Goals Study

The Acadience Reading benchmark goals, cut points for risk, and Composite Score were developed based upon data collected in a study conducted during the 2009–2010 school year. The benchmark goals are based on research that examined the predictive probability of a score on a measure at a particular point in time, compared to later Acadience Reading measures and external measures of reading proficiency and achievement. The external criterion measure of reading proficiency was the Group Reading and Diagnostic Evaluation (GRADE; Williams, 2001). The 40th percentile on the GRADE assessment was used as an indicator that the students had adequate early reading and/or reading skills for their grade. Data for the study were collected in thirteen elementary and middle schools in five states. Data collection included administering the Acadience Reading measures to participating students in grades K–6 in addition to the GRADE. Participants in the study were 3,816 students across grades K–6 from general education classrooms who were receiving English language reading instruction, including students with disabilities and students who were English language learners, provided they had the response capabilities to participate. The study included both students who were struggling in reading and those who were typically achieving. A subset of the total sample participated in the GRADE assessment (*n* = 1,306 across grades K–6). Additional information about the study is included in the *Acadience Reading Technical Manual*, available from https://acadiencelearning.org/.

Table 1. Likelihood of Meeting Later Reading Goals and Acadience Reading Benchmark Status

Likelihood of Meeting Later Reading Goals	Benchmark Status	Benchmark Status Including Above Benchmark	What It Means
>99% 95% 90% 80% 70% 60%	**At or Above Benchmark** *overall likelihood of achieving subsequent early literacy goals: 80% to 90%*	**Above Benchmark** *overall likelihood of achieving subsequent early literacy goals: 90% to 99%*	For students with scores in this range, the odds of achieving subsequent early literacy/reading goals are very good. These students likely need effective core instruction to meet subsequent early literacy/reading goals. Some students may benefit from instruction on more advanced skills.
		At Benchmark *overall likelihood of achieving subsequent early literacy goals: 70% to 85%*	For students with scores in this range, the odds are in favor of achieving subsequent early literacy/reading goals. The higher above the benchmark goal, the better the odds. These students likely need effective core instruction to meet subsequent early literacy/reading goals. Some students may require monitoring and strategic support on specific component skills as needed.
55% 50% 45% 40%	**Below Benchmark** *overall likelihood of achieving subsequent early literacy goals: 40% to 60%*	**Below Benchmark** *overall likelihood of achieving subsequent early literacy goals: 40% to 60%*	For students with scores in this range, the overall odds of achieving subsequent early literacy/reading goals are approximately even, and hard to predict. Within this range, the closer students' scores are to the benchmark goal, the better the odds; the closer students' scores are to the cut point, the lower the odds. These students likely need core instruction coupled with strategic support, targeted to their individual needs, to meet subsequent early literacy/reading goals. For some students whose scores are close to the benchmark goal, effective core instruction may be sufficient; students whose scores are close to the cut point may require more intensive support.
30% 20% 10% <5%	**Well Below Benchmark** *overall likelihood of achieving subsequent early literacy goals: 10% to 20%*	**Well Below Benchmark** *overall likelihood of achieving subsequent early literacy goals: 10% to 20%*	For students with scores in this range, the overall odds of achieving subsequent early literacy/reading goals are low. These students likely need intensive support in addition to effective core instruction. These students may also need support on prerequisite skills (i.e., below grade level) depending upon the grade level and how far below the benchmark their skills are.

The addition of the Above Benchmark status level has not changed the benchmark goals. A benchmark goal is still the point at which the odds are in the student's favor of meeting later reading goals (approximately 60% likelihood or higher). The higher above the benchmark goal the student scores, the better the odds. For students who are already at benchmark, the Above Benchmark status level also provides a higher goal to aim for.

"Overall likelihood" refers to the approximate percentage of students within the category who achieve later goals, although the exact percentage varies by grade, year, and measure (see *Acadience Reading Benchmark Goals and Composite Score Document*)..

Instructional decisions should be made based on students' patterns of performance across all measures, in addition to other available information on student skills, such as diagnostic assessment or in-class work.

Reading Composite Score

The Reading Composite Score is a combination of multiple Acadience Reading scores and provides the best overall estimate of students' early literacy skills and/or reading proficiency. Most data management services will calculate the Reading Composite Score for you, provided that all required measures necessary for calculating it have been administered. To calculate the Reading Composite Score yourself, see the *Reading Composite Score Worksheets* at the end of this document.

Benchmark goals and cut points for risk for the Reading Composite Score are based on the same logic and procedures as the benchmark goals for the individual Acadience Reading measures. However, because the Reading Composite Score provides the best overall estimate of a student's skills, it should generally be interpreted first. If a student earns a Reading Composite Score that is at or above the benchmark goal, the odds are in the student's favor of reaching later important reading outcomes. Some students who score At or Above Benchmark on the Reading Composite Score may still need additional support in one of the basic early literacy skills, as indicated by a Below Benchmark score on an individual Acadience Reading measure (FSF, PSF, NWF, ORF, or Maze). This potential need for additional support is especially true for a student whose Reading Composite Score is close to the benchmark goal.

The Acadience Reading measures that are used to calculate the Reading Composite Score vary by grade and time of year. As such, the Reading Composite Score is not comparable across different grades and does not provide a direct measure of growth across grades. For grades K through 2, the Reading Composite Score is also not comparable across different times of year and should not be used as an indicator of growth within a grade. However, because the logic and procedures used to establish benchmark goals are consistent across grades and times of year, the percent of students at different benchmark status levels can be compared, even though the mean scores are not comparable.

References

Bruck, M., Genesee, F., & Caravolas, M. (1997). A cross-linguistic study of early literacy acquisition. In B. Blachman (Ed.), *Foundations of reading acquisition and dyslexia: Implications for early intervention* (pp. 145–162). Mahwah, NJ: Lawrence Erlbaum Associates.

Kim, Y., Foorman, B., Petscher, Y., & Zhou, C. (2010). The contributions of phonological awareness and letter-name knowledge to letter-sound acquisition—a cross-classified multilevel model approach. *Journal of Educational Psychology, 102*(2), 313–326.

Mann, V.A., & Wimmer, H. (2002). Phoneme awareness and pathways into literacy: A comparison of German and American children. *Reading and Writing: An Interdisciplinary Journal, 15,* 653–682.

National Early Literacy Panel. (2008). *Developing early literacy: Report of the National Early Literacy Panel.* Washington, DC: National Institute for Literacy.

Piasta, S.B., & Wagner, R.K. (2010). Learning letter names and sounds: Effects of instruction, letter type, and phonological processing skill. *Journal of Experimental Child Psychology, 105,* 324–344.

Schatschneider, C., Fletcher, J. M., Francis, D. J., Carlson, C. D., & Foorman, B. R. (2004). Kindergarten Prediction of Reading Skills: A Longitudinal Comparative Analysis. *Journal of Educational Psychology, 96*(2), 265–282.

Walsh, D.J., Price, G.G., & Gillingham, M.G. (1988). The critical but transitory importance of letter naming. *Reading Research Quarterly, 23,* 108–122.

Williams, K.T. (2001). Group Reading Assessment and Diagnostic Evaluation (GRADE). New York: Pearson.

Acadience Reading: Summary of Benchmark Goals and Cut Points for Risk

Each cell shows three values: ABOVE BENCHMARK (top number) / BENCHMARK GOAL (bold middle number) / CUT POINT FOR RISK (bottom number).

Reading Composite Score

	K Beg	K Mid	K End	1 Beg	1 Mid	1 End	2 Beg	2 Mid	2 End	3 Beg	3 Mid	3 End	4 Beg	4 Mid	4 End	5 Beg	5 Mid	5 End	6 Beg	6 Mid	6 End
Above	38	156	152	129	177	208	202	256	287	289	349	405	341	383	446	386	411	466	435	461	478
Goal	**26**	**122**	**119**	**113**	**130**	**155**	**141**	**190**	**238**	**220**	**285**	**330**	**290**	**330**	**391**	**357**	**372**	**415**	**344**	**358**	**380**
Cut	13	85	89	97	100	111	109	145	180	180	235	280	245	290	330	258	310	340	280	285	324

First Sound Fluency (FSF)

	K Beg	K Mid
Above	16	43
Goal	**10**	**30**
Cut	5	20

Phoneme Segmentation Fluency (PSF)

	K End	1 Beg	1 Mid
Above	44	56	47
Goal	**20**	**40**	**40**
Cut	10	25	25

Nonsense Word Fluency (NWF)

Correct Letter Sounds

	K Mid	K End	1 Beg	1 Mid	1 End	2 Beg
Above	28	34	40	59	81	72
Goal	**17**	**28**	**27**	**43**	**58**	**54**
Cut	8	15	18	33	47	35

Whole Words Read

	1 Mid	1 End	2 Beg
Above	4	17	25
Goal	**1**	**8**	**13**
Cut	0	3	6

Oral Reading Fluency (ORF)

Words Correct

	1 Mid	1 End	2 Beg	2 Mid	2 End	3 Beg	3 Mid	3 End	4 Beg	4 Mid	4 End	5 Beg	5 Mid	5 End	6 Beg	6 Mid	6 End
Above	34	67	68	91	104	90	105	118	104	121	133	121	133	143	139	141	151
Goal	**23**	**47**	**52**	**72**	**87**	**70**	**86**	**100**	**90**	**103**	**115**	**111**	**120**	**130**	**107**	**109**	**120**
Cut	16	32	37	55	65	55	68	80	70	79	95	96	101	105	90	92	95

Accuracy

	1 Mid	1 End	2 Beg	2 Mid	2 End	3 Beg	3 Mid	3 End	4 Beg	4 Mid	4 End	5 Beg	5 Mid	5 End	6 Beg	6 Mid	6 End
Above	86%	97%	96%	99%	99%	98%	99%	99%	98%	99%	100%	99%	99%	100%	99%	99%	100%
Goal	**78%**	**90%**	**90%**	**96%**	**97%**	**95%**	**96%**	**97%**	**96%**	**97%**	**98%**	**98%**	**98%**	**99%**	**97%**	**97%**	**98%**
Cut	68%	82%	81%	91%	93%	89%	92%	94%	93%	94%	95%	95%	96%	97%	94%	94%	96%

Retell

	1 End	2 Beg	2 Mid	2 End	3 Beg	3 Mid	3 End	4 Beg	4 Mid	4 End	5 Beg	5 Mid	5 End	6 Beg	6 Mid	6 End
Above	17	25	31	39	33	40	46	36	39	46	40	46	52	43	48	50
Goal	**15**	**16**	**21**	**27**	**20**	**26**	**30**	**27**	**30**	**33**	**33**	**36**	**36**	**27**	**29**	**32**
Cut	0	8	13	18	10	18	20	14	20	24	22	25	25	16	18	24

Retell Quality of Response

	2 Mid	2 End	3 Beg	3 Mid	3 End	4 Beg	4 Mid	4 End	5 Beg	5 Mid	5 End	6 Beg	6 Mid	6 End
Goal	**2**	**2**	**2**	**2**	**3**	**2**	**2**	**3**	**2**	**3**	**3**	**2**	**2**	**3**
Cut	1	1	1	1	2	1	1	2	1	2	2	1	1	2

Maze Adjusted Score

	3 Beg	3 Mid	3 End	4 Beg	4 Mid	4 End	5 Beg	5 Mid	5 End	6 Beg	6 Mid	6 End
Above	11	16	23	18	20	28	21	21	28	27	30	30
Goal	**8**	**11**	**19**	**15**	**17**	**24**	**18**	**20**	**24**	**18**	**19**	**21**
Cut	5	7	14	10	12	20	12	13	18	14	14	15

Reading Composite Score: A combination of multiple Acadience Reading scores, which provides the best overall estimate of the student's reading proficiency. For information on how to calculate the composite score, see the *Acadience Reading Benchmark Goals and Composite Score* document available from http://acadiencelearning.org/.

ABOVE BENCHMARK (number above bold number in each box): Students scoring above the benchmark are highly likely to achieve important reading outcomes (approximately 90% to 99%). These scores are identified as *Above Benchmark*. While students scoring Above Benchmark are likely to need *Core Support*, some may benefit from instruction on more advanced skills.

BENCHMARK GOAL (large bold number in the middle of the box): Students scoring at or above the benchmark goal have the odds in their favor (approximately 80% to 90% overall) of achieving later important reading outcomes. These scores are identified as *At or Above Benchmark* and the students are likely to need *Core Support*.

CUT POINT FOR RISK (number below bold number in each box): Students scoring below the cut point for risk are unlikely (approximately 10%–20%) to achieve subsequent goals without receiving additional, targeted instructional support. These scores are identified as *Well Below Benchmark* and the students are likely to need *Intensive Support*.

Scores below the benchmark goal and at or above the cut point for risk are identified as *Below Benchmark*. In this range, a student's future performance is harder to predict, and these students are likely to need *Strategic Support*.

Note: There is no benchmark goal for Letter Naming Fluency (LNF).

This is a summary of the Acadience Reading benchmark goals. For a full description, see the *Acadience Reading Benchmark Goals and Composite Score* document available from http://acadiencelearning.org/.

Acadience is a registered trademark of Dynamic Measurement Group, Inc. This page is adapted from a chart developed by Cache County School District.

Kindergarten Benchmark Goals and Cut Points for Risk

Acadience Reading Measure	Benchmark Status	Likely Need for Support	Beginning of Year	Middle of Year	End of Year
Reading Composite Score	Above Benchmark	Likely to Need Core Support[a]	38 +	156 +	152 +
	At Benchmark	**Likely to Need Core Support**[b]	**26** - 37	**122** - 155	**119** - 151
	Below Benchmark	Likely to Need Strategic Support	*13 - 25*	*85 - 121*	*89 - 118*
	Well Below Benchmark	Likely to Need Intensive Support	0 - 12	0 - 84	0 - 88
FSF	Above Benchmark	Likely to Need Core Support[a]	16 +	43 +	
	At Benchmark	**Likely to Need Core Support**[b]	**10** - 15	**30** - 42	
	Below Benchmark	Likely to Need Strategic Support	*5 - 9*	*20 - 29*	
	Well Below Benchmark	Likely to Need Intensive Support	0 - 4	0 - 19	
PSF	Above Benchmark	Likely to Need Core Support[a]		44 +	56 +
	At Benchmark	**Likely to Need Core Support**[b]		**20** - 43	**40** - 55
	Below Benchmark	Likely to Need Strategic Support		*10 - 19*	*25 - 39*
	Well Below Benchmark	Likely to Need Intensive Support		0 - 9	0 - 24
NWF-CLS	Above Benchmark	Likely to Need Core Support[a]		28 +	40 +
	At Benchmark	**Likely to Need Core Support**[b]		**17** - 27	**28** - 39
	Below Benchmark	Likely to Need Strategic Support		*8 - 16*	*15 - 27*
	Well Below Benchmark	Likely to Need Intensive Support		0 - 7	0 - 14

The benchmark goal is the number that is **bold**. The cut point for risk is the number that is *italicized*.
[a] Some students may benefit from instruction on more advanced skills.
[b] Some students may require monitoring and strategic support on component skills.

First Grade Benchmark Goals and Cut Points for Risk

Acadience Reading Measure	Benchmark Status	Likely Need for Support	Beginning of Year	Middle of Year	End of Year
Reading Composite Score	Above Benchmark	Likely to Need Core Support[a]	129 +	177 +	208 +
	At Benchmark	**Likely to Need Core Support**[b]	**113** - 128	**130** - 176	**155** - 207
	Below Benchmark	Likely to Need Strategic Support	*97* - 112	*100* - 129	*111* - 154
	Well Below Benchmark	Likely to Need Intensive Support	0 - 96	0 - 99	0 - 110
PSF	Above Benchmark	Likely to Need Core Support[a]	47 +		
	At Benchmark	**Likely to Need Core Support**[b]	**40** - 46		
	Below Benchmark	Likely to Need Strategic Support	*25* - 39		
	Well Below Benchmark	Likely to Need Intensive Support	0 - 24		
NWF-CLS	Above Benchmark	Likely to Need Core Support[a]	34 +	59 +	81 +
	At Benchmark	**Likely to Need Core Support**[b]	**27** - 33	**43** - 58	**58** - 80
	Below Benchmark	Likely to Need Strategic Support	*18* - 26	*33* - 42	*47* - 57
	Well Below Benchmark	Likely to Need Intensive Support	0 - 17	0 - 32	0 - 46
NWF-WWR	Above Benchmark	Likely to Need Core Support[a]	4 +	17 +	25 +
	At Benchmark	**Likely to Need Core Support**[b]	**1** - 3	**8** - 16	**13** - 24
	Below Benchmark	Likely to Need Strategic Support	0	*3* - 7	*6* - 12
	Well Below Benchmark	Likely to Need Intensive Support		0 - 2	0 - 5
ORF Words Correct	Above Benchmark	Likely to Need Core Support[a]		34 +	67 +
	At Benchmark	**Likely to Need Core Support**[b]		**23** - 33	**47** - 66
	Below Benchmark	Likely to Need Strategic Support		*16* - 22	*32* - 46
	Well Below Benchmark	Likely to Need Intensive Support		0 - 15	0 - 31
ORF Accuracy	Above Benchmark	Likely to Need Core Support[a]		86% +	97% +
	At Benchmark	**Likely to Need Core Support**[b]		**78%** - 85%	**90%** - 96%
	Below Benchmark	Likely to Need Strategic Support		*68%* - 77%	*82%* - 89%
	Well Below Benchmark	Likely to Need Intensive Support		0% - 67%	0% - 81%
Retell	Above Benchmark	Likely to Need Core Support[a]			17 +
	At Benchmark	**Likely to Need Core Support**[b]			**15** - 16
	Below Benchmark	Likely to Need Strategic Support			*0* - 14
	Well Below Benchmark	Likely to Need Intensive Support			

The benchmark goal is the number that is **bold**. The cut point for risk is the number that is *italicized*.
[a] Some students may benefit from instruction on more advanced skills.
[b] Some students may require monitoring and strategic support on component skills.

Second Grade Benchmark Goals and Cut Points for Risk

Acadience Reading Measure	Benchmark Status	Likely Need for Support	Beginning of Year	Middle of Year	End of Year
Reading Composite Score	Above Benchmark	Likely to Need Core Support[a]	202 +	256 +	287 +
	At Benchmark	**Likely to Need Core Support**[b]	**141** - 201	**190** - 255	**238** - 286
	Below Benchmark	Likely to Need Strategic Support	*109* - 140	*145* - 189	*180* - 237
	Well Below Benchmark	Likely to Need Intensive Support	0 - 108	0 - 144	0 - 179
NWF-CLS	Above Benchmark	Likely to Need Core Support[a]	72 +		
	At Benchmark	**Likely to Need Core Support**[b]	**54** - 71		
	Below Benchmark	Likely to Need Strategic Support	*35* - 53		
	Well Below Benchmark	Likely to Need Intensive Support	0 - 34		
NWF-WWR	Above Benchmark	Likely to Need Core Support[a]	21 +		
	At Benchmark	**Likely to Need Core Support**[b]	**13** - 20		
	Below Benchmark	Likely to Need Strategic Support	*6* - 12		
	Well Below Benchmark	Likely to Need Intensive Support	0 - 5		
ORF Words Correct	Above Benchmark	Likely to Need Core Support[a]	68 +	91 +	104 +
	At Benchmark	**Likely to Need Core Support**[b]	**52** - 67	**72** - 90	**87** - 103
	Below Benchmark	Likely to Need Strategic Support	*37* - 51	*55* - 71	*65* - 86
	Well Below Benchmark	Likely to Need Intensive Support	0 - 36	0 - 54	0 - 64
ORF Accuracy	Above Benchmark	Likely to Need Core Support[a]	96% +	99% +	99% +
	At Benchmark	**Likely to Need Core Support**[b]	**90%** - 95%	**96%** - 98%	**97%** - 98%
	Below Benchmark	Likely to Need Strategic Support	*81%* - 89%	*91%* - 95%	*93%* - 96%
	Well Below Benchmark	Likely to Need Intensive Support	0% - 80%	0% - 90%	0% - 92%
Retell	Above Benchmark	Likely to Need Core Support[a]	25 +	31 +	39 +
	At Benchmark	**Likely to Need Core Support**[b]	**16** - 24	**21** - 30	**27** - 38
	Below Benchmark	Likely to Need Strategic Support	*8* - 15	*13* - 20	*18* - 26
	Well Below Benchmark	Likely to Need Intensive Support	0 - 7	0 - 12	0 - 17
Retell Quality of Response	**At or Above Benchmark**	**Likely to Need Core Support**[b]		**2 +**	**2 +**
	Below Benchmark	Likely to Need Strategic Support		1	1
	Well Below Benchmark	Likely to Need Intensive Support			

The benchmark goal is the number that is **bold**. The cut point for risk is the number that is *italicized*.
[a] Some students may benefit from instruction on more advanced skills.
[b] Some students may require monitoring and strategic support on component skills.

Third Grade Benchmark Goals and Cut Points for Risk

Acadience Reading Measure	Benchmark Status	Likely Need for Support	Beginning of Year	Middle of Year	End of Year
Reading Composite Score	Above Benchmark	Likely to Need Core Support[a]	289 +	349 +	405 +
	At Benchmark	**Likely to Need Core Support**[b]	**220** - 288	**285** - 348	**330** - 404
	Below Benchmark	Likely to Need Strategic Support	*180* - 219	*235* - 284	*280* - 329
	Well Below Benchmark	Likely to Need Intensive Support	0 - 179	0 - 234	0 - 279
ORF Words Correct	Above Benchmark	Likely to Need Core Support[a]	90 +	105 +	118 +
	At Benchmark	**Likely to Need Core Support**[b]	**70** - 89	**86** - 104	**100** - 117
	Below Benchmark	Likely to Need Strategic Support	*55* - 69	*68* - 85	*80* - 99
	Well Below Benchmark	Likely to Need Intensive Support	0 - 54	0 - 67	0 - 79
ORF Accuracy	Above Benchmark	Likely to Need Core Support[a]	98% +	99% +	99% +
	At Benchmark	**Likely to Need Core Support**[b]	**95%** - 97%	**96%** - 98%	**97%** - 98%
	Below Benchmark	Likely to Need Strategic Support	*89%* - 94%	*92%* - 95%	*94%* - 96%
	Well Below Benchmark	Likely to Need Intensive Support	0% - 88%	0% - 91%	0% - 93%
Retell	Above Benchmark	Likely to Need Core Support[a]	33 +	40 +	46 +
	At Benchmark	**Likely to Need Core Support**[b]	**20** - 32	**26** - 39	**30** - 45
	Below Benchmark	Likely to Need Strategic Support	*10* - 19	*18* - 25	*20* - 29
	Well Below Benchmark	Likely to Need Intensive Support	0 - 9	0 - 17	0 - 19
Retell Quality of Response	**At or Above Benchmark**	**Likely to Need Core Support**[b]	**2 +**	**2 +**	**3 +**
	Below Benchmark	Likely to Need Strategic Support	*1*	*1*	*2*
	Well Below Benchmark	Likely to Need Intensive Support			1
Maze Adjusted Score	Above Benchmark	Likely to Need Core Support[a]	11 +	16 +	23 +
	At Benchmark	**Likely to Need Core Support**[b]	**8** - 10	**11** - 15	**19** - 22
	Below Benchmark	Likely to Need Strategic Support	*5* - 7	*7* - 10	*14* - 18
	Well Below Benchmark	Likely to Need Intensive Support	0 - 4	0 - 6	0 - 13

The benchmark goal is the number that is **bold**. The cut point for risk is the number that is *italicized*.
[a] Some students may benefit from instruction on more advanced skills.
[b] Some students may require monitoring and strategic support on component skills.

Fourth Grade Benchmark Goals and Cut Points for Risk

Acadience Reading Measure	Benchmark Status	Likely Need for Support	Beginning of Year	Middle of Year	End of Year
Reading Composite Score	Above Benchmark	Likely to Need Core Support[a]	341 +	383 +	446 +
	At Benchmark	**Likely to Need Core Support**[b]	**290** - 340	**330** - 382	**391** - 445
	Below Benchmark	Likely to Need Strategic Support	*245 - 289*	*290 - 329*	*330 - 390*
	Well Below Benchmark	Likely to Need Intensive Support	0 - 244	0 - 289	0 - 329
ORF Words Correct	Above Benchmark	Likely to Need Core Support[a]	104 +	121 +	133 +
	At Benchmark	**Likely to Need Core Support**[b]	**90** - 103	**103** - 120	**115** - 132
	Below Benchmark	Likely to Need Strategic Support	*70 - 89*	*79 - 102*	*95 - 114*
	Well Below Benchmark	Likely to Need Intensive Support	0 - 69	0 - 78	0 - 94
ORF Accuracy	Above Benchmark	Likely to Need Core Support[a]	98% +	99% +	100% +
	At Benchmark	**Likely to Need Core Support**[b]	**96%** - 97%	**97%** - 98%	**98%** - 99%
	Below Benchmark	Likely to Need Strategic Support	*93% - 95%*	*94% - 96%*	*95% - 97%*
	Well Below Benchmark	Likely to Need Intensive Support	0% - 92%	0% - 93%	0% - 94%
Retell	Above Benchmark	Likely to Need Core Support[a]	36 +	39 +	46 +
	At Benchmark	**Likely to Need Core Support**[b]	**27** - 35	**30** - 38	**33** - 45
	Below Benchmark	Likely to Need Strategic Support	*14 - 26*	*20 - 29*	*24 - 32*
	Well Below Benchmark	Likely to Need Intensive Support	0 - 13	0 - 19	0 - 23
Retell Quality of Response	**At or Above Benchmark**	**Likely to Need Core Support**[b]	**2 +**	**2 +**	**3 +**
	Below Benchmark	Likely to Need Strategic Support	*1*	*1*	*2*
	Well Below Benchmark	Likely to Need Intensive Support			1
Maze Adjusted Score	Above Benchmark	Likely to Need Core Support[a]	18 +	20 +	28 +
	At Benchmark	**Likely to Need Core Support**[b]	**15** - 17	**17** - 19	**24** - 27
	Below Benchmark	Likely to Need Strategic Support	*10 - 14*	*12 - 16*	*20 - 23*
	Well Below Benchmark	Likely to Need Intensive Support	0 - 9	0 - 11	0 - 19

The benchmark goal is the number that is **bold**. The cut point for risk is the number that is *italicized*.
[a] Some students may benefit from instruction on more advanced skills.
[b] Some students may require monitoring and strategic support on component skills.

Fifth Grade Benchmark Goals and Cut Points for Risk

Acadience Reading Measure	Benchmark Status	Likely Need for Support	Beginning of Year	Middle of Year	End of Year
Reading Composite Score	Above Benchmark	Likely to Need Core Support[a]	386 +	411 +	466 +
	At Benchmark	**Likely to Need Core Support**[b]	**357** - 385	**372** - 410	**415** - 465
	Below Benchmark	Likely to Need Strategic Support	*258* - 356	*310* - 371	*340* - 414
	Well Below Benchmark	Likely to Need Intensive Support	0 - 257	0 - 309	0 - 339
ORF Words Correct	Above Benchmark	Likely to Need Core Support[a]	121 +	133 +	143 +
	At Benchmark	**Likely to Need Core Support**[b]	**111** - 120	**120** - 132	**130** - 142
	Below Benchmark	Likely to Need Strategic Support	*96* - 110	*101* - 119	*105* - 129
	Well Below Benchmark	Likely to Need Intensive Support	0 - 95	0 - 100	0 - 104
ORF Accuracy	Above Benchmark	Likely to Need Core Support[a]	99% +	99% +	100%
	At Benchmark	**Likely to Need Core Support**[b]	**98%**	**98%**	**99%**
	Below Benchmark	Likely to Need Strategic Support	*95% - 97%*	*96% - 97%*	*97% - 98%*
	Well Below Benchmark	Likely to Need Intensive Support	0% - 94%	0% - 95%	0% - 96%
Retell	Above Benchmark	Likely to Need Core Support[a]	40 +	46 +	52 +
	At Benchmark	**Likely to Need Core Support**[b]	**33** - 39	**36** - 45	**36** - 51
	Below Benchmark	Likely to Need Strategic Support	*22* - 32	*25* - 35	*25* - 35
	Well Below Benchmark	Likely to Need Intensive Support	0 - 21	0 - 24	0 - 24
Retell Quality of Response	**At or Above Benchmark**	**Likely to Need Core Support**[b]	**2 +**	**3 +**	**3 +**
	Below Benchmark	Likely to Need Strategic Support	*1*	*2*	*2*
	Well Below Benchmark	Likely to Need Intensive Support		1	1
Maze Adjusted Score	Above Benchmark	Likely to Need Core Support[a]	21 +	21 +	28 +
	At Benchmark	**Likely to Need Core Support**[b]	**18** - 20	**20**	**24** - 27
	Below Benchmark	Likely to Need Strategic Support	*12* - 17	*13* - 19	*18* - 23
	Well Below Benchmark	Likely to Need Intensive Support	0 - 11	0 - 12	0 - 17

The benchmark goal is the number that is **bold**. The cut point for risk is the number that is *italicized*.
[a] Some students may benefit from instruction on more advanced skills.
[b] Some students may require monitoring and strategic support on component skills.

Sixth Grade Benchmark Goals and Cut Points for Risk

Acadience Reading Measure	Benchmark Status	Likely Need for Support	Beginning of Year	Middle of Year	End of Year
Reading Composite Score	Above Benchmark	Likely to Need Core Support[a]	435 +	461 +	478 +
	At Benchmark	**Likely to Need Core Support**[b]	*344* - 434	*358* - 460	*380* - 477
	Below Benchmark	Likely to Need Strategic Support	*280 - 343*	*285 - 357*	*324 - 379*
	Well Below Benchmark	Likely to Need Intensive Support	0 - 279	0 - 284	0 - 323
ORF Words Correct	Above Benchmark	Likely to Need Core Support[a]	139 +	141 +	151 +
	At Benchmark	**Likely to Need Core Support**[b]	**107** - 138	**109** - 140	**120** - 150
	Below Benchmark	Likely to Need Strategic Support	*90* - 106	*92* - 108	*95* - 119
	Well Below Benchmark	Likely to Need Intensive Support	0 - 89	0 - 91	0 - 94
ORF Accuracy	Above Benchmark	Likely to Need Core Support[a]	99% +	99% +	100%
	At Benchmark	**Likely to Need Core Support**[b]	**97%** - 98%	**97%** - 98%	**98%** - 99%
	Below Benchmark	Likely to Need Strategic Support	*94% - 96%*	*94% - 96%*	*96% - 97%*
	Well Below Benchmark	Likely to Need Intensive Support	0% - 93%	0% - 93%	0% - 95%
Retell	Above Benchmark	Likely to Need Core Support[a]	43 +	48 +	50 +
	At Benchmark	**Likely to Need Core Support**[b]	**27** - 42	**29** - 47	**32** - 49
	Below Benchmark	Likely to Need Strategic Support	*16 - 26*	*18 - 28*	*24 - 31*
	Well Below Benchmark	Likely to Need Intensive Support	0 - 15	0 - 17	0 - 23
Retell Quality of Response	**At or Above Benchmark**	**Likely to Need Core Support**[b]	**2 +**	**2 +**	**3 +**
	Below Benchmark	Likely to Need Strategic Support	*1*	*1*	*2*
	Well Below Benchmark	Likely to Need Intensive Support			1
Maze Adjusted Score	Above Benchmark	Likely to Need Core Support[a]	27 +	30 +	30 +
	At Benchmark	**Likely to Need Core Support**[b]	**18** - 26	**19** - 29	**21** - 29
	Below Benchmark	Likely to Need Strategic Support	*14 - 17*	*14 - 18*	*15 - 20*
	Well Below Benchmark	Likely to Need Intensive Support	0 - 13	0 - 13	0 - 14

The benchmark goal is the number that is **bold**. The cut point for risk is the number that is *italicized*.
[a] Some students may benefit from instruction on more advanced skills.
[b] Some students may require monitoring and strategic support on component skills.

Kindergarten Percentage of Students Who Meet Later Outcomes on the Reading Composite Score Based On Benchmark Status on Individual Acadience Reading Measures

Acadience Reading Measure	Benchmark Status	Percent of students **At or Above** Benchmark on **middle-of-year** Reading Composite Score based on beginning-of-year status	Percent of students **Above** Benchmark on **middle-of-year** Reading Composite Score based on beginning-of-year status	Percent of students **At or Above** Benchmark on **end-of-year** Reading Composite Score based on middle-of-year status	Percent of students **Above** Benchmark on **end-of-year** Reading Composite Score based on middle-of-year status
Reading Composite Score	**At or Above Benchmark**	**85%**	58%	**93%**	59%
	Above Benchmark	91%	67%	98%	77%
	At Benchmark	70%	35%	85%	32%
	Below Benchmark	54%	24%	56%	13%
	Well Below Benchmark	32%	12%	18%	3%
FSF	**At or Above Benchmark**	**83%**	57%	**86%**	52%
	Above Benchmark	88%	64%	93%	65%
	At Benchmark	69%	36%	80%	41%
	Below Benchmark	56%	26%	54%	19%
	Well Below Benchmark	42%	18%	22%	5%
PSF	**At or Above Benchmark**	–	–	**86%**	52%
	Above Benchmark	–	–	94%	66%
	At Benchmark	–	–	79%	38%
	Below Benchmark	–	–	53%	18%
	Well Below Benchmark	–	–	26%	7%
NWF Correct Letter Sounds	**At or Above Benchmark**	–	–	**87%**	53%
	Above Benchmark	–	–	96%	72%
	At Benchmark	–	–	78%	31%
	Below Benchmark	–	–	47%	11%
	Well Below Benchmark	–	–	18%	4%

Note. This table shows the percent of students that are on track on the Reading Composite Score at the middle and end of the year based on the student's Acadience Reading measure score at the beginning and middle of the year. N = 441,923 students who had Acadience Reading data for the 2013–2014 school year. Data exported from mCLASS®, VPORT®, and Acadience Data Management.

First Grade Percentage of Students Who Meet Later Outcomes on the Reading Composite Score Based On Benchmark Status on Individual Acadience Reading Measures

Acadience Reading Measure	Benchmark Status	Percent of students **At or Above** Benchmark on **middle-of-year** Reading Composite Score based on beginning-of-year status	Percent of students **Above** Benchmark on **middle-of-year** Reading Composite Score based on beginning-of-year status	Percent of students **At or Above** Benchmark on **end-of-year** Reading Composite Score based on middle-of-year status	Percent of students **Above** Benchmark on **end-of-year** Reading Composite Score based on middle-of-year status
Reading Composite Score	**At or Above Benchmark**	**87%**	68%	**92%**	66%
	Above Benchmark	93%	79%	99%	85%
	At Benchmark	74%	44%	75%	20%
	Below Benchmark	59%	29%	36%	5%
	Well Below Benchmark	28%	11%	7%	1%
PSF	**At or Above Benchmark**	**77%**	56%	–	–
	Above Benchmark	79%	59%	–	–
	At Benchmark	74%	52%	–	–
	Below Benchmark	64%	43%	–	–
	Well Below Benchmark	36%	21%	–	–
NWF Correct Letter Sounds	**At or Above Benchmark**	**85%**	66%	**86%**	63%
	Above Benchmark	91%	77%	95%	81%
	At Benchmark	68%	37%	67%	28%
	Below Benchmark	49%	22%	43%	12%
	Well Below Benchmark	22%	8%	18%	4%
NWF Whole Words Read	**At or Above Benchmark**	83%	64%	**83%**	59%
	Above Benchmark	92%	78%	96%	80%
	At Benchmark	66%	36%	63%	25%
	Below Benchmark	37%	16%	36%	10%
	Well Below Benchmark	–	–	17%	5%
ORF Words Correct	**At or Above Benchmark**			**91%**	66%
	Above Benchmark			98%	83%
	At Benchmark			74%	24%
	Below Benchmark			35%	6%
	Well Below Benchmark			7%	1%
ORF Accuracy	**At or Above Benchmark**			**91%**	67%
	Above Benchmark			97%	80%
	At Benchmark			74%	27%
	Below Benchmark			43%	10%
	Well Below Benchmark			9%	2%

Note. This table shows the percent of students that are on track on the Reading Composite Score at the middle and end of the year based on the student's Acadience Reading measure score at the beginning and middle of the year. N = 452,530 students who had Acadience Reading data for the 2013–2014 school year. Data exported from mCLASS®, VPORT®, and Acadience Data Management.

Second Grade Percentage of Students Who Meet Later Outcomes on the Reading Composite Score Based On Benchmark Status on Individual Acadience Reading Measures

Acadience Reading Measure	Benchmark Status	Percent of students **At or Above** Benchmark on **middle-of-year** Reading Composite Score based on beginning-of-year status	Percent of students **Above** Benchmark on **middle-of-year** Reading Composite Score based on beginning-of-year status	Percent of students **At or Above** Benchmark on **end-of-year** Reading Composite Score based on middle-of-year status	Percent of students **Above** Benchmark on **end-of-year** Reading Composite Score based on middle-of-year status
Reading Composite Score	**At or Above Benchmark**	**93%**	64%	**91%**	64%
	Above Benchmark	99%	83%	98%	84%
	At Benchmark	85%	36%	77%	28%
	Below Benchmark	46%	8%	35%	7%
	Well Below Benchmark	11%	1%	8%	1%
NWF Correct Letter Sounds	**At or Above Benchmark**	**92%**	66%	–	–
	Above Benchmark	96%	76%	–	–
	At Benchmark	82%	46%	–	–
	Below Benchmark	61%	26%	–	–
	Well Below Benchmark	37%	13%	–	–
NWF Whole Words Read	**At or Above Benchmark**	**90%**	64%	–	–
	Above Benchmark	96%	76%	–	–
	At Benchmark	80%	43%	–	–
	Below Benchmark	57%	23%	–	–
	Well Below Benchmark	36%	13%	–	–
ORF Words Correct	**At or Above Benchmark**	**96%**	71%	**94%**	69%
	Above Benchmark	99%	84%	98%	84%
	At Benchmark	90%	42%	85%	40%
	Below Benchmark	64%	15%	54%	15%
	Well Below Benchmark	16%	2%	12%	2%
ORF Accuracy	**At or Above Benchmark**	**92%**	63%	**91%**	65%
	Above Benchmark	98%	79%	96%	77%
	At Benchmark	82%	37%	81%	44%
	Below Benchmark	45%	11%	44%	14%
	Well Below Benchmark	11%	2%	11%	4%
Retell	**At or Above Benchmark**	**89%**	63%	**84%**	60%
	Above Benchmark	94%	74%	91%	72%
	At Benchmark	80%	41%	71%	37%
	Below Benchmark	62%	22%	48%	18%
	Well Below Benchmark	33%	9%	24%	8%

Note. This table shows the percent of students that are on track on the Reading Composite Score at the middle and end of the year based on the student's Acadience Reading measure score at the beginning and middle of the year. N = 394,821 students who had Acadience Reading data for the 2013–2014 school year. Data exported from mCLASS®, VPORT®, and Acadience Data Management.

Third Grade Percentage of Students Who Meet Later Outcomes on the Reading Composite Score Based On Benchmark Status on Individual Acadience Reading Measures

Acadience Reading Measure	Benchmark Status	Percent of students **At or Above** Benchmark on **middle-of-year** Reading Composite Score based on beginning-of-year status	Percent of students **Above** Benchmark on **middle-of-year** Reading Composite Score based on beginning-of-year status	Percent of students **At or Above** Benchmark on **end-of-year** Reading Composite Score based on middle-of-year status	Percent of students **Above** Benchmark on **end-of-year** Reading Composite Score based on middle-of-year status
Reading Composite Score	**At or Above Benchmark**	**90%**	62%	**93%**	64%
	Above Benchmark	98%	82%	99%	84%
	At Benchmark	76%	29%	83%	29%
	Below Benchmark	43%	9%	46%	7%
	Well Below Benchmark	12%	2%	9%	1%
ORF Words Correct	**At or Above Benchmark**	**91%**	64%	**92%**	65%
	Above Benchmark	97%	82%	98%	83%
	At Benchmark	79%	35%	83%	36%
	Below Benchmark	49%	12%	50%	11%
	Well Below Benchmark	14%	2%	12%	2%
ORF Accuracy	**At or Above Benchmark**	**87%**	60%	**85%**	57%
	Above Benchmark	94%	75%	92%	69%
	At Benchmark	78%	42%	76%	39%
	Below Benchmark	46%	16%	38%	11%
	Well Below Benchmark	10%	3%	8%	2%
Retell	**At or Above Benchmark**	**79%**	53%	**82%**	55%
	Above Benchmark	89%	68%	91%	69%
	At Benchmark	65%	32%	69%	34%
	Below Benchmark	39%	14%	46%	16%
	Well Below Benchmark	22%	8%	25%	7%
Maze Adjusted Score	**At or Above Benchmark**	**89%**	65%	**90%**	65%
	Above Benchmark	94%	76%	96%	78%
	At Benchmark	78%	43%	80%	44%
	Below Benchmark	58%	23%	58%	22%
	Well Below Benchmark	29%	9%	26%	7%

Note. This table shows the percent of students that are on track on the Reading Composite Score at the middle and end of the year based on the student's Acadience Reading measure score at the beginning and middle of the year. N = 303,928 students who had Acadience Reading data for the 2013–2014 school year. Data exported from mCLASS®, VPORT®, and Acadience Data Management.

Fourth Grade Percentage of Students Who Meet Later Outcomes on the Reading Composite Score Based On Benchmark Status on Individual Acadience Reading Measures

Acadience Reading Measure	Benchmark Status	Percent of students **At or Above** Benchmark on **middle-of-year** Reading Composite Score based on beginning-of-year status	Percent of students **Above** Benchmark on **middle-of-year** Reading Composite Score based on beginning-of-year status	Percent of students **At or Above** Benchmark on **end-of-year** Reading Composite Score based on middle-of-year status	Percent of students **Above** Benchmark on **end-of-year** Reading Composite Score based on middle-of-year status
Reading Composite Score	**At or Above Benchmark**	**91%**	68%	**91%**	65%
	Above Benchmark	97%	84%	98%	83%
	At Benchmark	76%	32%	77%	29%
	Below Benchmark	45%	11%	45%	8%
	Well Below Benchmark	9%	2%	9%	1%
ORF Words Correct	**At or Above Benchmark**	**92%**	72%	**90%**	66%
	Above Benchmark	97%	82%	97%	82%
	At Benchmark	79%	41%	76%	33%
	Below Benchmark	54%	19%	42%	11%
	Well Below Benchmark	12%	2%	7%	1%
ORF Accuracy	**At or Above Benchmark**	**82%**	60%	**80%**	55%
	Above Benchmark	89%	69%	88%	66%
	At Benchmark	68%	39%	67%	35%
	Below Benchmark	46%	20%	36%	12%
	Well Below Benchmark	12%	4%	7%	2%
Retell	**At or Above Benchmark**	**79%**	58%	**81%**	57%
	Above Benchmark	86%	68%	88%	66%
	At Benchmark	63%	37%	66%	36%
	Below Benchmark	40%	18%	45%	20%
	Well Below Benchmark	17%	6%	19%	7%
Maze Adjusted Score	**At or Above Benchmark**	**89%**	68%	**88%**	67%
	Above Benchmark	94%	78%	95%	79%
	At Benchmark	73%	39%	75%	41%
	Below Benchmark	47%	19%	50%	20%
	Well Below Benchmark	14%	4%	18%	5%

Note. This table shows the percent of students that are on track on the Reading Composite Score at the middle and end of the year based on the student's Acadience Reading measure score at the beginning and middle of the year. N = 114,567 students who had Acadience Reading data for the 2013–2014 school year. Data exported from mCLASS®, VPORT®, and Acadience Data Management.

Fifth Grade Percentage of Students Who Meet Later Outcomes on the Reading Composite Score Based On Benchmark Status on Individual Acadience Reading Measures

Acadience Reading Measure	Benchmark Status	Percent of students **At or Above** Benchmark on **middle-of-year** Reading Composite Score based on beginning-of-year status	Percent of students **Above** Benchmark on **middle-of-year** Reading Composite Score based on beginning-of-year status	Percent of students **At or Above** Benchmark on **end-of-year** Reading Composite Score based on middle-of-year status	Percent of students **Above** Benchmark on **end-of-year** Reading Composite Score based on middle-of-year status
Reading Composite Score	**At or Above Benchmark**	**92%**	76%	**90%**	68%
	Above Benchmark	96%	84%	96%	82%
	At Benchmark	75%	41%	73%	32%
	Below Benchmark	37%	13%	35%	9%
	Well Below Benchmark	3%	1%	3%	1%
ORF Words Correct	**At or Above Benchmark**	**91%**	76%	**91%**	72%
	Above Benchmark	95%	83%	95%	81%
	At Benchmark	75%	46%	76%	42%
	Below Benchmark	56%	26%	47%	18%
	Well Below Benchmark	16%	5%	8%	2%
ORF Accuracy	**At or Above Benchmark**	**80%**	63%	**76%**	55%
	Above Benchmark	89%	76%	88%	74%
	At Benchmark	76%	57%	71%	48%
	Below Benchmark	42%	22%	38%	18%
	Well Below Benchmark	11%	4%	10%	4%
Retell	**At or Above Benchmark**	**76%**	59%	**75%**	55%
	Above Benchmark	82%	67%	83%	66%
	At Benchmark	60%	39%	59%	34%
	Below Benchmark	42%	23%	39%	19%
	Well Below Benchmark	18%	9%	17%	7%
Maze Adjusted Score	**At or Above Benchmark**	**86%**	69%	**91%**	74%
	Above Benchmark	91%	78%	92%	77%
	At Benchmark	67%	41%	77%	48%
	Below Benchmark	45%	22%	52%	25%
	Well Below Benchmark	15%	6%	14%	4%

Note. This table shows the percent of students that are on track on the Reading Composite Score at the middle and end of the year based on the student's Acadience Reading measure score at the beginning and middle of the year. N = 98,565 students who had Acadience Reading data for the 2013–2014 school year. Data exported from mCLASS®, VPORT®, and Acadience Data Management.

Sixth Grade Percentage of Students Who Meet Later Outcomes on the Reading Composite Score Based On Benchmark Status on Individual Acadience Reading Measures

Acadience Reading Measure	Benchmark Status	Percent of students **At or Above** Benchmark on **middle-of-year** Reading Composite Score based on beginning-of-year status	Percent of students **Above** Benchmark on **middle-of-year** Reading Composite Score based on beginning-of-year status	Percent of students **At or Above** Benchmark on **end-of-year** Reading Composite Score based on middle-of-year status	Percent of students **Above** Benchmark on **end-of-year** Reading Composite Score based on middle-of-year status
Reading Composite Score	**At or Above Benchmark**	**93%**	54%	**94%**	55%
	Above Benchmark	99%	82%	100%	83%
	At Benchmark	85%	20%	87%	21%
	Below Benchmark	32%	2%	35%	1%
	Well Below Benchmark	3%	0%	3%	0%
ORF Words Correct	**At or Above Benchmark**	**92%**	55%	**93%**	56%
	Above Benchmark	99%	80%	99%	80%
	At Benchmark	85%	26%	85%	27%
	Below Benchmark	44%	3%	50%	5%
	Well Below Benchmark	8%	0%	11%	1%
ORF Accuracy	**At or Above Benchmark**	**86%**	49%	**86%**	50%
	Above Benchmark	92%	61%	94%	66%
	At Benchmark	83%	45%	83%	43%
	Below Benchmark	46%	12%	46%	10%
	Well Below Benchmark	9%	2%	10%	1%
Retell	**At or Above Benchmark**	**85%**	50%	**86%**	51%
	Above Benchmark	93%	65%	95%	68%
	At Benchmark	75%	33%	76%	31%
	Below Benchmark	52%	15%	49%	10%
	Well Below Benchmark	26%	5%	21%	3%
Maze Adjusted Score	**At or Above Benchmark**	**89%**	51%	**90%**	53%
	Above Benchmark	98%	77%	99%	78%
	At Benchmark	78%	24%	81%	27%
	Below Benchmark	36%	4%	43%	6%
	Well Below Benchmark	13%	2%	12%	1%

Note. This table shows the percent of students that are on track on the Reading Composite Score at the middle and end of the year based on the student's Acadience Reading measure score at the beginning and middle of the year. N = 32,337 students who had Acadience Reading data for the 2013–2014 school year. Data exported from mCLASS®, VPORT®, and Acadience Data Management.

Percent of Students Who Met Outcomes on the GRADE

Acadience Reading Measure	End-of-Year Benchmark Status	Likelihood of Being on Track on the GRADE by Grade Level						
		K	1	2	3	4	5	6
Reading Composite Score	At or Above Benchmark	74%	90%	89%	90%	84%	87%	93%
	Below Benchmark	50%	48%	45%	48%	58%	45%	45%
	Well Below Benchmark	36%	10%	14%	7%	3%	7%	13%
FSF	At or Above Benchmark	70%						
	Below Benchmark	56%						
	Well Below Benchmark	50%						
PSF	At or Above Benchmark	74%	83%					
	Below Benchmark	63%	59%					
	Well Below Benchmark	20%	32%					
NWF Correct Letter Sounds	At or Above Benchmark		90%					
	Below Benchmark		42%					
	Well Below Benchmark		10%					
NWF Whole Words Read	At or Above Benchmark		89%					
	Below Benchmark		36%					
	Well Below Benchmark		13%					
ORF Words Correct	At or Above Benchmark		87%	89%	89%	85%	83%	90%
	Below Benchmark		62%	43%	50%	59%	57%	64%
	Well Below Benchmark			14%	18%	3%	11%	25%
ORF Accuracy	At or Above Benchmark			88%	87%	75%	82%	90%
	Below Benchmark			39%	38%	54%	55%	69%
	Well Below Benchmark			26%	19%	6%	16%	30%
Retell	At or Above Benchmark			86%	86%	83%	86%	90%
	Below Benchmark			56%	48%	53%	39%	60%
	Well Below Benchmark			19%	20%	12%	20%	25%
Retell Quality of Response	At or Above Benchmark			81%	87%	87%	83%	92%
	Below Benchmark			41%	60%	52%	38%	68%
	Well Below Benchmark				15%	19%	11%	25%
Maze Adjusted Score	At or Above Benchmark				90%	80%	82%	90%
	Below Benchmark				48%	65%	61%	57%
	Well Below Benchmark				14%	14%	20%	20%

Note. This table shows the likelihood of being on track on the GRADE assessment administered at the end of the year, based on the student's individual end-of-year Acadience Reading measure benchmark status. The 40th percentile for the GRADE assessment was used to indicate whether the student was on track.

K *Kindergarten Reading Composite Score Worksheet*
© Dynamic Measurement Group, Inc. / October 15, 2018

The Reading Composite Score is used to interpret student results for Acadience Reading. Most data management services will calculate the composite score for you. If you do not use a data management service or if your data management service does not calculate it, you can use this worksheet to calculate the composite score.

Name: _____   **Class:** _____

Beginning of Year Benchmark

FSF Score = _____ [1]

LNF Score = _____ [2]

Reading Composite Score (add values 1–2) = [_____]

Do not calculate the composite score if any of the values are missing.

Middle of Year Benchmark

FSF Score = _____ [1]

LNF Score = _____ [2]

PSF Score = _____ [3]

NWF CLS Score = _____ [4]

Reading Composite Score (add values 1–4) = [_____]

Do not calculate the composite score if any of the values are missing.

End of Year Benchmark

LNF Score = _____ [1]

PSF Score = _____ [2]

NWF CLS Score = _____ [3]

Reading Composite Score (add values 1–3) = [_____]

Do not calculate the composite score if any of the values are missing.

1 *First Grade Reading Composite Score Worksheet*
© Dynamic Measurement Group, Inc. / October 15, 2018

The Reading Composite Score is used to interpret student results for Acadience Reading. Most data management services will calculate the composite score for you. If you do not use a data management service or if your data management service does not calculate it, you can use this worksheet to calculate the composite score.

Name: _____　　**Class:** _____

Beginning of Year Benchmark

LNF Score　**=**　_____ [1]

PSF Score　**=**　_____ [2]

NWF CLS Score　**=**　_____ [3]

Reading Composite Score (add values 1–3)　= ☐

Do not calculate the composite score if any of the values are missing.

Middle of Year Benchmark

NWF CLS Score　**=**　_____ [1]

NWF WWR Score　**=**　_____ [2]

ORF Words Correct　**=**　_____ [3]

ORF Accuracy Percent: _____ %
100 x (Words Correct / (Words Correct + Errors))

Accuracy Value from Table　**=**　_____ [4]

Reading Composite Score (add values 1–4)　= ☐

Do not calculate the composite score if any of the values are missing.

End of Year Benchmark

NWF WWR Score _____ **x 2**　**=**　_____ [1]

ORF Words Correct　**=**　_____ [2]

ORF Accuracy Percent: _____ %
100 x (Words Correct / (Words Correct + Errors))

Accuracy Value from Table　**=**　_____ [3]

Reading Composite Score (add values 1–3)　= ☐

Do not calculate the composite score if any of the values are missing.

Middle of Year	
ORF Accuracy Percent	Accuracy Value
0% – 49%	0
50% – 52%	2
53% – 55%	8
56% – 58%	14
59% – 61%	20
62% – 64%	26
65% – 67%	32
68% – 70%	38
71% – 73%	44
74% – 76%	50
77% – 79%	56
80% – 82%	62
83% – 85%	68
86% – 88%	74
89% – 91%	80
92% – 94%	86
95% – 97%	92
98% – 100%	98

End of Year	
ORF Accuracy Percent	Accuracy Value
0% – 64%	0
65% – 66%	3
67% – 68%	9
69% – 70%	15
71% – 72%	21
73% – 74%	27
75% – 76%	33
77% – 78%	39
79% – 80%	45
81% – 82%	51
83% – 84%	57
85% – 86%	63
87% – 88%	69
89% – 90%	75
91% – 92%	81
93% – 94%	87
95% – 96%	93
97% – 98%	99
99% – 100%	105

2 Second Grade Reading Composite Score Worksheet

© Dynamic Measurement Group, Inc. / October 15, 2018

The Reading Composite Score is used to interpret student results for Acadience Reading. Most data management services will calculate the composite score for you. If you do not use a data management service or if your data management service does not calculate it, you can use this worksheet to calculate the composite score.

Name: _____   **Class:** _____

Beginning of Year	
ORF Accuracy Percent	Accuracy Value
0% – 64%	0
65% – 66%	3
67% – 68%	9
69% – 70%	15
71% – 72%	21
73% – 74%	27
75% – 76%	33
77% – 78%	39
79% – 80%	45
81% – 82%	51
83% – 84%	57
85% – 86%	63
87% – 88%	69
89% – 90%	75
91% – 92%	81
93% – 94%	87
95% – 96%	93
97% – 98%	99
99% – 100%	105

Beginning of Year Benchmark

NWF WWR Score _____ **x 2** = _____ [1]

ORF Words Correct = _____ [2]

ORF Accuracy Percent: _____ %
100 x (Words Correct / (Words Correct + Errors))

Accuracy Value from Table = _____ [3]

Reading Composite Score (add values 1–3) = [   ]

Do not calculate the composite score if any of the values are missing.

Middle and End of Year	
ORF Accuracy Percent	Accuracy Value
0% – 85%	0
86%	8
87%	16
88%	24
89%	32
90%	40
91%	48
92%	56
93%	64
94%	72
95%	80
96%	88
97%	96
98%	104
99%	112
100%	120

Middle of Year Benchmark

ORF Words Correct = _____ [1]

Retell Score _____ **x 2** = _____ [2]

ORF Accuracy Percent: _____ %
100 x (Words Correct / (Words Correct + Errors))

Accuracy Value from Table = _____ [3]

Reading Composite Score (add values 1–3) = [   ]

If ORF is below 40 and Retell is not administered, use 0 for the Retell value only for calculating the Reading Composite Score. Do not calculate the composite score if any of the values are missing.

End of Year Benchmark

ORF Words Correct = _____ [1]

Retell Score _____ **x 2** = _____ [2]

ORF Accuracy Percent: _____ %
100 x (Words Correct / (Words Correct + Errors))

Accuracy Value from Table = _____ [3]

Reading Composite Score (add values 1–3) = [   ]

If ORF is below 40 and Retell is not administered, use 0 for the Retell value only for calculating the Reading Composite Score. Do not calculate the composite score if any of the values are missing.

3 Third Grade Reading Composite Score Worksheet
© Dynamic Measurement Group, Inc. / October 15, 2018

The Reading Composite Score is used to interpret student results for Acadience Reading. Most data management services will calculate the composite score for you. If you do not use a data management service or if your data management service does not calculate it, you can use this worksheet to calculate the composite score.

Name: _____   **Class:** _____

Beginning, Middle, and End of Year	
ORF Accuracy Percent	Accuracy Value
0% – 85%	0
86%	8
87%	16
88%	24
89%	32
90%	40
91%	48
92%	56
93%	64
94%	72
95%	80
96%	88
97%	96
98%	104
99%	112
100%	120

Beginning of Year Benchmark

ORF Words Correct **=** _____ [1]

Retell Score _____ **x 2** **=** _____ [2]

Maze Adjusted Score _____ **x 4** **=** _____ [3]

ORF Accuracy Percent: _____ %
100 x (Words Correct / (Words Correct + Errors))

Accuracy Value from Table **=** _____ [4]

Reading Composite Score (add values 1–4) = ☐

If ORF is below 40 and Retell is not administered, use 0 for the Retell value only for calculating the Reading Composite Score. Do not calculate the composite score if any of the values are missing.

Middle of Year Benchmark

ORF Words Correct **=** _____ [1]

Retell Score _____ **x 2** **=** _____ [2]

Maze Adjusted Score _____ **x 4** **=** _____ [3]

ORF Accuracy Percent: _____ %
100 x (Words Correct / (Words Correct + Errors))

Accuracy Value from Table **=** _____ [4]

Reading Composite Score (add values 1–4) = ☐

If ORF is below 40 and Retell is not administered, use 0 for the Retell value only for calculating the Reading Composite Score. Do not calculate the composite score if any of the values are missing.

End of Year Benchmark

ORF Words Correct **=** _____ [1]

Retell Score _____ **x 2** **=** _____ [2]

Maze Adjusted Score _____ **x 4** **=** _____ [3]

ORF Accuracy Percent: _____ %
100 x (Words Correct / (Words Correct + Errors))

Accuracy Value from Table **=** _____ [4]

Reading Composite Score (add values 1–4) = ☐

If ORF is below 40 and Retell is not administered, use 0 for the Retell value only for calculating the Reading Composite Score. Do not calculate the composite score if any of the values are missing.

4 Fourth Grade Reading Composite Score Worksheet
© Dynamic Measurement Group, Inc. / October 15, 2018

The Reading Composite Score is used to interpret student results for Acadience Reading. Most data management services will calculate the composite score for you. If you do not use a data management service or if your data management service does not calculate it, you can use this worksheet to calculate the composite score.

Name: _____   **Class:** _____

Beginning, Middle, and End of Year	
ORF Accuracy Percent	Accuracy Value
0% – 85%	0
86%	8
87%	16
88%	24
89%	32
90%	40
91%	48
92%	56
93%	64
94%	72
95%	80
96%	88
97%	96
98%	104
99%	112
100%	120

Beginning of Year Benchmark

ORF Words Correct = _____ [1]

Retell Score _____ **x 2** = _____ [2]

Maze Adjusted Score _____ **x 4** = _____ [3]

ORF Accuracy Percent: _____ %
100 x (Words Correct / (Words Correct + Errors))

Accuracy Value from Table = _____ [4]

Reading Composite Score (add values 1–4) = [_____]

If ORF is below 40 and Retell is not administered, use 0 for the Retell value only for calculating the Reading Composite Score. Do not calculate the composite score if any of the values are missing.

Middle of Year Benchmark

ORF Words Correct = _____ [1]

Retell Score _____ **x 2** = _____ [2]

Maze Adjusted Score _____ **x 4** = _____ [3]

ORF Accuracy Percent: _____%
100 x (Words Correct / (Words Correct + Errors))

Accuracy Value from Table = _____ [4]

Reading Composite Score (add values 1–4) = [_____]

If ORF is below 40 and Retell is not administered, use 0 for the Retell value only for calculating the Reading Composite Score. Do not calculate the composite score if any of the values are missing.

End of Year Benchmark

ORF Words Correct = _____ [1]

Retell Score _____ **x 2** = _____ [2]

Maze Adjusted Score _____ **x 4** = _____ [3]

ORF Accuracy Percent: _____ %
100 x (Words Correct / (Words Correct + Errors))

Accuracy Value from Table = _____ [4]

Reading Composite Score (add values 1–4) = [_____]

If ORF is below 40 and Retell is not administered, use 0 for the Retell value only for calculating the Reading Composite Score. Do not calculate the composite score if any of the values are missing.

5 Fifth Grade Reading Composite Score Worksheet
© Dynamic Measurement Group, Inc. / October 15, 2018

The Reading Composite Score is used to interpret student results for Acadience Reading. Most data management services will calculate the composite score for you. If you do not use a data management service or if your data management service does not calculate it, you can use this worksheet to calculate the composite score.

Name: _____   **Class:** _____

Beginning, Middle, and End of Year	
ORF Accuracy Percent	Accuracy Value
0% – 85%	0
86%	8
87%	16
88%	24
89%	32
90%	40
91%	48
92%	56
93%	64
94%	72
95%	80
96%	88
97%	96
98%	104
99%	112
100%	120

Beginning of Year Benchmark

ORF Words Correct **=** _____ [1]

Retell Score _____ **x 2** **=** _____ [2]

Maze Adjusted Score _____ **x 4** **=** _____ [3]

ORF Accuracy Percent: _____ %
100 x (Words Correct / (Words Correct + Errors))

Accuracy Value from Table **=** _____ [4]

Reading Composite Score (add values 1–4) = [_____]

If ORF is below 40 and Retell is not administered, use 0 for the Retell value only for calculating the Reading Composite Score. Do not calculate the composite score if any of the values are missing.

Middle of Year Benchmark

ORF Words Correct **=** _____ [1]

Retell Score _____ **x 2** **=** _____ [2]

Maze Adjusted Score _____ **x 4** **=** _____ [3]

ORF Accuracy Percent: _____ %
100 x (Words Correct / (Words Correct + Errors))

Accuracy Value from Table **=** _____ [4]

Reading Composite Score (add values 1–4) = [_____]

If ORF is below 40 and Retell is not administered, use 0 for the Retell value only for calculating the Reading Composite Score. Do not calculate the composite score if any of the values are missing.

End of Year Benchmark

ORF Words Correct **=** _____ [1]

Retell Score _____ **x 2** **=** _____ [2]

Maze Adjusted Score _____ **x 4** **=** _____ [3]

ORF Accuracy Percent: _____ %
100 x (Words Correct / (Words Correct + Errors))

Accuracy Value from Table **=** _____ [4]

Reading Composite Score (add values 1–4) = [_____]

If ORF is below 40 and Retell is not administered, use 0 for the Retell value only for calculating the Reading Composite Score. Do not calculate the composite score if any of the values are missing.

6 Sixth Grade Reading Composite Score Worksheet
© Dynamic Measurement Group, Inc. / October 15, 2018

The Reading Composite Score is used to interpret student results for Acadience Reading. Most data management services will calculate the composite score for you. If you do not use a data management service or if your data management service does not calculate it, you can use this worksheet to calculate the composite score.

Name: _____   **Class:** _____

Beginning, Middle, and End of Year	
ORF Accuracy Percent	Accuracy Value
0% – 85%	0
86%	8
87%	16
88%	24
89%	32
90%	40
91%	48
92%	56
93%	64
94%	72
95%	80
96%	88
97%	96
98%	104
99%	112
100%	120

Beginning of Year Benchmark

ORF Words Correct **=** _____ [1]

Retell Score _____ **x 2** **=** _____ [2]

Maze Adjusted Score _____ **x 4** **=** _____ [3]

ORF Accuracy Percent: _____%
100 x (Words Correct / (Words Correct + Errors))

Accuracy Value from Table **=** _____ [4]

Reading Composite Score (add values 1–4) = [_____]

If ORF is below 40 and Retell is not administered, use 0 for the Retell value only for calculating the Reading Composite Score. Do not calculate the composite score if any of the values are missing.

Middle of Year Benchmark

ORF Words Correct **=** _____ [1]

Retell Score _____ **x 2** **=** _____ [2]

Maze Adjusted Score _____ **x 4** **=** _____ [3]

ORF Accuracy Percent: _____%
100 x (Words Correct / (Words Correct + Errors))

Accuracy Value from Table **=** _____ [4]

Reading Composite Score (add values 1–4) = [_____]

If ORF is below 40 and Retell is not administered, use 0 for the Retell value only for calculating the Reading Composite Score. Do not calculate the composite score if any of the values are missing.

End of Year Benchmark

ORF Words Correct **=** _____ [1]

Retell Score _____ **x 2** **=** _____ [2]

Maze Adjusted Score _____ **x 4** **=** _____ [3]

ORF Accuracy Percent: _____%
100 x (Words Correct / (Words Correct + Errors))

Accuracy Value from Table **=** _____ [4]

Reading Composite Score (add values 1–4) = [_____]

If ORF is below 40 and Retell is not administered, use 0 for the Retell value only for calculating the Reading Composite Score. Do not calculate the composite score if any of the values are missing.

Appendix B: Pronunciation Guide

The Pronunciation Guide is a reference for giving and scoring the Acadience Reading measures. The phonemes and examples should guide how the sounds are spoken to students during the assessment, and also should guide how to score the measures. The sounds listed in this guide are shown in the initial, medial, and final position in words when possible. Multiple spellings (or the most common spellings) for each sound are shown. Different regions of the country use different dialects of American English. Any regional or dialectal pronunciation of the sound is acceptable.

Phoneme	Phoneme Example	Phoneme	Phoneme Example
/b/	**b**us, **b**aby, tu**b**	/TH/	**th**em, fea**th**er, brea**the**
/d/	**d**ig, **d**oll, la**dd**er, hi**d**	/ng/	wi**ng**, spinni**ng**, thi**n**k, ru**ng**
/f/	**f**ox, be**f**ore, lau**gh**, gra**ph**	/a/	**a**nt, t**a**p, h**a**t
/g/	**g**o, le**g**, so**gg**y, ho**g**	/e/	**e**cho, h**e**n, m**e**t
/h/	**h**im, a**h**ead	/i/	**i**s, s**i**t, b**i**g, w**i**th
/j/	**j**ar, le**dg**e, **j**ump, a**g**ile	/o/	**o**ff, s**a**w, d**o**t, w**a**ffle
/k/	**c**ap, **k**ite, ba**k**ing, e**ch**o, sta**ck**	/u/	**u**p, **a**llow, ab**o**ve, m**o**ther
/l/	**l**ap, **l**ight, ho**ll**ow, pu**ll**	/ai/	**a**ce, r**ai**l, m**a**de, h**ay**
/m/	**m**ess, **m**e, ha**mm**er, su**m**, a**m**	/ea/	**ea**t, fl**ee**t, sh**e**
/n/	**n**ot, di**nn**er, o**n**	/ie/	**i**ce, tr**i**ed, finall**y**, p**ie**, l**igh**t, fl**y**
/p/	**p**ie, a**pp**le, ho**p**	/oa/	**oa**k, s**oa**p, h**o**pe
/r/	**r**un, t**r**ee, **wr**ite, a**rr**ow	/oo/	b**oo**t, sh**oe**, val**ue**, neph**ew**
/s/	**s**ap, **c**ity, li**s**ten, ra**ce**	/uu/	w**oo**d, sh**ou**ld, p**u**t
/t/	**t**ot, ho**tt**er, ma**t**	/ow/	h**ou**se, c**ow**
/v/	**v**est, **v**ase, se**v**en, mo**v**e	/oy/	**oi**l, p**oi**nt, ch**oi**ce, t**oy**
/w/	**w**in, a**w**ay, **wh**eel, some**wh**ere	/ar/ (1 phoneme)	**ar**t, h**ear**t, st**ar**t
/y/	**y**es, on**i**on	/er/ (1 phoneme)	f**er**n, f**ir**st, l**ear**n, t**ur**n, g**ir**l
/z/	**z**ip, ea**s**y, i**s**	/or/ (1 phoneme)	s**or**t, bef**ore**
/ch/	**ch**icken, fu**t**ure, swi**tch**	/e/ /r/ (2 phonemes)	p**air**, sh**are**
/sh/	**sh**op, **sh**ow, mo**ti**on, hu**sh**	/i/ /r/ (2 phonemes)	h**ear**
/zh/	trea**s**ure, bei**ge**	/uu/ /r/ (2 phonemes)	t**our**, l**ure**
/th/	**th**ink, no**th**ing, sou**th**		

Note: For the intent and purpose of assessing beginning phonemic awareness skills in students in kindergarten and first grade, we do not distinguish between the /w/ sound in "win" and the /wh/ sound in "where" or between the /o/ sound in "hop" and the /aw/ sound in "saw."

Bibliography

Adams, M. J. (1990). *Beginning to read: Thinking and learning about print.* Cambridge, MA: MIT Press.

Badian, N.A. (1995). Predicting reading ability over the long term: The changing role of letter naming, phonological awareness and orthographic processing. *Annals of Dyslexia, 45,* 79–96. doi:10.1007/BF02648213

Bruck, M., Genesee, F., & Caravolas, M. (1997). A cross-linguistic study of early literacy acquisition. In B. Blachman (Ed.), *Foundations of reading acquisition and dyslexia: Implications for early intervention* (pp. 145–162). Mahwah, NJ: Lawrence Erlbaum Associates.

Buck, J. & Torgesen, J. K. (2003). *The Relationship Between Performance on a Measure of Oral Reading Fluency and Performance on the Florida Comprehensive Assessment Test.* Technical Report #1, Florida Center for Reading Research, Tallahassee, FL.

Cummings, K. D., Kaminski, R. A., Good, R. H., & O'Neil, M. E. (2011). Assessing phonemic awareness in preschool and kindergarten: Development and initial validation of First Sound Fluency. *Assessment for Effective Intervention, 36*(2), 94–106. doi:10.1177/1534508410392209

Daane, M.C., Campbell, J.R., Grigg, W.S., Goodman, M.J., & Oranje, A. (2005). *Fourth-Grade Students Reading Aloud: NAEP 2002 Special Study of Oral Reading* (NCES 2006–469). U.S. Department of Education. Institute of Education Sciences, National Center for Education Statistics. Washington, DC: Government Printing Office. Available http://nces.ed.gov/nationsreportcard/pdf/studies/2006469.pdf. Accessed 6/22/2010.

Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children, 52*(3), 219–232.

Deno, S. L. (1989). Curriculum-based measurement and special education services: A fundamental and direct relationship. In M. R. Shinn (Ed.), *Curriculum-based measurement: Assessing special children* (pp. 1–17). New York: Guilford Press.

Deno, S. L., & Mirkin, P. K. (1977). *Data-based program modification: A manual.* Reston, VA: Council for Exceptional Children.

Deno, S. L., & Fuchs, L. S. (1987). Developing curriculum-based measurement systems for data-based special education problem solving. *Focus on Exceptional Children, 19*(8), 1–15.

Dewey, E. N., Latimer, R. J., Kaminski, R. A., & Good, R. H. (2011). *DIBELS Next Development: Findings from Beta 2 Validation Study* (Tech. Report No. 10). Eugene, OR: Dynamic Measurement Group. Available: https://acadiencelearning.org/.

Dewitz, P., & Dewitz, P. K. (2003). They can read the words, but they can't understand: Refining comprehension assessment. *Reading Teacher, 56,* 422–435.

Ehri, L.C. (1983). A critique of five studies related to letter-name knowledge and learning to read. In L. Gentile, M. Kamil, & J. Blanchard (Eds.), *Reading research revisited* (pp. 143–153). Columbus, OH: C.E. Merrill.

Fuchs, D., & Fuchs, L.S. (2006). Introduction to responsiveness-to-intervention: What, why, and how valid is it? *Reading Research Quarterly, 41*, 92–99. doi:10.1598/RRQ.41.1.4

Fuchs, L. S., & Deno, S. L. (1991). Paradigmatic distinctions between instructionally relevant measurement models. *Exceptional Children, 57*(6), 488–500.

Good, R. H., & Kaminski, R. A. (1996). Assessment for instructional decisions: Toward a proactive/prevention model of decision making for early literacy skills. *School Psychology Quarterly, 11*, 326– 336. doi:10.1037/h0088938

Good, R. H., III, Kaminski, R. K., Cummings, K., Dufour-Martel, C., Petersen, K., Powell-Smith, K., Stollar, S., & Wallin, J. (2011). *Acadience Reading Assessment Manual.* Eugene, OR: Dynamic Measurement Group. Available: https://acadiencelearning.org/

Good, R. H., Simmons, D. C., & Smith, S. (1998). Effective academic interventions in the United States: Evaluating and enhancing the acquisition of early reading skills. *School Psychology Review, 27*, 45–56.

Gray, J. S., Warnock, A. N., Kaminski, R. A., & Good, R. H. (2018). *Acadience Reading National Norms 2014–2015* (Technical Report No. 23). Eugene, OR: Dynamic Measurement Group. Available: https://acadiencelearning.org/.

Haynes, S. N., Richard, D. R., & Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment, 7*, 238–247. doi:10.1037/1040-3590.7.3.238

Hopkins, W. G. (2002). A scale of magnitudes for the effect statistics. In *A review of statistics*. Retrieved Aug 11, 2010 from http://www.sportsci.org/resource/stats/effectmag.html

Hornby, A. S., Wehmeier, S., McIntosh, C., & Turnbull, J. (2005) *Oxford advanced learner's dictionary of current English.* London: Oxford University Press.

Hudson, R. F., Lane, H. B., & Pullen, P. C. (2005). Reading fluency assessment and instruction: What, why, and how? *The Reading Teacher, 58*(8), 702–714.

Individuals with Disabilities Education Improvement Act of 2004 (IDEA), Pub.L.No.108–446, 118 Stat. 2647 (2004), [Amending 20 U.S.C. § § 1400 et seq.].

Kame'enui, E. J., Carnine, D. W., Dixon, R. C., Simmons, D. C., & Coyne, M. D. (2002). *Effective teaching strategies that accommodate diverse learners* (2nd ed.). Upper Saddle River, NJ: Merrill Prentice Hall.

Kaminski, R. A. (1992). *Assessment for the primary prevention of early academic problems: Utility of curriculum-based measurement prereading tasks.* Unpublished doctoral dissertation. University of Oregon, Eugene.

Kaminski, R.A., Baker, S.K., Chard, D., Clarke, B., Smith, S. (2006). *Final report: Reliability, Validity, and Sensitivity of Houghton Mifflin Early Growth Indicators* (Technical Report). Eugene, OR: Dynamic Measurement Group and Pacific Institutes for Research.

Kaminski, R.A. & Cummings, K. D. (2007). Assessment for learning: Using general outcomes measures. *Threshold*, Winter, 2007, 26–28. Available http://ciconline.org/threshold.

Kaminski, R. A., Cummings, K. D., Powell-Smith, K. A., & Good, R. H. III (2008). Best practices in using Dynamic Indicators of Basic Early Literacy Skills (DIBELS®) for formative assessment and evaluation. In A. Thomas, & J. Grimes (Eds.), *Best practices in school psychology-V, (pp. 1181–1204).* Bethesda, MD: National Association of School Psychologists.

Kaminski, R. A., & Good, R. H. (1996). Toward a technology for assessing basic early literacy skills. *School Psychology Review, 25,* 215–227.

Kaminski, R. A., & Good, R. H., III. (1998). Assessing early literacy skills in a problem-solving model: Dynamic Indicators of Basic Early Literacy Skills. In M. R. Shinn (Ed.), *Advanced applications of curriculum-based measurement* (pp. 113–142). New York: Guilford Press.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychology Review, 95*, 492–527. doi:10.1037/0033-295X.95.4.492

Logan, G. D. (1997). Automaticity and reading: Perspectives from the instance theory of automatization. *Reading & Writing Quarterly: Overcoming Learning Difficulties, 13*, 123–46. doi:10.1080/1057356970130203

Mann, V.A., & Wimmer, H. (2002). Phoneme awareness and pathways into literacy: A comparison of German and American children. *Reading and Writing: An Interdisciplinary Journal, 15*, 653–682. doi:10.1023/A:1020984704781

Messick, S. (1993). Validity. In R. L. Linn (Ed.). *Educational measurement* (pp. 105–146). Phoenix, AZ: Oryx Press.

National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction.* Reports of the subgroups. Washington, DC: National Institute of Child Health and Human Development.

National Research Council. (1998). *Preventing reading difficulties in young children.* Washington, DC: National Academy Press.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.

Powell-Smith, K. A., Good, R. H., & Atkins, T. (2010). *DIBELS Next Oral Reading Fluency Readability Study* (Tech. Report No. 7). Eugene, OR: Dynamic Measurement Group. Available: https://acadiencelearning.org/.

Powell-Smith, K. A., Good, R. H., Latimer, R. J., Dewey, E. N., Kaminski, R. A., & Cummings, K. D. (2011). *DIBELS Next Benchmark Goals Study* (Tech. Report No. 11). Eugene, OR: Dynamic Measurement Group. Available: https://acadiencelearning.org/.

Rathvon, N., (2004). *Early reading assessment: A practitioner's handbook.* New York: Guilford.

Salvia, J., Ysseldyke, J., & Bolt, S. (2007). *Assessment in special and inclusive education* (Tenth Edition). Boston: Houghton Mifflin Company.

Shanahan, T. (2005). Review of the DIBELS: Dynamic Indicators of Basic Early Literacy Skills, Sixth Edition. In R. A. Spies, B. S. Plake, & L. L. Murphy (Eds.), *The sixteenth mental measurements yearbook (pp. 308–310).* Lincoln, NE: The Buros Institute of Mental Measurements.

Shinn, M. R. (1995). Best practices in using curriculum-based measurement in a problem-solving model. In J. G. A. Thomas (Ed.), *Best practices in school psychology III* (Vol. 3, pp. 671–697). Silver Springs, MD: National Association of School Psychologists.

Shuttleworth, Martyn (2009). Convergent Validity and Discriminant Validity. Retrieved Nov 16, 2010 from Experiment Resources: http://www.experiment-resources.com/convergent-validity.html

Simmons, D. C., & Kame'enui, E. J. (Eds.). (1998). *What reading research tells us about children with diverse learning needs: Bases and basics.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Tilly, W. D., III, (2008). The evolution of school psychology to science-based practice. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology V* (pp. 18–32). Washington, DC: National Association of School Psychologists.

Torgesen, J. K., Wagner, R. K., Rashotte, C.A., Rose, E., Lindamood, P., Conway, T. , & Garvin, C. (1999). Preventing reading failure in young children with phonological processing disabilities: Group and individual responses to instruction. *Journal of Educational Psychology, 91*, 579–593. doi:10.1037/0022- 0663.91.4.579

U.S. Dept. of Education, National Center for Education Statistics, CCD (2006–07). Demographic Information for participating school district (Study D). Retrieved October 19, 2009, from http://nces.ed.gov/.

U.S. Dept. of Education, National Center for Education Statistics, CCD (2007–08). Demographic Information for participating school districts (Studies A and B). Retrieved from http://nces.ed.gov/.

U.S. Dept. of Education, National Center for Education Statistics, CCD (2008-09) & PSS (2007–08). Demographic Information for participating school districts (Study C). Retrieved February 7, 2011, from http://nces.ed.gov/.

Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1999). Comprehensive test of phonological processing. Austin, TX: PRO-ED.

Walsh, D.J., Price, G.G., & Gillingham, M.G. (1988). The critical but transitory importance of letter naming. *Reading Research Quarterly, 23*, 108–122. doi:10.2307/747907

Williams, K. T. (2001). Group Reading and Diagnostic Evaluation (GRADE). New York: Pearson.

Wilson, J. (2005). *The relationship of Dynamic Indicators of Basic Early Literacy Skills (DIBELS) Oral Reading Fluency to performance on Arizona Instrument to Measure Standards (AIMS).* Tempe, AZ: Tempe School District No. 3. Available: https://dibels.uoregon.edu/research/techreports/

Yopp, H. K. (1988). The validity and reliability of phonemic awareness tests. *Reading Research Quarterly, 23*(2), 159–177. doi:10.2307/747800

Zeno, S.M., Ivens, S.H., Millard, R.T., Duvvuri, R. (1995). *The educator's word frequency guide.* Brewster, NY: Touchstone Applied Science Associates.