Measuring What Matters: Understanding What Reading Assessments Really Tell Us

by Jessica R. Toste, Daniel Espinas, Eric Oslund, Amy Elleman, and Gina Biancarosa

Teaching is a bit like navigation—continually adjusting course based on student progress, needs, and an evolving classroom context. Every instructional decision, at its core, is a hypothesis about how students learn, and assessment data provide the evidence needed to test and refine those hypotheses. These data serve as the map and compass of reading instruction, helping teachers understand where students are in their reading development, evaluate the impact of teaching, and determine the best direction for continued learning. Yet, the term "data" encompasses many different kinds of information collected for a variety of purposes. Some assessments capture broad patterns across schools or districts, while others focus narrowly on specific reading skills. Recognizing these distinctions is essential for using data meaningfully rather than mechanically.

Federal legislation, including the Every Student Succeeds Act (2015)—the most recent reauthorization of the Elementary and Secondary Education Act—and the Individuals with Disabilities Education Act (IDEA, 2004), underscores the central role of assessment data in educational decision-making. These laws require schools to systematically collect, analyze, and use evidence of student performance to guide instruction, monitor progress, and promote equitable educational outcomes. ESSA continues the federal emphasis on accountability and continuous improvement across all students and subgroups, while IDEA ensures that assessment data inform individualized educational planning and the evaluation of special education services. However, compliance alone does not guarantee effective data use.

The most productive use of assessment begins with asking "What information do I need, and why?" As Hamilton and colleagues (2009) describe in the cycle of instructional improvement, data use is an iterative process of collecting evidence, interpreting patterns, generating hypotheses, and adjusting instruction in response. Because no single assessment can answer every question about student learning, teachers must integrate data from multiple sources to gain a full picture of reading development (Mandinach & Gummer, 2016). Two overarching questions guide this paper:

1. What are the different types of reading assessment data commonly available in schools?
2. How does each type inform instructional or other decision-making?

To answer these questions, we describe five key sources of reading assessment data: summative state assessments, universal screening, curriculum-embedded assessments, progress monitoring, and diagnostic assessments. Each source serves a unique function within a school's assessment system, varying in purpose, frequency, scope, and the types of instructional decisions it supports. Collectively, these sources form a coherent framework for understanding students' reading performance across time and contexts.

The Concept of Grain Size

A useful way to think about assessment data is through the concept of grain size—the level of specificity or detail captured by a measure (Coburn & Turner, 2011). Grain size reflects the *scale* at which a question can be answered. Large-grained assessments (e.g., annual state tests) yield broad insights about overall achievement or program effectiveness. In contrast, fine-grained assessments (e.g., progress monitoring or diagnostic tools) provide detailed evidence about individual skills or instructional targets.

The grain size of an assessment determines the types of instructional or adminis-

trative decisions it can support. Large-grained data may reveal that a school's reading scores have improved overall, but they cannot identify *why* or *for whom* improvement occurred. Conversely, fine-grained data can show that a student has mastered consonant–vowel–consonant (CVC) decoding patterns but still struggles with multisyllabic words—information that directly informs instructional planning. Understanding grain size helps educators align their questions with the right type of data. For example, if a teacher wants to know whether a student's reading fluency is improving in response to intervention, a measure of oral reading rate (fine-grained) is appropriate. If a district leader wants to evaluate whether a new reading program is improving outcomes system-wide, end-of-year summative assessments (large-grained) provide the relevant evidence.

Importantly, no single grain size—or single assessment—can meet all data needs. A comprehensive reading assessment system incorporates multiple sources that vary in scope and purpose. When viewed together, these data allow educators to monitor student growth, evaluate program effectiveness, and make informed instructional adjustments at multiple levels (Hamilton et al., 2009; Mandinach & Gummer, 2016).

Finally, when progress data indicate that a student is not responding as expected, diagnostic assessment serves as a process of hypothesis generation and testing—collecting additional evidence to determine why learning needs are not being met and how instruction should be individualized (Deno, 2003; Fuchs & Fuchs, 2002).

In the sections that follow, we examine the five major sources of reading assessment data presented in Table 1. For each, we describe its primary purpose, the questions it can answer, and how it contributes to the cycle of instructional improvement. Taken together, these data sources demonstrate how assessment, when used intentionally and strategically, can lead to improved reading outcomes for all students.

Summative State Assessments

State assessments serve as summative measures of academic performance and are administered once annually, typically in the spring. These standardized tests are designed to evaluate student achievement against state academic standards and to provide data for system-level accountability. Although often referred to as "high stakes" because results may influence district ratings, teacher evaluations, or funding, their central purpose is programmatic rather than diagnostic. Most states administer these assessments electronically and rely on scaled scores and proficiency levels to gauge whether students meet grade-level expectations. They are commonly referred to as statewide summative or accountability assessments, which emphasize data-driven accountability (Hamilton et al., 2002).

Administered once per year in Grades 3 to 8 and once in high school, state assessments generate data used for accountability reporting, longitudinal tracking, and resource planning. Because of their scale, they provide consistent metrics for evaluating performance across years and populations. These data inform system-level evaluation and planning by addressing questions such as the following:

- How are students performing relative to state standards?
- Where do achievement gaps exist across schools or student subgroups?
- Which curricular programs, initiatives, or schedules are associated with higher/lower outcomes at scale?
- Where should we target resources (e.g., coaching, intervention staff, materials) to address persistent weaknesses?

It is equally important to recognize what summative assessment data are *not* designed to do. Summative assessment data are not most appropriate for informing day-to-day instruction or evaluating short-cycle interventions, as these assessments are administered infrequently and are not sensitive to small, near-term changes in learning. Likewise, because these assessments provide only broad indicators of achievement, they should not be used for diagnostic or placement decisions at the individual student level.

Research demonstrates that well-designed state assessments yield reliable and valid measures of achievement at the system level. Research examining large samples of students has found that these assessments are highly reliable and measure what they are intended to measure, especially when aligned with grade-level standards (Hamilton et al., 2002). These assessments are useful for program evaluation and monitoring equity across student groups.

Case Example

After reviewing state assessment results, a district curriculum director noticed that fewer than half of the students in Grades 4 to 8 met proficiency benchmarks in reading. The data revealed that certain schools and student sub-

Table 1

Sources of Reading Assessment Data Used in Schools

Data Source	Description	Purpose	Other Common Names
Summative State Assessments	· Annual, comprehensive assessment of curriculum across an entire year · Capture achievement levels of student groups in broad content areas	· Determine if students meet grade-level benchmarks · Evaluate effectiveness of schoolwide curriculum and instruction · Inform accountability and equity monitoring at district or state level	· High-stakes tests · Accountability assessments · Statewide assessments
Universal Screening	· Typically administered three times per year (beginning, middle, end) · Captures broad academic performance and progress across students	· Evaluate effectiveness of the school's core curriculum, environment, and instruction · Identify students not making expected progress who may need additional intervention	· Benchmark assessments · Interim assessments
Curriculum-Embedded Assessments	· Cover course-specific content directly tied to curriculum · Administered after exposure to targeted material (e.g., chapters, lessons, units)	· Evaluate students' understanding of explicitly taught content · Inform grading and classroom instruction	· Unit (end-of-unit) tests · Student assignments or independent work · End-of-semester grades
Progress Monitoring	· Short, frequent probes of targeted academic skills · Administered as part of instruction or intervention	· Assess students' performance relative to long-term learning goals · Evaluate responsiveness to instruction and need for instructional adjustment	· Curriculum-based measurement · General outcome measures · Mastery measurement · Progress probes
Diagnostic Assessments	· Administered when progress monitoring data show limited growth or when a student's learning profile is unclear · Provide detailed information about specific skill strengths and weaknesses	· Determine underlying reasons for academic difficulties · Inform individualized intervention planning	· Diagnostic testing · Error analysis · Phonics inventory

groups exhibited particularly low performance, suggesting that current curricula or instructional approaches were not meeting the needs of all learners. In response, the district allocated resources to strengthen reading intervention services at the start of the school year, ensuring that students performing below proficiency receive additional supports.

Universal Screening

Universal screening or benchmark assessments are designed to identify students who may be at risk for current or future reading difficulties. Within school-based tiered prevention systems, such as multi-tiered systems of support (MTSS) or response to intervention (RTI), screening data serve as an early-warning indicator that helps ensure timely intervention. Screeners are brief, standardized, and efficient—often taking less than ten minutes per student. They can be criterion-referenced (e.g., mastery of specific letter–sound correspondences) or norm-referenced (e.g., percentile ranks compared to peers). Common formats include curriculum-based measurement (CBM), such as oral reading fluency, and com-

Screeners are brief, standardized, and efficient—often taking less than ten minutes per student.

puter-adaptive testing (CAT). Because screening results can influence access to support, their technical adequacy is critical. High-quality screeners demonstrate strong reliability, validity, and classification accuracy (Petscher & Suhr, 2022). The National Center on Intensive Intervention (NCII; intensiveintervention.org) publishes independent evaluations of many commercial tools.

Universal screening typically occurs two to three times per year—most commonly in the fall, winter, and spring. These repeated administrations enable educators to evaluate whether instructional changes have an impact on risk status over time and ensure ongoing responsiveness to student needs. Screening data address questions such as the following:

- Which students are meeting, at risk of not meeting, or not meeting grade-level reading benchmarks?
- Is the core (Tier 1) reading instruction effective for the majority (e.g., ≥80%) of students?
- Who may require additional Tier 2 or Tier 3 intervention?
- How can screening data guide grouping decisions or instructional focus within intervention blocks?
- How do screening outcomes compare to prior years or other schools in the district?

Although screening assessments provide essential information to guide prevention and early intervention, they are *not* intended to serve as diagnostic tools or to determine special education eligibility. Screening data are used to identify *who* may be at risk, but not *why* a student is struggling or which specific skills require attention. Similarly, they are less useful for evaluating the effects of short-term interventions or for making fine-grained instructional adjustments. When used as part of a multi-tiered system, screening results guide timely follow-up assessments and targeted instruction; however, they should not be interpreted in isolation.

A robust research base supports the technical adequacy of universal screening for identifying reading difficulties. Research shows that different versions of these reading measures give consistent results and do a good job of predicting students' later reading success (Schatschneider et al., 2008). Meta-analytic reviews confirm that oral reading fluency scores are strong indicators of which students are likely to experience reading difficulties (Reschly et al., 2009). Overall, screening has strong empirical support as a reliable, valid, and efficient means of identifying risk and allocating instructional resources.

Case Example

All first grade students in a school completed the winter oral reading fluency (ORF) screening. Seventeen of the twenty students in one classroom meet benchmark expectations, indicating that Tier 1 instruction is effective. Three students scored below the 25th percentile, prompting the teacher to form a small decoding intervention group for six weeks. In another classroom, however, half of the students scored below benchmark, leading the MTSS team to review and strengthen Tier 1 instruction in that classroom.

Curriculum-Embedded Assessments

Curriculum-embedded assessments are integrated directly within classroom instruction to provide immediate feedback about students' mastery of recently taught content. They are formative in nature and used to monitor understanding, guide reteaching, and inform short-term instructional decisions. Unlike external standardized tests, curriculum-embedded assessment activities are derived directly from instructional materials (e.g., weekly spelling checks, lesson-end reading tasks, comprehension questions tied to class texts). They are brief, teacher-administered, and scored immediately, providing actionable classroom data. Curriculum-embedded assessment data is collected frequently—daily, weekly, or at the end of instructional units—depending on curric-

Unlike external standardized tests, curriculum-embedded assessment activities are derived directly from instructional materials (e.g., weekly spelling checks, lesson-end reading tasks, comprehension questions tied to class texts). They are brief, teacher-administered, and scored immediately, providing actionable classroom data.

ulum design and pacing of instruction. These data help teachers answer questions such as the following:

- To what extent did students master the specific concept, skill, or text that was taught in this lesson or unit?
- Is the current instructional approach, pacing, or grouping effective for most students?
- Does student performance suggest that the instructional sequence or materials should be adjusted (e.g., reordering lessons to better build prerequisite knowledge)?
- For students who easily meet expectations, what opportunities exist to extend or enrich instruction to maintain engagement and growth?

While curriculum-embedded assessments provide valuable, real-time feedback to guide instruction, they are *not* designed to measure long-term growth or broader patterns of achievement. Instead, they capture learning tied closely to specific lessons or units and therefore cannot substitute for more comprehensive measures of reading development. Because they are developed within a particular curriculum, results may not generalize beyond that instructional context or allow for valid comparisons across classrooms or schools. In addition, curriculum-embedded assessments may lack the technical rigor of standardized measures, making them less suitable for program evaluation or high-stakes decision-making.

Empirical evidence supports the use of curriculum-embedded tests and other formative assessments to enhance teaching and learning. Meta-analyses demonstrate that formative assessment practices yield moderate to large positive effects on student achievement when teachers use results to adapt instruction (Black & Wiliam, 1998; Fuchs & Fuchs, 1986). Research also emphasizes that alignment with instructional content is critical to reliability and validity—curriculum-embedded assessments function best when they measure skills and knowledge explicitly taught (Heritage, 2010). In classrooms where such assessments are used consistently, teachers report higher instructional responsiveness and improved student engagement.

Case Example

As part of a second grade unit on long vowel spelling patterns, a teacher administered a brief skills check, during which students read and spelled words containing "a_e," "ai," and "ay" patterns. Most students read and spelled "a_e" words such as "made" and "gate" accurately, but many confused "ai" and "ay" in final positions. Based on these results, the teacher decided to provide additional practice distinguishing "ai" and "ay" before introducing the next set of vowel patterns. By using curriculum-embedded assessment data to adjust pacing and focus, she reinforced mastery and prevented cumulative errors in later lessons.

Meta-analyses demonstrate that formative assessment practices yield moderate to large positive effects on student achievement when teachers use results to adapt instruction.

Progress Monitoring

Progress monitoring involves the frequent repeated measurement of student performance to evaluate whether instruction or intervention is producing desired growth. These assessments are sensitive to small changes over time, helping educators determine whether instructional adjustments are needed. Two primary approaches are used for this purpose. *Mastery measures* focus on tracking discrete skills that students are learning (e.g., pronouncing short vowel sounds or reading affixes). *General outcome measures* (GOMs), in contrast, assess overall reading proficiency by sampling performance from a broad skill domain (e.g., the number of words read correctly per minute). While both approaches can be used for progress monitoring, GOMs are particularly valuable because they reflect cumulative learning and can be used to compare student growth over time and across contexts. According to Fuchs and Deno (1991), GOMs combine the best parts of formal testing and classroom observation, giving teachers a practical way to track learning, adjust instruction, and understand how a student's progress compares to others.

In practice, progress monitoring is most often implemented through curriculum-based measurement (CBM), a well-established GOM approach that uses brief, standardized tasks such as oral reading fluency or maze comprehension (Deno, 2003; Fuchs, 2004). Each administration typically takes only a few minutes and yields data that can be plotted to evaluate the student's rate of improvement against expected benchmarks. Progress monitoring is typically conducted on a weekly or biweekly

basis for students receiving targeted intervention (Tier 2 or Tier 3). Progress monitoring data support instructional decisions by addressing questions such as the following:

- Is the current intervention or instructional approach producing meaningful growth in student performance over time?
- Is the student's rate of improvement sufficient to meet grade-level benchmarks or individualized goals within the expected timeframe?
- Do results suggest a need to intensify, modify, or maintain the current intervention (e.g., adjust duration, group size, or instructional focus)?
- Are patterns across students or groups indicating that the intervention program or instructional materials should be revised or replaced?

Although progress monitoring measures are highly sensitive to short-term growth and instructional responsiveness, they are not designed to provide a comprehensive picture of overall achievement or to assess mastery of specific content standards. These measures typically rely on brief, general outcome indicators (e.g., oral reading fluency, assessed as the number of words read correctly per minute) that reflect progress within a skill domain rather than detailed profiles of strengths and weaknesses. As a result, progress monitoring data should not be used to diagnose the source of reading difficulties, evaluate broad program effectiveness, or make high-stakes placement decisions. Their strength lies in tracking change over time and informing ongoing instructional adjustments—not in replacing more comprehensive assessments.

Extensive empirical evidence supports the reliability, validity, and instructional utility of progress monitoring measures, particularly CBM. Research demonstrates that CBM oral reading fluency and maze comprehension tasks show strong alternate-form reliability ($r > .90$) and predictive validity for later reading outcomes (Fuchs et al., 2001; Wayman et al., 2007). Teachers who use CBM data to guide instruction see greater student achievement gains than those who do not (Stecker et al., 2005). Progress monitoring thus serves as a core evidence-based practice for data-based individualization and continuous instructional improvement.

Case Example

A second grade student participating in a small-group decoding intervention completed a weekly oral reading fluency probe. After six weeks, progress monitoring data showed minimal growth relative to the expected benchmark slope. In response, the teacher added an additional session per week with increased emphasis on blending and automaticity. Following an additional four weeks, the student's fluency trajectory aligned with grade-level expectations, suggesting that the instructional adjustment was effective.

Diagnostic Assessments

Diagnostic assessments provide in-depth information about a student's specific reading strengths and weaknesses and are used to individualize instruction. They are typically administered when progress monitoring data indicate inadequate response to intervention or when a student's learning profile is unclear. Diagnostic tools may assess discrete academic skills (e.g., decoding, vocabulary, comprehension) as well as related behavioral, motivational, or strategic factors that influence reading (Truckenmiller & Toste, 2025). These measures are usually administered individually and yield detailed information about skill mastery and underlying processes. Diagnostic data inform decisions such as the following:

- Which specific component reading skills or underlying processes are contributing to the student's reading difficulties?
- What hypotheses can be generated about a student's learning needs?
- How should instruction be tailored or intensified to meet those needs?
- Are additional assessments (e.g., cognitive, language, behavioral) warranted?

Diagnostic assessments are designed to provide detailed, individualized information about student learning, but they are not appropriate for frequent use or for evaluating overall program effectiveness. Because they emphasize depth rather than breadth, these measures are time-intensive and impractical for large-scale administration. Their usefulness depends on skilled interpretation and clear instructional hypotheses; when used in isolation, they risk overassessment or misinterpretation. Diagnostic data are most effective when integrated into a comprehensive assessment system to inform individualized planning.

Diagnostic assessments are used as needed rather than on a fixed schedule, often following screening or progress monitoring when a student fails to make expected gains.

A solid research base supports the technical adequacy and instructional value of diagnostic assessments in reading. Well-designed diagnostic assessments have been shown to

Each assessment type contributes unique information: summative data evaluate program impact, screening data identify risk, curriculum-embedded tests inform immediate instructional choices, progress monitoring tracks growth, and diagnostic assessments guide individualization. Together, they transform isolated data points into a dynamic feedback system for teaching and learning.

produce consistent results and accurately measure key aspects of reading, including decoding, fluency, and comprehension (Torgesen, 2005). Studies also highlight how different comprehension tests measure different underlying skills, reinforcing the need to match assessment purpose to instructional questions. Implementation studies within data-based individualization frameworks demonstrate that diagnostic assessment enhances the precision of interventions and student outcomes (Fuchs et al., 2021). Additionally, research on reading difficulty profiles illustrates how understanding individual patterns of word-level and language comprehension deficits can guide differentiated instruction (Nation, 2019; Spear-Swerling, 2019).

Case Example

A fourth-grade student receiving Tier 3 reading intervention showed limited progress despite several months of intensive support. The reading specialist administered diagnostic assessments of decoding and morphological awareness, which revealed difficulty with reading multisyllabic words. Instruction was adjusted to include explicit teaching of affixes and a "peel-off" word reading strategy for analyzing longer words. Six weeks later, progress monitoring data showed clear improvement in reading fluency and accuracy, confirming that the targeted instructional change addressed the underlying skill deficit.

Integrating Assessment Data for Instructional Decision-Making

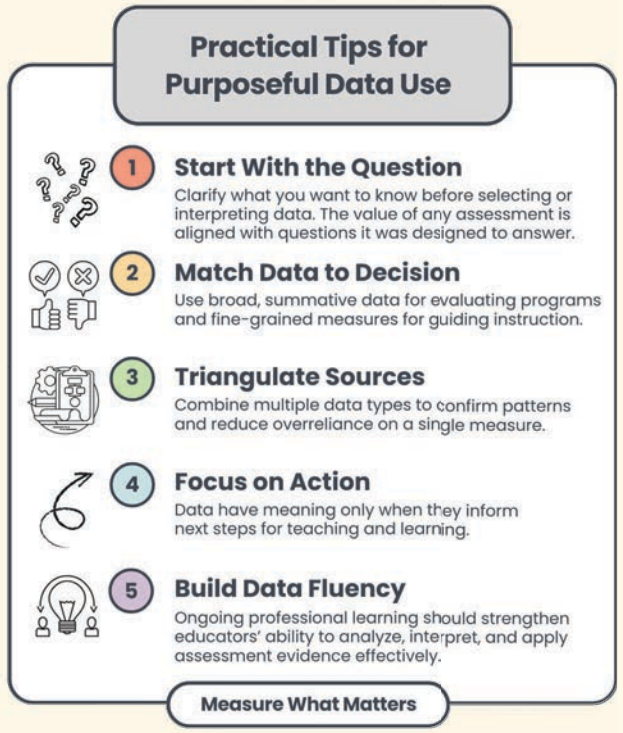
Although each assessment type serves a distinct purpose, their true value lies in how they work together to form a coherent system of instructional decision-making. No single data source—whether a statewide test or a curriculum-embedded measure—can provide the full picture of a student's reading development. When used in combination, however, these tools provide information at multiple levels of grain size, from broad indicators of system effectiveness to fine-grained evidence of skill mastery. This integration allows educators to link policy-level accountability, schoolwide planning, and day-to-day instruction within a continuous cycle of improvement.

The cycle of instructional improvement (Hamilton et al., 2009) offers a useful frame for understanding this process. Educators collect data from multiple sources, interpret data to identify student strengths and needs, generate hypotheses about how to improve outcomes, and adjust instruction accordingly. Subsequent data collection then provides evidence about whether those adjustments were effective. Within this continuous cycle, each assessment type contributes unique information: summative data evaluate program impact, screening data identify risk, curriculum-embedded tests inform immediate instructional choices, progress monitoring tracks growth, and diagnostic assessments guide individualization. Together, they transform isolated data points into a dynamic feedback system for teaching and learning.

Of course, this system's effectiveness depends not merely on the presence of data but on educators' ability to interpret and act on them appropriately. Technical adequacy—reliability, validity, and classification accuracy—matters, but so does teacher expertise. The best assessments can be misused when their purpose is misunderstood or when data are applied outside their intended scope. The most meaningful data are those that are collected intentionally to answer specific instructional questions (see Figure 1). Even high-quality assessments lose value when used to address the wrong question or decision. Professional learning that develops data literacy, particularly in aligning data use with instructional questions, is therefore essential. Evidence-informed decision-making is a skill in itself, requiring both statistical understanding and pedagogical reasoning.

Finally, it is worth reconsidering what we mean by *evidence* in reading instruction. Evidence is not limited to test scores; it encompasses the multiple, intersecting indicators that reveal how students are progressing. Data have different grain sizes—some offering broad trends, others providing finely detailed

Figure 1



insights—and each is valuable when used for the right purpose. Ultimately, the usefulness of any given assessment depends on whether it was designed and administered to answer the questions of interest. When educators view assessment data not as mandates to be complied with but as mirrors reflecting student learning, data become instruments of inquiry and improvement. In this way, assessment is transformed from an external requirement into an integral part of responsive, evidence-informed teaching. ■

Author Note

The authors would like to acknowledge the Lorentz Center at Leiden University (Leiden, The Netherlands) for sponsoring the international scientific meeting that led to this collaborative work.

References

Black, P., & Wiliam, D. (1998). *Assessment and classroom learning. Assessment in Education: Principles, Policy, & Practice, 5*(1), 7-74. https://doi.org/10.1080/0969595980050102

Coburn, C. E., & Turner, E. O. (2011). Research on data use: A framework and analysis. *Measurement: Interdisciplinary Research and Perspectives, 9*(4), 173-206. https://doi.org/10.1080/15366367.2011.626729

Deno, S. L. (2003). Curriculum-based measures: Development and perspectives. *Assessment for Effective Intervention, 28*(3–4), 3–12. https://doi.org/10.1177/073724770302800302

Every Student Succeeds Act, 20 U.S.C. § 6301 (2015).

Fuchs, L. S. (2004). The past, present, and future of curriculum-based measurement research. *School Psychology Review, 33*(2), 188-192. https://doi.org/10.1080/02796015.2004.12086241

Fuchs, L. S., & Deno, S. L. (1991). Paradigmatic distinctions between instructionally relevant measurement models. *Exceptional Children, 57*(6), 488-500. https://doi.org/10.1177/001440299105700603

Fuchs, L. S., & Fuchs, D. (2002). What is scientifically-based research on progress monitoring? *Washington, DC: National Center on Student Progress Monitoring.*

Fuchs, L. S., & Fuchs, D. (1986). Effects of formative evaluation: A meta-analysis. *Exceptional Children, 53*(3), 199-208. https://doi.org/10.1177/001440298605300301

Fuchs, L. S., Fuchs, D., Hamlett, C. L., & Stecker, P. M. (2021). Bringing data-based individualization to scale: A call for the next-generation technology of teacher supports. *Journal of Learning Disabilities, 54*(5), 319-333. https://doi.org/10.1177/0022219420950654

Hamilton, L. S., Halverson, R., Jackson, S., Mandinach, E., Supovitz, J., & Wayman, J. (2009). *Using student achievement data to support instructional decision making* (NCEE 2009-4067). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://ies.ed.gov/ncee/wwc/publications/practiceguides/

Hamilton, L. S., Stecher, B. M., & Klein, S. P. (Eds.). (2002). *Making sense of test-based accountability in education.* RAND.

Heritage, M. (2010). *Formative assessment and next-generation assessment systems: Are we losing an opportunity?* National Center for Research on Evaluation, Standards, and Student Testing (CRESST) and the Council of Chief State School Officers (CCSSO).

Individuals with Disabilities Education Act, 20 U.S.C. § 1400 (2004).

Mandinach, E. B., & Gummer, E. S. (2016). What does it mean for teachers to be data literate: Laying out the skills, knowledge, and dispositions. *Teaching and Teacher Education, 60*, 366-376. https://doi.org/10.1016/j.tate.2016.07.011

Nation, K. (2019). Children's reading difficulties, language, and reflections on the simple view of reading. *Australian Journal of Learning Difficulties, 24*(1), 47-73. https://doi.org/10.1080/19404158.2019.1609272

Petscher, Y., & Suhr, M. (2022). Considerations for choosing and using screeners for students with disabilities. In C. J. Lemons, S. R. Powell, K. L. Lane, & T. C. Aceves (Eds.), *Handbook of special education research, Volume II: Research-based practices in and intervention innovations* (pp. 83-96). Routledge.

Reschly, A. L., Busch, T. W., Betts, J., Deno, S. L., & Long, J. D. (2009). Curriculum-based measurement oral reading as an indicator of reaching achievement: A meta-analysis of the correlational evidence. *Journal of School Psychology, 47*(6), 427-469. https://doi.org/10.1016/j.jsp.2009.07.001

Schatschneider, C., Wagner, R. K., & Crawford, E. C. (2008). The importance of measuring growth in response to intervention models: Testing a core assumption. *Learning and Individual Differences, 18*(3), 308-315. https://doi.org/10.1016/j.lindif.2008.04.005

Spear-Swerling, L. (2019). Structured literacy and typical literacy practices: Understanding differences to create instructional opportunities. *TEACHING Exceptional Children, 51*(3), 201-211. https://doi.org/10.1177/0040059917750160

Stecker, P. M., Fuchs, L. S., & Fuchs, D. (2005). Using curriculum-based measurement to improve student achievement: Review of research. *Psychology in the Schools, 42*(8), 795-819. https://doi.org/10.1002/pits.20113

Torgesen, J. K. (2005). *Assessment to prevent reading failure: A response to intervention approach*. In S. R. Jimerson, M. K. Burns, & A. M. VanDerHeyden (Eds.), *Handbook of Response to Intervention: The Science and Practice of Assessment and Intervention* (pp. 15-25). Springer.

Truckenmiller, A. J., & Toste, J. R. (2025). Using fine-grained literacy analysis to guide decision-making for instructional intensification. *Intervention in School and Clinic, 61*(1), 3-7. https://doi.org/10.1177/10534512251345798

Wayman, M. M., Wallace, T., Wiley, H. I., Tichá, & Espin, C. A. (2007). Literature synthesis on curriculum-based measurement in reading. *The Journal of Special Education, 41*(2), https://doi.org/10.1177/00224669070410020401

Jessica Toste

Jessica R. Toste, Ph.D., is an associate professor in the Department of Special Education and Distinguished Teaching Professor at The University of Texas at Austin. She holds research affiliations with the Meadows Center for Preventing Educational Risk and Center for Research to Community Impact. Her research is focused on methods for intensifying intervention for students with persistent reading challenges and reading disabilities.

Daniel Espinas

Dr. Daniel R. Espinas, Ph.D., is an assistant professor in the Division of Special Education in the College of Education and Human Development at George Mason University. His research focuses on academic development and interventions for children and adolescents with disabilities.

Eric Oslund

Eric Oslund, Ph.D., is a professor and chair of the Department of Elementary and Special Education at Middle Tennessee State University. Dr. Oslund's research centers on employing advanced applied statistics to explore various facets of education, including early reading measures, adolescent comprehension theories, and nationwide data-based decision-making.

Amy Elleman

Amy Elleman, Ph.D., earned a B.S. in psychology at the University of Maryland in Heidelberg, Germany and a Ph.D. at Vanderbilt University. Dr. Elleman's research focuses on the examination of factors central to reading comprehension, including vocabulary, inference generation, and knowledge acquisition. She has expertise in meta-analysis and conducting reading comprehension interventions.

Gina Biancarosa

Gina Biancarosa, Ph.D., is the Ann Swindells Chair in Education and a full professor in the Department of Special Education and Clinical Sciences in the College of Education at the University of Oregon. Biancarosa's research centers on questions related to more precise and informative measurement of reading, especially of reading comprehension, and modeling of growth in reading over time.