K–2 Assessment Systems Enable Early Intervention to Foster Student Success

Joanne L. Jensen, WestEd
Jessica Goldstein, Boulevard Research Partners LLC
Matthew A. Brunetti, WestEd

Across the United States, statewide assessments in English language arts and mathematics are federally mandated each school year in grades 3 through 8 and once in high school. The intent is to help educators, policymakers, and parents directly gauge how students and their school systems are performing against state standards. Missing nationwide, however, is any systematic state-level attempt to evaluate students' ongoing progress in grades K–2, the grades that lay the foundation for all later learning.

As noted in the companion paper in this series, as many as 41 states (Center for Standards, Assessment, and Accountability, 2021; Weisenfeld et al., 2020) assess incoming kindergarteners to understand how prepared each child is to participate in kindergarten curricula — a critical development, since research shows that, without effective intervention, performance gaps among students as they enter kindergarten persist into third grade (Duncan et al., 2007; Connor et al., 2011; Neuman & Dickinson, 2001).

However, for districts and schools to effectively intervene to change the trajectories for these young students, ongoing evaluation of students' progress throughout kindergarten, first, and second grades is essential. And although many states provide different kinds of early grade assessments, no state currently has a multidimensional statewide K–2 system. When and how to use available K–2 assessments is largely left up to local jurisdictions, with mixed results in terms of teachers' ability to effectively identify and attend to students' learning needs.

This paper focuses on this lag in K–2 assessment systems and how states can act to address it. We first review a range of assessment types and their utility for supporting learning in the early years of schooling. We then discuss research findings on state K–2 assessment policies that provide insights for other policy leaders to consider as they work to build K–2 assessment systems that effectively help districts and schools support academic success for their youngest students.

ABOUT THIS BRIEF

Many children start school needing extra support to thrive academically in grades K–2 — the foundation for success as they move up the grades. This paper discusses designing early grade assessment systems that enable educators to intervene throughout the K–2 years to help students achieve success. A companion paper explains how states can lay the groundwork for addressing readiness gaps by identifying, at kindergarten entry, those children who may need extra support to thrive in the early grades.

WestEd
WestEd.org

Types of Assessments and Their Use for K–2

In a 2018 summary of recent state policies on early grade assessment, the Council of Chief State School Officers (CCSSO) delineated five types of early grade assessments: summative, interim, screener, diagnostic, and formative (2019). The paper reported that 35 states offered one or more K–2 assessments, with the majority requiring that all students be assessed. Some of these states, however, offer optional assessments or encourage districts to use some form of assessment to monitor student progress. Only two of the 35 states provided summative K–2 assessments.

Twenty states administered assessments for diagnostic/screening purposes; four provided only formative/interim assessments; and seven administered assessments for both diagnostic and formative purposes. While all state K–2 assessments targeted reading and literacy skills, only 11 of the 35 also offered a mathematics assessment. Six other states offered an assessment that included mathematics and an additional subject area.

Table 1 below, drawn from information presented in the 2019 CCSSO publication, describes the purpose, frequency, question addressed, and system-level use for each type of assessment.

Table 1. Assessment Types, Purposes, Frequency, Questions Addressed, and System-Level Uses

Assessment Type	Purpose	Frequency	Questions Addressed	System-Level Use
Summative	Evaluates whether students have met grade-level standards	Once, typically toward the end of the school year	Have these students met grade-level standards?	• State • District • School • Classroom
Interim	Evaluates whether students are advancing toward achievement of grade-level standards	At key points throughout the year	Are students on track to meet grade-level standards by the end of the year?	• District • School • Classroom
Screener	Identifies those who may need extra support to attain desired learning outcomes	Typically, at the beginning of the school year or as needed	Do students require additional support or further evaluation?	• School • Classroom
Diagnostic	Determines the eligibility of students for specialized programming or services	As needed, typically based on the results from other assessments	What are students' strengths and areas of specific need? Can learning needs be diagnosed by additional focused assessment?	• Student
Formative	Checks students' understanding during the course of instruction to guide teaching and learning	Daily	Are students learning what was planned for them to learn? If not, how can understanding be improved to meet learning goals?	• Classroom • Student

Note. Although summative results are not typically available until after the conclusion of the school year, the results can inform classroom teaching practices for the subsequent school year.

More specifically, each of these assessment types has strengths and limitations, as follows:

Interim and summative assessments. Interim and summative assessments are designed to periodically evaluate students' progress toward, and achievement of, grade-level learning standards. Interim assessments are typically administered at key points throughout the year, while summative assessments are typically administered once toward the end of the school year. The uses of data from these assessments overlap considerably, as outlined in Table 2.

There are two distinct uses for interim assessments: to predict performance on summative assessments and to evaluate student learning and progress toward end-of-year goals.

Predictive interim assessments may include any grade-level learning standards. While they can help school- and district-level staff to identify students who are not on track to meet learning expectations required for promotion or meet the performance expectation associated with the summative assessment, they can also prompt teachers to identify curricula and approaches that may not be supportive of learning for all students. There is a risk that students may be assessed on content they have not yet had the opportunity to learn, whether through direct instruction or their own discovery. Consequently, a low score may reflect the pacing of instruction or lack of opportunity to learn rather than actual student learning.

Table 2. Uses of Data From Interim and Summative Assessments

Use	Interim Assessment	Summative Assessment	Next Steps
Accountability	How well is the system currently serving students?	How well did the system serve students?	• Undertake program improvement, as needed • Implement professional learning, as needed • Make policy and funding decisions
Progress and Preparation	How well are students learning this year's grade-level standards?	Did students achieve the grade-level expectations? What are students' relative strengths and weaknesses based on reported subdomains?	• Triangulate data with other sources • Identify students' strengths and learning needs • Target instruction • Communicate with families about students' learning progress • Review curricula and pacing • Identify needed professional learning
Promotion	Might grade retention be a likely possibility for any of these students?	Are these students adequately prepared for the next grade?	• Plan for and make decisions about grade retention

Source. Authors.

To avert this pitfall and more definitively evaluate progress toward mastery of standards, states may require districts to systematize the delivery of instruction by following a pacing guide. A pacing guide ensures consistency in the order in which instructional topics are taught. When a pacing guide is used, interim assessments can be designed to assess only recently instructed content. The assessments thus evaluate whether a student has met the specified portion of grade-level standards, and those particular standards can be evaluated with greater depth than can a predictive approach.

For interim assessments to be useful to educators and policymakers, assessment validity is fundamentally required — that is, alignment between the reason for using the assessment and the purpose for which the assessment was designed (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014; Kane, 2006). For instructional usefulness, timely reporting of results is also critical, enabling teachers to adjust strategies such as grouping or targeted instruction to meet individual student needs. Timeliness of reporting is less essential for program improvement and policy planning.

Screeners and diagnostic assessments. In a school setting, *screening assessments* — commonly referred to as screeners — are brief, typically low-cost assessments given to all students to provide a basic understanding of students' discrete skills and identify students who may be in need of further evaluation. Screeners differ from interim assessments in that they are brief measures of targeted skills (e.g., phonological awareness and oral reading fluency). They differ from formative assessment in that they are not context dependent (i.e., the same screener is administered to all students), and they are used for identification rather than checking for understanding. Literacy and numeracy screeners are usually administered at the start of the year, and districts may also choose to screen students multiple times throughout the year. This flexibility is especially relevant in the K–2 context because learning expectations evolve quickly over the course of the year. Schools may also use behavioral screeners to identify students who might benefit from additional social or emotional supports in the classroom.

Importantly, screeners used for academic purposes need to be developmentally appropriate. For example, kindergarten screeners would target critical early literacy skills such as letter identification and phonological awareness, whereas screening for reading comprehension is more appropriate for students in second grade. In first grade, oral reading fluency may be more appropriate toward the end of the school year than at the beginning.

Diagnostic assessments are used as a follow-up for individual students for whom the screening process identified the need for additional support. Diagnostic assessments can confirm or add detail to initial screening results and provide additional information related to eligibility for specialized programming or services.

Recently, online curriculum providers — whose services include providing screening, diagnostic, and interim assessments — have prompted expanded use of diagnostic assessments. Rather than using such assessments only for students whose screening identified the need for greater support, an increasingly common approach is to administer diagnostic assessments to all students in order to inform decisions about the most appropriate content for each student, whether that student is below, at, or above grade level.

Formative assessment. *Formative assessment* is a process, a planned and ongoing exchange between teacher and student in which each looks for real-time evidence of how learning is progressing that informs the need for adjustments to teaching and learning. In this process, teachers and students alike are learners. Teachers use what they learn about individual and collective student progress to guide their instruction. Students use what they learn about their own progress to guide their ongoing learning efforts (CCSSO, 2018).

Effective formative assessment is intentional and continuous. The process hinges on the teacher having clear learning goals and communicating those to students, along with clear criteria for knowing when students have met those goals. Everyone, in short, has a clear picture of the various ways in which students might demonstrate mastery. Teachers must also be prepared to respond when students are not progressing as expected or hoped. Teachers' use of formative assessment will necessarily vary, depending on the individual student or students. The process also includes student agency — that is, a learner's willingness and ability to engage in self-assessment and to both give and be open to accepting peer feedback. Given the complexities of formative assessment, classroom educators and administrators will likely profit from ongoing professional learning opportunities on how to use it effectively to improve student outcomes.

Although the formative assessment process is nearly universally viewed as a core component of the learning process (Andrade & Cizek, 2010; Popham, 2013), little quantitative research has been done on its efficacy, particularly in terms of studies specific to children in the earliest grades (Turner & Coburn, 2012; Riley-Ayers, 2014). However, one recent review of research on the effects of formative assessment in grades K–5 provides some preliminary guidance on where and how to use it most effectively. Researchers evaluated 23 studies, published between 1988 and 2014, that met the criteria for What Works Clearinghouse evidence standards and procedures (Klute et al., 2017). Across the reviewed studies, students who participated in formative assessment performed better on measures of academic achievement than those who did not. Formative assessment had larger positive effects when used during mathematics instruction as compared to reading and writing instruction. There was also evidence that student-directed formative assessment, including self- and peer assessment, was effective specifically for mathematics instruction, while educator- or computer-program-directed approaches were effective for both mathematics and reading.

Considerations for Building a K–2 Assessment System

How can states build K–2 assessment systems that are coherent and effective? The starting point is purpose. Educators and policymakers need to ask themselves: What do we want to learn about our students and for what reason? What question(s) do we want answered and to what end? Within the overall purpose of narrowing the readiness gap and improving children's early grade achievement, specific purposes may include summarizing learning from a school year to support instruction for individual students, planning classroom curriculum, providing population-level data intended to improve program quality, or guiding resource allocations. To serve this range of purposes, existing early-grade assessment systems include multiple types of assessments (Goldstein & Flake, 2016).

As noted earlier, the use of assessment results must be aligned to the assessment's intended purpose. That match is essential for assessment validity — the foundation of an effective assessment system. In addition, states fundamentally need to address where, when, and how data resulting from assessments will be used to improve learning for K–2 students who are in need of greater support.

As policy leaders strive to create an effective K–2 assessment system, state experiences to date offer considerations that can help inform their efforts. These include the following:

Summative assessments provide a common statewide performance metric but alone are insufficient. A common end-of-year metric that reflects the depth and breadth of students' content learning aligned to the state standards is important for measuring achievement of those standards. When adapted for the early grades (K–2), summative assessments can inform policy and program decisions and identify opportunities for professional learning by showing how well student performance aligns to grade-level expectations.

Systematic monitoring of student learning based on a common metric is especially critical in states with laws that require the retention of students who were not proficient in reading by third grade. As of 2018, 16 states were developing or had instituted such laws (Weyer, 2018). These statutes show a focus on literacy over other aspects of learning and development. They also point to a growing need for state K–2 assessment systems that include identification of students who face the likelihood of retention, along with an evaluation of their strengths and weaknesses, so that educators can tailor instructional support to specific learning needs. While summative assessments can identify districts and schools that serve students in need of greater support to succeed, they do not provide enough data on individual students to guide student-specific action. Moreover, summative assessment results typically arrive after students have moved on from the assessed grade.

Of the 35 states in the CCSSO report that offered some type of statewide assessment in grades K–2, only four — Georgia, Indiana, Michigan, and Tennessee — included end-of-year assessments (CCSSO, 2019). In Georgia, kindergarten teachers complete a year-long performance-based assessment, GKIDS, designed to provide ongoing diagnostic information about kindergarten students' developing skills in English language arts, mathematics, science, social studies, personal/social development, and approaches to learning. GKIDS is characterized both as a summative assessment, since it provides a summary of student performance in English language arts and mathematics at the end of the kindergarten school year, and also as a formative assessment, since it supports kindergarten teachers to plan instruction throughout the year (Georgia Department of Education, 2018).

Tennessee provides an optional summative assessment of early literacy and mathematics skills for second-grade students, designed to inform both second- and third-grade teachers of students' mastery of the standards and to support schools and districts as they measure their progress toward the goal of having 75 percent of third graders reading on grade level by 2025 (Tennessee Department of Education, 2017). Indiana and Michigan both had summative assessments in the primary grades as of the CCSSO's 2019 reporting, but early grade summative assessments are no longer part of those states' assessment systems. In Indiana, the Indiana Reading Evaluation and Determination (IREAD-3) is a summative assessment

administered to third graders that provides a measure of mastery of foundational reading standards through grade 3 (Indiana Department of Education, 2019). In Michigan, there are benchmark (interim) assessments of early literacy and mathematics skills administered in the fall, winter, and spring in kindergarten and first and second grade. The state provides the assessments for this use, but districts also have the option to select their own assessments (Michigan Department of Education, n.d.-a; Michigan Department of Education, 2019).

Interim assessments may have greater utility for tracking the development of early reading and mathematics skills. In the absence of state summative assessments, interim assessments can support district and school staff in assessing student learning. Interim assessments may, in fact, be more useful to educators because they are administered more frequently and typically structured with more targeted content than summative assessments. States can mandate a specific commercially or locally developed assessment or offer districts a list of state-approved assessments, as is done in Louisiana and Michigan (Louisiana Department of Education, n.d.-a; Michigan Department of Education, n.d.-b).

State-approved lists have the advantage of allowing districts to align assessments to their students' needs. Different assessments, however, may measure slightly different constructs — for example, the conceptualization of early reading skills may vary somewhat across assessments. States that permit districts to provide their own interim solution should require evidence that the chosen assessments are aligned to state standards and measure the depth and breadth of the assessable state standards.

CCSSO reported that approximately 70 percent of the 26 states that offered an interim assessment required districts to report results to the state (CCSSO, 2019). But comparing those results statewide can be a challenge if district-selected assessments lack a common metric for "proficiency" or "on track to proficiency." One solution is for states to institute a standard-setting procedure that supports a common target for the identification of students who are on track for achieving proficiency and those who are not.

An alternative to interim assessments for gauging early reading and mathematics skills is the use of screeners. A growing number of states are integrating reading screeners into state systems, a shift that coincides with a rising number of third-grade reading retention laws across the country (CCSSO, 2019). One advantage of screeners is that they measure targeted skills that are demonstrably associated with the learning outcome of interest. Because screener assessments are targeted, they tend to be less time-consuming than interim or summative assessments.

A system of screeners at key points between kindergarten and second grade can support districts in identifying students who, without added support, face the likelihood of retention in third grade. Further, states should consider using mathematics as well as reading screeners in the early grades, since early mathematics skills have a demonstrated association with academic success in later elementary grades (Watts et al., 2014).

Including support for formative assessment in a state system helps enable real-time classroom intervention and remediation. Assessments that involve a time lag of weeks or months before results are available have limited instructional usefulness, since remediation requires that teachers return to content taught long before. In contrast, the formative assessment process

allows teachers to adjust instruction in real time, based on student progress (CCSSO, 2018). In recognition of the importance of formative assessment in state-level assessment systems, the Smarter Balanced Assessment Consortium provides member states with activities and lessons to support districts in using the formative assessment process (Smarter Balanced, 2020). Similarly, Tennessee is expanding its Tennessee Comprehensive Assessment Program to include state-provided summative, interim, and formative components (Tennessee Department of Education, 2020). (For the purposes of its assessment system, Tennessee defines formative assessments as short assessments that cover a limited number of the Tennessee Academic Standards in each assessment, which can be administered by a teacher as desired.)

Conclusion

Robust assessment systems in K–2 can provide data to evaluate school and district performance, gauge young students' progress against state standards, and help educators identify and intervene early with students who may need added support to achieve desired learning outcomes. Model state assessment systems for the early grades include summative and interim assessments as well as screeners and formative assessment. Summative assessments offer schools and districts an annual evaluation of overall student performance and can serve as an indicator of the need to focus supports on individual students who, without intervention, would appear to be on track for poor performance in the upcoming year. Interim assessments are more instructionally useful since they are frequent and more targeted. Screeners, similarly, can help teachers identify students who need specific kinds of academic support. The utility of these tools is enriched throughout the year with formative assessment. The key is designing assessment systems that include multiple kinds of assessments that balance out the strengths and limitations of each type in the service of helping the state improve learning for all K–2 students.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing.* American Educational Research Association.

Andrade, H. L., & Cizek, G. J. (2010). *Handbook of formative assessment.* Routledge.

CCSSO. (2018). *Revising the definition of formative assessment.* https://ccsso.org/sites/default/files/2018-06/Revising%20the%20Definition%20of%20Formative%20Assessment.pdf

CCSSO. (2019). *K–2 assessments: An update on state adoption and implementation.* https://ccsso.org/sites/default/files/2019-06/K-2%20Assessments%20Paper%20FINAL.pdf

Center for Standards, Assessment, and Accountability (CSAA). (2021). *State of the states: Pre-K/K assessment.* https://csaa.wested.org/tools/state-of-states/

Connor, C. M., Morrison, F. J., Schatschneider, C., Toste, J. R., Lundblom, E., Crowe, E. C., & Fishman, B. (2011). Effective classroom instruction: Implications of child characteristics by reading instruction interactions on first graders' word reading achievement. *Journal of Research on Educational Effectiveness, 4*(3), 173–207.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., Pagani, L. S., Feinstein, L., Engel, M., Brooks-Gunn, J., Sexton, H., Duckworth, K., & Japel, C. (2007). School readiness and later achievement. *Developmental Psychology, 43*, 1428–1446.

Georgia Department of Education. (2018). *Georgia Kindergarten Inventory of Developing Skills: Administration manual*. https://www.gadoe.org/Curriculum-Instruction-and-Assessment/Assessment/Documents/GKIDS/Resources_18-19/GKIDS_Administration_Manual_2018-19.pdf

Goldstein, J., & Flake, J. (2016). Towards a framework for the validation of early childhood assessment systems. *Educational Assessment, Evaluation and Accountability, 28*(3), 273–293.

Indiana Department of Education. (2019). *2019-2020 Indiana assessments policy manual.* Office of Student Assessment.

Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17–64). National Council on Measurement in Education & American Council on Education.

Klute, M., Apthorp, H., Harlacher, J., & Reale, M. (2017). *Formative assessment and elementary school student academic achievement: A review of the evidence (REL 2017–259).* U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Central. https://www.sri.com/publication/formative-assessment-and-elementary-school-student-academic-achievement-a-review-of-the-evidence/

Louisiana Department of Education. (n.d.). *Kindergarten through 2nd grade assessments.* https://www.louisianabelieves.com/measuringresults/assessments-in-kindergarten-through-2nd-grade

Michigan Department of Education. (n.d.-a). *Michigan's requirements for assessing learners in early elementary grades.* https://www.michigan.gov/documents/mde/Early_Literacy_and_Mathematics_Benchmark_Assessment_Overview_663446_7.pdf

Michigan Department of Education. (n.d.-b). *2021-22 Early literacy MCL: 380.1280f initial assessments.* https://www.michigan.gov/documents/mde/17-18_Initial_Assessment_List_560866_7.pdf

Michigan Department of Education. (2019). *Benchmark assessments: Early literacy and mathematics 2021-2022 test administration manual.* https://www.michigan.gov/documents/mde/Early_Literacy_and_Mathematics_TAM_663326_7.pdf

Neuman, S. B., & Dickinson, D. (Eds.). (2001). *The handbook of early literacy research.* Guilford.

Popham, W. J. (2013). Waving the flag for formative assessment. *Education Week, 32*(15), 29.

Riley-Ayers, S. (2014). *Formative assessment: Guidance for early childhood policymakers* [CEELO Policy Report]. Center on Enhancing Early Learning Outcomes.

Smarter Balanced. (2020). *Understanding the formative assessment process.* https://smartertoolsforteachers.org/resource/242

Tennessee Department of Education. (2017). *Grade 2 assessment.* https://www.tn.gov/education/assessment/tnready/grade-2-assessment.html

Tennessee Department of Education. (2020). *Request for proposals for assessment item and test form development.* https://www.tn.gov/content/dam/tn/generalservices/documents/cpo/rfp-updates/33111-00320/RFP_33111-00320%20Assessment%20Item%20and%20Test%20Form%20Development.pdf

Turner, E. O., & Coburn, C. E. (2012). Interventions to promote data use: An introduction. *Teachers College Record, 114*, 1–13.

Watts, T. W., Duncan, G. J., Siegler, R. S., & Davis Kean, P. E. (2014). What's past is prologue: Relations between early mathematics knowledge and high school achievement. *Educational Researcher, 43*(7), 352–360.

Weisenfeld, G. G., Garver, K., & Hodges, K. (2020). Federal and state efforts in the implementation of kindergarten entry assessments (2011–2018). *Early Education and Development.* Rutgers University. https://doi.org/10.7282/t3-cjs2-k115

Weyer, M. (2018). A look at third-grade reading retention policies. *LegisBrief, 26,* 21. National Conference of State Legislatures.

Suggested Citation: Jensen, J. L., Goldstein, J., and Brunetti, M. A. (2021). *K–2 Assessment Systems Enable Early Intervention to Foster Student Success.* WestEd.

WestEd is a nonpartisan, nonprofit research, development, and service agency that partners with education and other communities throughout the United States and abroad to promote excellence, achieve equity, and improve learning for children, youth, and adults. WestEd has more than a dozen offices nationwide, from Massachusetts, Vermont, Georgia, and Washington, DC, to Arizona and California, with headquarters in San Francisco. For more information, visit WestEd.org or call 415.565.3000.