



## REL Appalachia Ask A REL Response

Data Use, Early Childhood, Literacy, Math  
February 2019

### Question:

What are state policies regarding reading and math assessments in grades K–3? Is there any evidence that some K–3 assessments are better than others at accurately measuring student learning and informing instruction?

### Response:

Thank you for your request to our REL Reference Desk regarding evidence-based information about state policies concerning reading and mathematics assessments for students in grades K–3. Ask A REL is a collaborative reference desk service provided by the 10 Regional Educational Laboratories (RELs) that, by design, functions much in the same way as a technical reference library. Ask A REL provides references, referrals, and brief responses in the form of citations in response to questions about available education research.

Following an established REL Appalachia research protocol, we searched for peer-reviewed articles and other research reports on reading and math assessments in grades K–3. We focused on identifying resources that specifically addressed state policies regarding K–3 reading and math assessments, the characteristics of good early childhood assessments, and research on particular K–3 screening, diagnostic, and formative assessments. The sources included ERIC and other federally funded databases and organizations, research institutions, academic research databases, and general Internet search engines. For more details, please see the methods section at the end of this document.

The research team did not evaluate the quality of the resources provided in this response; we offer them only for your reference. Also, the search included the most commonly used research databases and search engines to produce the references presented here, but the references are not necessarily comprehensive, and other relevant references and resources may exist. References are listed in alphabetical order, not necessarily in order of relevance.

### References

Atkins-Burnett, S. (2007). *Measuring children's progress from preschool through third grade*. Washington, DC: Mathematica Policy Research. Retrieved from <https://www.mathematica-mpr.com/-/media/publications/pdfs/measchildprogress.pdf>

*From the introduction:* “This paper will discuss the measurement of child outcomes in the context of evaluating the effectiveness of preschool programs for children. Little is known about how individual districts and states are evaluating early childhood programs, so this discussion will highlight some of the ways in which this challenge is being addressed. After a brief discussion of the importance of focusing on the whole child rather than just their language and cognitive domains, most of the paper will explore what is known about current assessment methods used with young children. Problems related to relying solely on traditional, on-demand standardized tests to assess achievement of young children will be explained. Although young children who are English Language Learners (ELL) represent an increasing proportion of preschool children, it is beyond the scope of this paper to discuss in-depth the issues involved in assessing these children (see Lazarin, 2006 for some discussion of K–12 efforts). Observational measures that span the preschool to elementary age range offer an alternative to direct testing. The use of these measures in formative evaluation efforts will be discussed with the caution that high stakes should never be attached to these measures. Using a multimethod approach would provide a richer portrayal of children’s performance. Innovative and alternative approaches to assessment used by some states will be highlighted, and concerns about reliability of teacher judgments discussed. The paper concludes with a brief discussion of measuring classroom quality and recommendations for next steps.”

Brown, R. S., & Coughlin, E. (2007). *The predictive validity of selected benchmark assessments used in the Mid-Atlantic Region* (Issues & Answers Report, REL 2007–No. 017). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Mid-Atlantic. Retrieved from <https://eric.ed.gov/?id=ED499099>

*From the summary:* “This report examines the availability and quality of predictive validity data for a selection of benchmark assessments identified by state and district personnel as in use within Mid-Atlantic Region jurisdictions. Based on a review of practices within the school districts in the region, this report details the benchmark assessments being used, in which states and grade levels, and the technical evidence available to support the use of these assessments for predictive purposes. The report also summarizes the findings of conversations with test publishing company personnel and of technical reports, administrative manuals, and similar materials. The study investigates the evidence provided to establish a relationship between district and state test scores, and between performance on district-administered benchmark assessments and proficiency levels on state assessments. When particular district benchmark assessments cover only a subset of state test content, the study sought evidence of whether district tests correlate not only with overall performance on the state test, but also with relevant subsections of the state test. While the commonly used benchmark assessments in the Mid-Atlantic Region jurisdictions may possess strong internal psychometric characteristics, the report finds that evidence is generally lacking of their predictive validity with respect to the required state or summative assessments. To provide the jurisdictions with additional information on the predictive validity of the benchmark assessments currently used, further research is needed linking these assessments and the state tests currently in use. Additional research could help to

develop the type of predictive validity evidence school districts need to make informed decisions about which benchmark assessments correspond to state assessment outcomes, increasing potential success of instructional decisions meant to improve student learning as measured by state tests.”

Clements, D. H., Saram, J., Day-Hess, C., Germeroth, C., Ganzar, J., Pugia, A., & Barker, J. (2017). *Comprehensive review of assessments of early childhood mathematics competencies*. Unpublished manuscript. Retrieved from <https://dreme.stanford.edu/publications/comprehensive-review-assessments-early-childhood-mathematics-competencies>

*From the abstract:* “Mathematics-specific instruments focusing on the preschool through second grade age range were identified from comprehensive reviews of the literature and recommendations from a survey of experts in the field, resulting in extensive coding of 16 mathematics assessments. Data were collected on these assessments in a variety of categories, including: author(s), purpose/goals, grade/age, number of items, content areas covered, languages, time of administration, platform, administration format, training requirements, and psychometric information. Our findings indicate that there are instruments that meet criteria for different purposes, but they differ substantially in their logistical and content-coverage characteristics, so they should be considered and compared carefully before being selected to meet specific goals within the domains of policy, practice, and research. More specifically, we found that most widely-used instruments are useful for certain purposes, but are limited in terms of the content areas covered and the ability to capture children’s level of thinking. These findings, and the detailed content coding made available in this work, provide a far more comprehensive tool for the selection of instruments than has previously been available.”

“*Note:* This is a brief summary of a much more extensive report to be submitted for publication.”

Croft, M. (2016). *Issue brief: State adoption and implementation of K–2 assessments*. Iowa City, IA: ACT Research & Policy. Retrieved from [https://www.act.org/content/dam/act/unsecured/documents/5738\\_Issue\\_Brief\\_State\\_Adoption\\_of\\_K-2\\_Assess\\_WEB\\_secure.pdf](https://www.act.org/content/dam/act/unsecured/documents/5738_Issue_Brief_State_Adoption_of_K-2_Assess_WEB_secure.pdf)

*From the introduction:* “Learning gaps often start prior to elementary school. For instance, a recent study found that gaps in preliteracy skills emerged between Mexican American and White children by age 2. Similarly, researchers have found that by the time children are two years old, low-income children’s vocabularies are six months behind high-income children’s vocabularies. These gaps persist and often widen throughout students’ educational careers, and it is very difficult for students—particularly at-risk students—to catch up. The early presence of these learning gaps has drawn considerable attention in the last few years. In 2011, the Obama administration funded the Race to the Top Early Learning Challenge competitive grant program for states. A key component of the program was the collection of data to measure outcomes and progress, particularly through the creation of two types of assessments, which were designed to inform preschool instruction and to measure

readiness for kindergarten, respectively. These assessments were designed to provide a 'critical link' to the K–12 system; however, part of this link is currently missing—in grades K–2. Once students start elementary school, they are not required under federal law to test using a standardized assessment until grade 3. Although states are not required to administer standardized assessments in grades K–2, they may opt to do so to better gauge student progress, to help identify students who may be academically lagging so that they can receive remediation, to help with program evaluation and continuous improvement, and/or as part of a state accountability system....Given these differing perspectives on the administration of standardized assessments in grades K–2, I conducted a study to better understand how states that are implementing such assessments are using them. To do so, I relied on publicly available information from the websites of state departments of education, coding assessments into categories indicating the purpose of the assessment; how the assessment instrument was chosen for use; the grade levels and subject areas/domains covered; whether the assessment is mandatory or optional; and whether the results of the assessment must be reported to the state. The study focuses on assessments that would be administered during or at the end of the school year, but does not include kindergarten readiness assessments that are administered in the first months of kindergarten.”

Diffey, L. (2018). *50-state comparison: State kindergarten-through-third-grade policies*. Denver, CO: Education Commission of the States. Retrieved from <https://www.ecs.org/kindergarten-policies/>

*From the introduction:* “High-quality, early elementary years offer a critical opportunity for development and academic learning for all children. Key components of a quality, K–3 experience include kindergarten, qualified teachers, seamless transitions, appropriate assessments and interventions, family engagement, social-emotional supports and academic supports. Education Commission of the States has researched the policies that guide these key components in all 50 states to provide this comprehensive resource. Click on the questions below for 50-State Comparisons, showing how all states approach specific policies, or view a specific state’s approach by going to the individual state profiles page.”

Gersten, R., Clarke, B. S., Haymond, K., & Jordan, N. C. (2011). *Screening for mathematics difficulties in K–3 students. Second edition*. Portsmouth, NH: Center on Instruction at RMC Research Corporation. Retrieved from <https://eric.ed.gov/?id=ED524577>

*From the preface:* “Since 2007, when this technical report was originally issued, the assessment field has made considerable progress in developing valid and reliable screening measures for early mathematics difficulties. This update includes new research published since 2007. It focuses on valid and reliable screening measures for students in kindergarten and first grade. However, the authors also examined data on screening tests for second and third grades because the goal of screening is to identify students who might struggle to learn mathematics during their initial school years. Appended are: (1) Summary of the Evidence Base on Early Screening Measures as of December 2010; and (2) Procedure for Reviewing the Literature on Early Screening in Mathematics.”

Jordan, N. C., Glutting, J., Ramineni, C., & Watkins, M. W. (2010). Validating a number sense screening tool for use in kindergarten and first grade: Prediction of mathematics proficiency in third grade. *School Psychology Review*, 39(2), 181–195. Abstract retrieved from <https://eric.ed.gov/?id=EJ891847>; full text available at <http://edpsychassociates.com/Papers/NumberSenseScreen%282010%29.pdf>

*From the abstract:* “Using a longitudinal design, children were given a brief number sense screener (NSB) screener ( $N = 204$ ) over six time points, from the beginning of kindergarten to the middle of first grade. The NSB is based on research showing the importance of number competence (number, number relations, and number operations) for success in mathematics. Children’s mathematics achievement on a validated high-stakes state test was measured 3 years later, at the end of third grade. Test-retest reliability estimates were obtained for the NSB. Two criterion groups were then formed on the basis of the third-grade achievement test (children who met and who did not meet mathematics standards). Diagnostic validity analyses for the NSB were completed using repeated measures analyses of variance and receiver operator curve analyses. Results from all analyses revealed that scores on the NSB in kindergarten and first grade predicted mathematics proficiency in third grade. Areas under the receiver operator curve indicated that the NSB has high diagnostic accuracy (areas under the receiver operator curve = 0.78–0.88). Findings suggest that kindergarten and first-grade performance on the NSB is meaningful for predicting which children experience later mathematics difficulties.”

Rouse, H. L., & Fantuzzo, J. W. (2006). Validity of the Dynamic Indicators for Basic Early Literacy Skills as an indicator of early literacy for urban kindergarten children. *School Psychology Review*, 35(3), 341–235. Abstract retrieved from <https://eric.ed.gov/?id=EJ788273>; full text available at <https://pdfs.semanticscholar.org/6e00/9f97f6f7b28032ce8e1a2ccf14d30007e9bb.pdf>

*From the abstract:* “The validity of three subtests of the Dynamic Indicators for Basic Early Literacy Skills (DIBELS) was investigated for kindergarten children in a large urban school district. A stratified, random sample of 330 participants was drawn from an entire cohort of kindergarten children. Letter Naming Fluency, Phoneme Segmentation Fluency, and Nonsense Word Fluency evidenced significant concurrent and predictive validity when compared to general reading ability measured by teacher report, individual assessments, and group-administered nationally standardized tests. Evidence for convergent and discriminant validity was also found when comparing these subtests to measures of specific literacy, cognitive, and social-behavioral constructs.”

Shapiro, E. S., & Gibbs, D. P. (2014). *Comparison of progress monitoring with computer adaptive tests and curriculum based measures*. Bethlehem, PA: Center for Promoting Research to Practice, Lehigh University. Retrieved from <http://doc.renlearn.com/KMNet/R0057324CE9DD5FD.pdf>

*From the abstract:* “The purpose of this study was to compare both rates of reading achievement growth and predictive power of two widely-used assessments representing two different approaches to measurement – a computer adaptive assessment called STAR

Reading and a curriculum based measurement called AIMSweb. A total of 117 students from a school district in Tennessee were included in the sample. Data collection spanned two school years, and included students who were progress monitored (taking a minimum of 4 tests per year) in grades 1 through 4 in one year, and in grades 2 through 5 the subsequent year. Across the two years, interventions for both groups of students were consistent. The results of this study indicate that both measures were able to detect incremental change, and provide further support that both computer adaptive measures such as STAR Reading and CBMs such as AIMSweb R-CBM are acceptable for progress monitoring. Of the two measures, only STAR Reading achieved a significant correlation with the state reading assessment.”

Snow, C. E., & Van Hemel, S. B. (Eds.). (2008). *Early childhood assessment: Why, what, and how*. Washington, DC: The National Academies Press. Abstract retrieved from <https://eric.ed.gov/?id=ED555247>; full text available at <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.467.1059&rep=rep1&type=pdf>

*From the abstract:* “The assessment of young children’s development and learning has recently taken on new importance. Private and government organizations are developing programs to enhance the school readiness of all young children, especially children from economically disadvantaged homes and communities and children with special needs. Well-planned and effective assessment can inform teaching and program improvement, and contribute to better outcomes for children. This book affirms that assessments can make crucial contributions to the improvement of children’s well-being, but only if they are well designed, implemented effectively, developed in the context of systematic planning, and are interpreted and used appropriately. Otherwise, assessment of children and programs can have negative consequences for both. The value of assessments therefore requires fundamental attention to their purpose and the design of the larger systems in which they are used. ‘Early Childhood Assessment’ addresses these issues by identifying the important outcomes for children from birth to age 5 and the quality and purposes of different techniques and instruments for developmental assessments.”

### **Additional Ask A REL Responses to Consult**

Ask A REL Appalachia at SRI International. (2017). *What does the research say about the equivalency of scales on early grades (PK–4) literacy universal screening and progress monitoring measures (specifically: AIMSweb, easyCBM, STAR, DIBELS, iReady, NWEA MAP)?* Retrieved from <https://ies.ed.gov/ncee/edlabs/regions/appalachia/askarel/aar01.asp>

Ask A REL Midwest at American Institutes for Research. (2017). *Effectiveness of Dynamic Indicators of Basic Early Literacy Skills (DIBELS)*. Retrieved from <https://ies.ed.gov/ncee/edlabs/regions/midwest/askarel/2017/effectiveness-of-dibels.aspx>

Ask A REL Southwest at American Institutes for Research. (2018). *K–2 indicators predictive of later performance*. Retrieved from <https://ies.ed.gov/ncee/edlabs/regions/southwest/ask-a-rel/k-2-indicators-predictive-later-performance.aspx>

## Additional Organizations to Consult

Center on Enhancing Early Learning Outcomes (CEELO): <http://ceelo.org/>

*From the website:* “As one of 22 Comprehensive Centers funded by the U.S. Department of Education’s Office of Elementary and Secondary Education, the Center on Enhancing Early Learning Outcomes (CEELO) is designed to strengthen the capacity of State Education Agencies (SEAs) to lead sustained improvements in early learning opportunities and outcomes. We do this work through strategic and responsive technical assistance, working with SEAs, state and local early childhood leaders, and other federal and national technical assistance (TA) providers to promote innovation and accountability.”

- P–3 Alignment Projects: <http://ceelo.org/ceelo-technical-assistance/p-3-alignment/>
- State-By-State Early Learning Guidelines: <http://ceelo.org/state-map/>

Center on Standards & Assessment Implementation: <https://www.csai-online.org/>

*From the website:* “CSAI, managed by WestEd and the Center for Research on Evaluation, Standards, & Student Testing (CRESST), is one of seven content centers that provide research-based technical assistance and support to 15 Regional Comprehensive Centers (RCCs), and to the states that they serve, around each of the following areas: building state capacity and productivity; college and career readiness and success; enhancing early learning outcomes; great teachers and leaders; innovations in learning; school turnaround; and standards and assessment implementation. As a content center, CSAI is committed to providing high-quality technical assistance, research support, tools, and other resources to RCCs and to state education agencies (SEAs) to help inform decisions about standards, assessment, and accountability.”

- Interim assessment: <https://www.csai-online.org/collection/2932>
- Formative assessment: <https://www.csai-online.org/collection/2620>
- Formative assessment: State resources: <https://www.csai-online.org/spotlight/formative-assessment-state-resources>

Development and Research in Early Math Education (DREME): <https://dreme.stanford.edu/>

*From the website:* “The DREME Network was created in 2014 to advance the field of early mathematics research and improve young children’s opportunities to develop math skills. The Network focuses on math from birth through age eight years, with an emphasis on the preschool years. Network members and affiliates collaborate to conduct basic and applied research and develop innovative tools that address high-priority early math topics and inform and motivate other researchers, educators, policymakers and the public.”

## Methods

### Keywords and Search Strings

The following keywords and search strings were used to search the reference databases and other sources:

- (“K–3” OR “K–2” OR “early childhood”) AND assess\* AND (math\* OR literacy OR reading)
- (“K–3” OR “K–2” OR “early childhood”) AND assess\* AND (formative OR screen\* OR diagnostic)

### Databases and Resources

We searched ERIC, a free online library of more than 1.6 million citations of education research sponsored by the Institute of Education Sciences (IES), for relevant resources. Additionally, we searched the academic database ProQuest, Google Scholar, and the commercial search engine Google.

### Reference Search and Selection Criteria

In reviewing resources, Reference Desk researchers consider—among other things—these four factors:

- Date of the publication: Searches cover information available within the last ten years, except in the case of nationally known seminal resources.
- Reference sources: IES, nationally funded, and certain other vetted sources known for strict attention to research protocols receive highest priority. Applicable resources must be publicly available online and in English.
- Methodology: The following methodological priorities/considerations guide the review and selection of the references: (a) study types—randomized controlled trials, quasi experiments, surveys, descriptive data analyses, literature reviews, policy briefs, etc., generally in this order; (b) target population, samples (representativeness of the target population, sample size, volunteered or randomly selected), study duration, etc.; (c) limitations, generalizability of the findings and conclusions, etc.
- Existing knowledge base: Vetted resources (e.g., peer-reviewed research journals) are the primary focus, but the research base is occasionally slim or nonexistent. In those cases, the best resources available may include, for example, reports, white papers, guides, reviews in non-peer-reviewed journals, newspaper articles, interviews with content specialists, and organization websites.

Resources included in this document were last accessed on January 29, 2019. URLs, descriptions, and content included here were current at that time.

This memorandum is one in a series of quick-turnaround responses to specific questions posed by education stakeholders in the Appalachia region (Kentucky, Tennessee, Virginia, and West Virginia), which is served by the Regional Educational Laboratory Appalachia (REL AP) at SRI International. This Ask A REL response was developed by REL AP under Contract ED-IES-17-C-0004 from the U.S. Department of Education, Institute of Education Sciences, administered by SRI International. The content does not necessarily reflect the views or policies of IES or the U.S. Department of Education, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. government.