REL Appalachia Ask-A-REL Response
Data Use, Educator Effectiveness, Math
June 2017

Question:

What available research or resources are related to designing and implementing performance-based assessments in science and mathematics at the elementary school level?

Response:

Thank you for your request to our REL Reference Desk regarding evidence-based information about designing and implementing performance-based assessments in elementary-level science and mathematics. 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 research reports and descriptive study articles on designing and implementing performance-based assessments in science and mathematics. 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.

Research References

Abbott, A. L. (2016). Locally developed performance assessments: One state's decision to supplant standardized tests with alternative measures. *Journal of Organizational and Educational Leadership, 2*(1), Article 5. Retrieved from https://eric.ed.gov/?id=EJ1131521

From the abstract: "The purpose for this study was to develop a descriptive account of one large Virginia school district's plan for implementation of alternative, locally developed assessments designed to supplant standardized measures. As policy reform with alternative assessments has been under-researched for the past 30 years, there is a need for studies conducted at the district/state level that examine new methods and procedures to assess cognitive growth and complex skill sets (Darling-Hammond, et al., 2013). This research was timely as the action plan implemented by the Longridge City Public School division during the first trial, 2014/15 school year, was reported. For this descriptive case study (Stake, 1995; Yin, 2014), the research questions are as follows: (1) What are the processes of a

school division's leadership team for the development and enactment of alternative assessments? and (2) What are the needs and challenges of a school division's leadership team during the development and enactment of alternative assessment? The present study holds significance as it extends previous investigations of how state policy reform is linked with educational practice at the district level. In light of the reform, placing the focus on support and leadership at the local level was purposeful in discovering the considerations and actions necessary to meet the needs of numerous stakeholder groups across Longridge City Public Schools. To answer the research questions, the Virginia Department of Education 2014 legislative guidelines served as a framework to guide the development of this study. The tradition of case study (Yin, 2014) as an analytical approach was used to uncover the processes enacted by central office personnel (i.e., district leaders) to develop and implement alternative assessment, in the format of 'performance-based assessments,' or PBAs. The findings reveal the district's ability to comply with the state's legislative reform mandates, while to navigate new territory of 'local control.'"

Best, J., & Winslow, E. (2015). *Curriculum-Embedded Performance Assessments (CEPAs): Policy considerations for meaningful accountability.* Denver, CO: McREL International. Retrieved from https://eric.ed.gov/?id=ED557604

From the abstract: "Educational assessments provide data that give policymakers a 'snapshot' of how students are performing and serve as a means of holding teachers, schools, and districts accountable. Many contend, however, that assessments could do more to promote deeper learning in K–12 environments. One interesting possibility is the use of Curriculum-Embedded Performance Assessments (CEPAs). CEPAs are instructional units designed to promote subject matter learning and the acquisition of skill sets while providing data that can be used for both summative and formative purposes. The ultimate goal of CEPAs is to maintain consistency between what is taught, assessed, and how teachers are prepared. This brief provides policymakers an overview of CEPAs, gives examples of successful CEPA application at the classroom, school, district, and state levels, and suggests ways that CEPAs can improve policy-driven outcomes. Questions to consider and recommendations for policymakers are included."

Davey, T., Ferrara, S., Holland, P. W., Shavelson, R., Webb, N. M., & Wise, L. L. (2015). *Psychometric considerations for the next generation of performance assessment*. Princeton, NJ: Center for K–12 Assessment & Performance Management at ETS. Retrieved from https://www.ets.org/research/policy_research_reports/publications/report/2015/jubf

From the introduction: "The purpose of this report is to shine a light on the elephant in the room, exposing the conceptual and methodological challenges of performance assessment. We do this by reviewing what is known about performance assessment and recent psychometric developments that might address some of the challenges, and by identifying areas for new developments. To this end, the next chapter, Chapter II, takes up the definition of performance assessment. The two chapters that follow Chapter II enumerate challenges and possible remedies in producing reliable, comparable scores on performance assessments. One chapter focuses on an individual's performance (Chapter III) and the

other on performance in and of a group (Chapter IV). Chapter V takes responses from performance assessments and discusses alternative ways of modeling them to produce reliable and valid scores. And Chapter VI, the final chapter, summarizes the issues identified in the report and recommendations for addressing them along with speculations for future psychometric developments."

Kauble, A., & Wise, D. (2015). Leading instructional practices in a performance-based system. *Education Leadership Review of Doctoral Research, 2*(2), 88–104. Retrieved from https://eric.ed.gov/?id=EJ1105756

From the abstract: "Given the shift to Common Core, educational leaders are challenged to see new directions in teaching and learning. The purpose of this study was to investigate the instructional practices which may be related to the effectiveness of a performance-based system (PBS) and their impact on student achievement, as part of a thematic set of dissertations that examined different aspects of a PBS system in three separate school systems in different areas of the continental US. This specific study examined the role of instructional strategies in implementing and sustaining a performance-based system in order to better understand how instructional strategies can improve the implementation of an innovative school reform as well as support a sustainable outcome that improves student academic achievement. In the study, a questionnaire was utilized to measure instructional strategy perceptions. Next, instructional strategy actions and perceptions were explored through face-to-face focus groups with participants. Finally, classroom observations were conducted to determine which components of instructional practices are commonly used in a PBS. The design for this mixed method study integrated both qualitative and quantitative methods. The results of the study indicated that there were some differences in the perceptions and usage of instructional practices across grade levels and districts. It was found the participants believed that the individualized nature of a PBS along with instilling student self-motivation is what promotes student achievement, not the use of specific instructional practices."

Ketelhut, D. J., Nelson, B., Schifter, C., & Kim, Y. (2013). Improving science assessments by situating them in a virtual environment. *Education Sciences, 3*(2), 172–192. Retrieved from https://eric.ed.gov/?id=EJ1117991

From the abstract: "Current science assessments typically present a series of isolated fact-based questions, poorly representing the complexity of how real-world science is constructed. The National Research Council asserts that this needs to change to reflect a more authentic model of science practice. We strongly concur and suggest that good science assessments need to consist of several key factors: integration of science content with scientific inquiry, contextualization of questions, efficiency of grading and statistical validity and reliability. Through our Situated Assessment using Virtual Environments for Science Content and inquiry (SAVE Science) research project, we have developed an immersive virtual environment to assess middle school children's understanding of science content and processes that they have been taught through typical classroom instruction. In the virtual environment, participants complete a problem-based assessment by exploring a

game world, interacting with computer-based characters and objects, collecting and analyzing possible clues to the assessment problem. Students can solve the problems situated in the virtual environment in multiple ways; many of these are equally correct while others uncover misconceptions regarding inference-making. In this paper, we discuss stage one in the design and assessment of our project, focusing on our design strategies for integrating content and inquiry assessment and on early implementation results. We conclude that immersive virtual environments do offer the potential for creating effective science assessments based on our framework and that we need to consider engagement as part of the framework."

Kim, K. H., VanTassel-Baska, J., Bracken B. A., Feng, A., & Stambaugh, T. (2014). Assessing science reasoning and conceptual understanding in the primary grades using standardized and performance-based assessments. *Journal of Advanced Academics, 25*(1).

From the abstract: "Project Clarion, a Jacob K. Javits-funded project, focused on the scale-up of primary-grade science curricula. Curriculum units, based on an Integrated Curriculum Model (ICM), were developed for high-ability learners, but tried out with all students in Title I settings to study the efficacy of the units with all learners. The units focus on the development of students' conceptual understanding to undergird science content attainment. Teaching and learning models, such as concept formation and concept mapping, were used to scaffold science learning and reasoning for appropriate curriculum differentiation. Science content mastery was measured using the Metropolitan Achievement Tests. Reasoning skills were measured using the Test of Critical Thinking. Understanding of macro-concepts and content attainment were measured by curriculum-embedded performance-based assessments. Students with the ICM outperformed students without the specialized curriculum in science content and reasoning skills, and showed greater growth in both conceptual understanding and content attainment."

Lane, S., Parke, C. S., & Stone, C. A. (2002). The impact of a state performance-based assessment and accountability program on mathematics instruction and student learning: Evidence from survey data and school performance. *Educational Assessment, 8*(4), 279–315. Retrieved from https://eric.ed.gov/?id=EJ761416

From the abstract: "The purpose of this study was to examine the impact of the Maryland School Performance Assessment Program (MSPAP) and the Maryland Learning Outcomes (MLOs) on mathematics classroom instruction and assessment practices, professional development, and student learning. The data sources included questionnaires for principals, mathematics teachers, and students, as well as student performance on MSPAP over a 5-year period. Ninety elementary and middle schools in Maryland participated in the study. The results indicate that principals and teachers tended to support MSPAP as a tool for making changes in instruction, teachers were making some positive changes in mathematics instruction because of MSPAP (based on the questionnaire data), and the schools for which teachers reported that MSPAP had a greater impact on their mathematics instruction had greater MSPAP performance gains in mathematics over the 5 years."

Marion, S., & Leather. P. (2015). Assessment and accountability to support meaningful learning. *Education Policy Analysis Archives, 23*(9), 1–19. Retrieved from https://eric.ed.gov/?id=EJ1051681

From the abstract: "This paper presents an overview of New Hampshire's efforts to implement a pilot accountability system designed to support deeper learning for students and powerful organization change for schools and districts. The accountability pilot, referred to as Performance Assessment of Competency Education or PACE, is grounded in a competency based educational approach designed to ensure that students have meaningful opportunities to achieve critical knowledge and skills. These opportunities are judged by the outcomes students achieve and not by inputs such as seat time. Therefore, students must achieve these competencies before moving on to the next major learning targets and/or graduating from high school. High quality performance assessments play a crucial role in the PACE system because of the need to have assessments that measure the depths of student understanding of these complex learning targets. Performance assessments are used as both summative and interim measures in the PACE system as a way to document student learning of the competencies and to support remediation or extension interventions. The paper describes the system of assessments being implemented as part of the PACE pilot as well as providing a discussion of the technical quality issues the state is working to address as part of this accountability pilot. For example, being able to produce valid and comparable annual determinations for all students each year is a considerable technical challenge as well as documenting the degree to which all students are held to the same threshold expectations (equity). The paper concludes by relating the PACE initiative to the push for deeper and more meaningful learning for students."

National Research Council. (2014). *Developing Assessments for the Next Generation Science Standards*. Washington, DC: The National Academies Press.

From the introduction: "The committee will make recommendations for strategies for developing assessments that validly measure student proficiency in science as laid out in the new K–12 science education framework. The committee will review recent and current, ongoing work in science assessment to determine which aspects of the necessary assessment system for the framework's vision can be assessed with available techniques and what additional research and development is required to create an overall assessment system for science education in K–12. The committee will prepare a report that includes a conceptual framework for science assessment in K–12 and will make recommendations to state and national policy makers, research organizations, assessment developers, and study sponsors about the steps needed to develop valid, reliable, and fair assessments for the framework's vision of science education. The committee's report will discuss the feasibility and cost of its recommendations."

Pellegrino, J. W. (2013). Proficiency in science: Assessment challenges and opportunities. *Science, 30*(6130), 320–323. Retrieved from http://science.sciencemag.org/content/340/6130/320

From the abstract: "Proficiency in science is being defined through performance expectations that intertwine science practices, cross-cutting concepts, and core content knowledge. These descriptions of what it means to know and do science pose challenges for assessment design and use, whether at the classroom instructional level or the system level for monitoring the progress of science education. There are systematic ways to approach assessment development that can address design challenges, as well as examples of the application of such principles in science assessment. This Review considers challenges and opportunities that exist for design and use of assessments that can support science teaching and learning consistent with a contemporary view of what it means to be proficient in science."

Stanford School Redesign Network. (2008). *What is performance-based assessment? Informational booklet*. Stanford, CA: Stanford University School of Education. Retrieved from https://edpolicy.stanford.edu/sites/default/files/events/materials/2011-06-linked-learning-performance-based-assessment.pdf

From the introduction: "More than standardized tests of content knowledge, performance-based tasks are able to measure students' habits of mind. Performance-based assessment requires students to use high-level thinking to perform, create, or produce something with transferable real-world application. Research has shown that such assessment provides useful information about student performance to students, parents, teachers, principals, and policymakers. Research on thinking and learning processes also shows that performance-based assessment propels the education system in a direction that corresponds with how individuals actually learn."

Tucker, C. G. (2015). *Psychometric considerations for performance assessment with implications for policy and practice.* Princeton, NJ: Center for K–12 Assessment & Performance Management at ETS. Retrieved from https://www.ets.org/research/policy_research_reports/publications/report/2015/jubd

From the introduction: "Performance assessments are used to measure performance in education, work, and everyday life. Perhaps the public's most commonly experienced performance assessment is the driver's examination—a combination of multiple-choice questions probing knowledge of driving laws (etc.) and performance tasks measuring actual driving under real-world conditions. In education, content standards for guiding K–12 education systems have been revised recently to better support the preparation of our students for the current and future expectations of post-secondary college and career. States are likewise collaborating to develop and implement common assessment systems with a corresponding focus on critical-thinking, problem-solving, and analytical skills, a focus that is increasingly bringing performance assessment into mainstream K–12 educational assessment."

Additional Organizations to Consult

The Center on Standards & Assessment Implementation: http://www.csai-online.org/

From the website: "CSAI provides state education agencies (SEAs) and Regional Comprehensive Centers (RCCs) with research support, technical assistance, tools, and other resources to help inform decisions about standards, assessment, and accountability."

Next Generation Science Assessment: http://nextgenscienceassessment.org/

From the website: "The Next Generation Science Assessment (NGSA) group is a multi-institutional collaborative that is applying the evidence-centered design approach to create classroom-ready assessments for teachers to use formatively to gain insights into their students' progress on achieving the NGSS performance expectations."

Additional Ask-A-REL Responses to Consult

Ask-A-REL Northeast & Islands at EDC. (2013). *What does the research say about performance-based assessments?* Retrieved from http://www.relnei.org/wp-content/referencedocs/RELNEI_RD0209_Performance_Based_Assessments.pdf

Methods

Keywords and Search Strings

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

- (Performance-based assessment OR performance assessment) AND (design OR implement*)
- (Performance-based assessment OR performance assessment) AND elementary AND (math OR science)

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 the most current information (i.e., within the last five years), except in the case of nationally known seminal resources.
- Search priorities of reference sources: Search priorities include IES, nationally funded, and certain other vetted sources known for strict attention to research protocols. 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 May 18, 2017. URLs, descriptions, and content included here were current at that time.