International Society of the Learning Sciences

Insights from Using a Systematic Design Process to Develop Classroom-based Assessment Resources for Measuring Elementary Students' Science and Literacy Proficiencies

Gary Weiser, WestEd, gweiser@wested.org
Alison K Billman, The Lawrence Hall of Science, alison_billman@berkeley.edu
Christopher J Harris, WestEd, christopher.harris@wested.org
Lauren M Brodsky, The Lawrence Hall of Science, brodsky@berkeley.edu
Daniel Damelin, Concord Consortium, ddamelin@concord.org

Abstract: The *Framework* and NGSS bring to the forefront the role of language in doing science and in learning from doing science. Yet, most existing science assessments for elementary learners do not integrate or attend to aspects of scientific language and literacy that are essential components of science proficiency. Accordingly, there is a need for high-quality classroom resources capable of providing teachers with information on where students are in the process of developing proficiency in science and scientific-specific literacy. In this paper, we present our design process for developing assessment resources – tasks, rubrics, and teacher support materials – and describe initial insights from using the process to develop resources for elementary grade classrooms. We highlight challenges and opportunities that have emerged from our work attending to the role of literacy assessing multi-dimensional science learning, incorporating equity and inclusion considerations in assessment design, and embracing the affordances of technology.

Introduction

High-quality, classroom-based assessment resources can provide critical information that helps teachers in making informed instructional decisions to further student science learning (Shepard et al., 2018). However, the potential of classroom-based assessments to increase students' opportunities to learn can only be realized if students can access and engage with tasks productively and if accompanying resources provide teachers with insights on students' developing proficiencies. This sets a high bar for knowledge-in-use assessments that aim to measure students' multidimensional proficiencies as articulated in science education frameworks and standards such as the National Research Council's Framework for K-12 Science Education (henceforth *Framework*; NRC, 2012) and the accompanying Next Generation Science Standards in the United States (NGSS Lead States, 2013). The challenge is especially acute when designing assessment resources for early elementary students who are just beginning to develop the cognitive endurance, attention span, physical dexterity, and language and literacy skills they must have to demonstrate what they know and can do (Billman et al., 2021). For example, as young students use and apply their knowledge to understand and explain phenomena, they must learn science-specific ways of talking, reading, and writing to communicate findings and provide explanations (Billman & Pearson, 2018). Unfortunately, most existing science assessments for elementary learners do not integrate or attend to aspects of scientific language and literacy that are essential components of science proficiency. Accordingly, there is a need for classroom resources capable of providing teachers with information on students' developing proficiencies in science and science-specific literacy.

In this paper, we describe how we use the Next Generation Science Assessments for Young Scientists (NGSA-YS) design process to fill the need for equitable, technology-based assessment tasks that integrate science and literacy proficiencies appropriate for elementary students. The NGSA-YS process (Billman et al., 2021) was specifically developed to meet the challenge of measuring students' integrated science and literacy proficiencies by following the argumentative reasoning of evidence centered design (ECD; Mislevy & Haertel, 2006) to create classroom-based assessment tasks that support NGSS teaching and learning. In accordance with the principles of ECD, NGSA-YS is a systematic design and development process that builds from the Next Generation Science Assessment (NGSA) approach (Harris et al., 2019; Harris et al., in press). NGSA-YS aims to result in coherent assessments by unpacking the science and literacy dimensions to be measured, developing evidence statements that describe student performances, and creating tasks that measure specific knowledge and skills, which taken all together form a basis of an argument about what students know. After describing the design process, we highlight challenges and opportunities that have emerged from our work.

Theoretical framework

Three-dimensional science learning and the integrated role of literacy

The vision of science learning in the *Framework* emphasizes three dimensions – disciplinary core ideas, crosscutting concepts, and science and engineering practices – and the idea of knowledge-in-use where students use and apply the dimensions to explain phenomena and/or design solutions to problems. Importantly, the *Framework* also brings to the forefront the role of language in doing and learning science. For example, students develop proficiency in scientific ways of talking, reading, and writing when they engage in the science and engineering practice of argumentation to construct claims based on evidence or to exchange ideas that advance, improve, and defend explanations. Likewise, applying the crosscutting concept of systems and system models requires students to generate coherent narrative accounts of science phenomena, linking ideas and describing relationships among components in systems. Recent research shows that integrated science and literacy instruction increases student learning in both domains (Billman & Pearson, 2018; Maarouf, 2019). It follows that classroom-based assessment resources should also attend to this integration to best support science instruction.

Equity and inclusion considerations in the design process

The *Framework* makes the call that science education must be for all students. The design of classroom-based, knowledge-in-use assessments must likewise heed that call, ensuring that assessments are inclusive of all students and are equitable in the measure of all students' proficiency (NRC, 2014). Recent guidance for designing knowledge-in-use assessments that attend to and promote inclusion of all learners includes recommendations that tasks use scenarios and phenomena that are likely to have interest or universal relevance for students, explicitly attend to accessibility of language, include scaffolds to make expectations explicit for students, and be accompanied by tools and routines to support teacher enactment (Furtak et al., 2020). This is an emerging area in NGSS-aligned assessment, where researchers are dedicating attention to how assessments can and should support meaningful science learning for a wide range of students.

Affordances of technology for high-quality formative assessment

Technology-delivered assessments can provide unique opportunities for students to demonstrate knowledge-in-use, exhibit emerging literacy proficiency, and can also provide equitable and inclusive ways for students to access and engage with tasks. Related to this last point, technology can provide options for students to use multiple modalities and representations, affording students with diverse abilities and language backgrounds opportunities to demonstrate their proficiencies that are not possible in print-based assessment (Henderson et al., 2021). For teachers, technology can make regular use of formative assessment more feasible by providing support for implementation, data collection, interpretation, and instructional next steps (Zhai et al., 2020).

The NGSA-YS design process

The NGSA-YS design process consists of four iterative steps (Figure 1). Each step includes activities that achieve coherence across the science and literacy proficiencies underlying an NGSS performance expectation (PE), that incorporate guidance for equity and inclusion, and tend to the design of technology-based affordances.

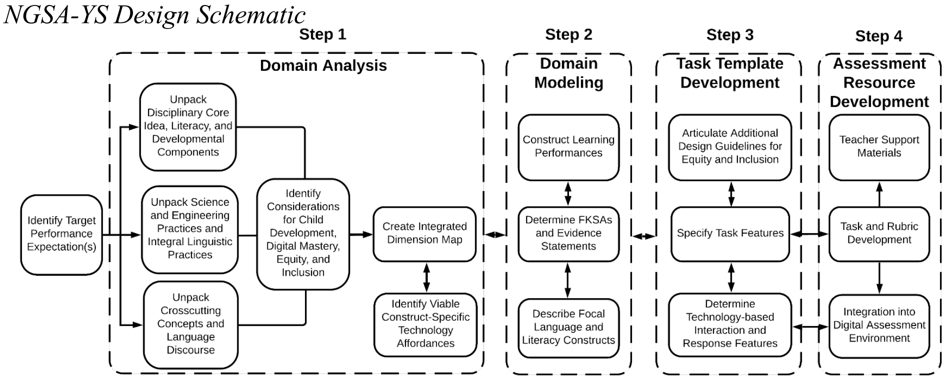
The first step is a purposeful analysis of the domain covered by a PE or PE bundle. In this step, we identify the focal components of each dimension including the science-specific language and aspects of literacy required to demonstrate proficiency as well as the aspects of language and literacy that may inhibit students' ability to demonstrate proficiency. Information obtained through the domain analysis is brought back together in an *integrated dimension map* (IDM) that illustrates relationships between each aspect of the three dimensions of the NGSS called out in the PE along with key elements of science-specific language and literacy. The IDM helps identify ways in which the three dimensions can work together to define proficiency and lays the groundwork for specifying the knowledge, skills, and abilities that will be measured. Alongside our IDMs, we set forth an identification of viable technology affordances specific to the constructs articulated in the mapping.

The second step, domain modeling, uses the IDM to articulate and refine the multiple sub-performances (referred to as learning performances) that students need to demonstrate to meet a PE or PE bundle. A single learning performance is crafted as a knowledge-in-use statement that is smaller in scope and covers a designated area of an IDM. Multiple learning performances are needed to fully cover the conceptual terrain of a PE. Learning performances are a keystone in the process of developing tasks that assess for building toward a PE.

The design goal of **the third step** is to develop task templates – detailed design tools that guide the creation of families of tasks and rubrics. Task templates articulate information about how the task is presented to the student, what content is covered in the task, options for how students can respond, and preliminary scoring guidelines. Templates also bring attention to language and literacy features, to equity and inclusion considerations, and to intended technology interactions such as how students will interact with any digital components as part of the task activities or response modes.

The final step involves operationalizing the assessment argument and specifications from the task template to develop tasks, scoring rubrics, associated digital assessment environments, and teacher resources. In general, each task consists of a set of prompts tied together by a common, developmentally appropriate scenario that establishes a relevant phenomenon or context for student problem solving aligned to the associated learning performance. The prompts embedded within the task ensure that there are opportunities for students to demonstrate integrated proficiency with the dimensions of the learning performance and with science-specific language and literacy. Rubrics target the science and science-specific literacy aspects of student responses that should be scored and show how they relate to proficiency with the learning performance. To complement the assessment tasks and rubrics, we develop materials to support teachers' use of the assessments and their facility with implementing formative assessment practices in the midst of science instruction. A key resource is a practice guide to support administration of the tasks, scoring and interpretation of student responses, and decision-making for instructional next steps.

Figure 1
NGSA-YS Design Schematic



Challenges and opportunities in using the NGSA-YS process

Our use of the NGSA-YS process revealed new challenges in addressing the needs of teachers and students. It also revealed new opportunities to push the assessment development field forward in attending to literacy and equity within multi-dimensional science assessment design while realizing the affordances of technology.

All students, especially in the early grades, benefit from opportunities to practice and develop a broad range of language and literacy skills in the context of engaging with the NGSS. In our work, we recognize that these opportunities must also extend to classroom-based assessment. Consequently, it is important in assessment design to identify measurable components of language and literacy proficiency that are essential elements of the target science performance. One challenge we wrestled with was determining the degree to which students can express their understandings with everyday language while recognizing that science has domain-specific words, some that are essential for engaging with an idea or concept, but which students are unlikely to encounter outside of the classroom (Osborne et al., 2016). For example, "push and pull" is sometimes sufficient for expressing ideas about forces, but the domain-specific use of "forces" can include additional meanings. We have found, too, that it is of value to identify general language and literacy proficiency targets. For example, student proficiency with writing within an organizational structure can be targeted (e.g., students' ability to describe the steps in an investigation). This adds new challenges in unpacking an NGSS PE, but also improves the instructional supportiveness of the assessment in both science and language arts domains.

Equity is interwoven throughout our work using the NGSA-YS process. In attempting to follow the emerging guidance on equity and inclusion, we have wrestled with early grade students' lived experiences and developmental profiles in relationship to expectations put forth in the NGSS. For example, providing students with different options when developing and using models without neglecting the conventions and expectations in the design and format of visual representations in the scientific community. Additionally, in identifying phenomena and constructing task scenarios that are relevant to students' lived experiences, we have wrestled with the tension between scenarios that have universal relevance and utility for a wide range of students and those that leverage the specific identities of learners.

As we work to design assessments that will be delivered on a digital platform, we have focused on affordances that fall into two categories: (1) general enhancements applicable to all tasks; and (2) interactive technology-based components directly related to an assessment target. General enhancements include building text-to-speech capabilities to facilitate students' reading of the assessment tasks and speech-to-text capabilities to assist students in recording their responses in ways that do not require keyboarding skills. Interactive components include design patterns that allow students to express ideas in modalities other than writing and to engage in science practices in ways not possible with paper-based assessments. These might be simulations that students manipulate and annotate, tools that facilitate construction of a model diagram, or drag-and-drop interactives that provide a context-specific opportunity to express concepts and ideas with both words and images. Evaluating when affordances from these two categories are useful in a particular context has raised new questions regarding what elementary students can do within a digital system and what supports they need.

Conclusion

Our use of the NGSA-YS approach has contributed to our understanding of knowledge-in-use assessment design for the early elementary grades by focusing attention on the science-specific language and literacy practices that are integral parts of multi-dimensional science learning. Importantly, considerations of language and literacy inform the tasks, the rubrics, and the teaching resources we develop. New technology affords opportunities for students of wide-ranging proficiency to showcase what they know and can do. These insights fit into an emerging appreciation of the integrated role of science and literacy that push our assessment work forward in designing instructionally beneficial assessments for elementary teachers and their young students.

References

Billman, A. K., & Pearson, P. D. (2018). *First grade second language: Uniting science knowledge and literacy development for English language learners (Technical Report).* Berkeley, CA: Regents of UC.

Billman, A. K., Rutstein, D., & Harris, C. J. (2021). *Articulating a transformative approach for designing tasks that measure young learners' developing proficiencies in integrated science and literacy.* Berkeley, CA: The Lawrence Hall of Science, University of California, Berkeley.

Furtak, E., Kang, H., Pellegrino, J., Harris, C., Krajcik, J., Morrison, D. …Wingert, K. (2020). Emergent design heuristics for three-dimensional classroom assessments that promote equity. In M. Gresalfi & I. S. Horn (Eds.), *The Interdisciplinarity of the Learning Sciences, 14th International Conference of the Learning Sciences (ICLS) 2020*, Volume 1. Nashville, Tennessee: International Society of the Learning Sciences.

Harris, C. J., Krajcik, J. S., Pellegrino, J. W., & DeBarger, A. H. (2019). Designing knowledge-in-use assessments to promote deeper learning. *Educational Measurement: Issues and Practice*, *38*(2)53–67.

Harris, C. J., Krajcik, J. S., & Pellegrino, J. W. (Eds.) (in press, 2022). *Creating and using instructionally supportive assessments in NGSS classrooms.* NSTA Press.

Henderson, J. B., Zillmer, N., Holton, A., Weiner, S., Greenwald, E., Goss, M. ...McNeill, K. L. (2021). How Science Teachers DiALoG Classrooms: Towards a Practical and Responsive Formative Assessment of Oral Argumentation. *Journal of Science Education and Technology*, *30*(6), 803–815.

Maarouf, S. A. (2019). Supporting academic growth of English language learners: Integrating reading into STEM curriculum. *World Journal of Education*, *9*(4), 83.

Mislevy, R. J., & Haertel, G. D. (2006). Implications of evidence-centered design for educational testing. *Educational Measurement: Issues and Practice*, *25*(4), 6–20.

National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. Washington, DC: National Academies Press.

National Research Council. (2014). *Developing assessments for the Next Generation Science Standards.* National Academies Press.

NGSS Lead States. (2013). *Next Generation Science Standards: For states, by state*s. National Academies Press.

Osborne, J., Sedlacek, Q. C., Lemmi, C., & Friend, M. (2016). Learning to Read Science. *Science Scope*, *40*(3), 36–42.

Shepard, L. A., Penuel, W. R., & Pellegrino, J. W. (2018). Using learning and motivation theories to coherently link formative assessment, grading practices, and large-scale assessment. *Educational Measurement: Issues and Practice*, *37*(1), 21–34.

Zhai, X., Yin, Y., Pellegrino, J. W., Haudek, K. C., & Shi, L. (2020). Applying machine learning in science assessment: a systematic review. *Studies in Science Education*, *56*(1), 111–151.

Acknowledgements

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A210290. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education. We gratefully acknowledge the foundational work of the Next Generation Science Assessment Project.