Three-Dimensional Assessment of NGSS Upper Elementary Engineering Design Performance Expectations

Kevin W. McElhaney, Satabdi Basu, Tallie Wetzel, Jared Boyce
Education Division, SRI International

Paper presented at the 2019 NARST Annual International Conference

Abstract

The *Framework for K-12 Science Education* and the Next Generation Science Standards (NGSS) articulate a vision for integrating engineering with the natural science disciplines of physical, life, and Earth science. We describe a principled approach using evidence-centered design for developing assessment tasks aligned with the NGSS engineering design performance expectations (PEs) in either a *science-independent* or *science-bundled* way. We illustrate the design process using example tasks developed to measure student learning outcomes from a curriculum unit integrating upper elementary Earth science and engineering PEs. We present preliminary empirical reliability evidence for a science-bundled assessment based on a classroom pilot study. This work informs the design of summative and formative NGSS-aligned engineering assessments and illustrates how the use of a systematic design process can help ensure the alignment of assessment tasks to specific NGSS PEs.

Three-Dimensional Assessment of NGSS Upper Elementary Engineering Design Performance Expectations

Problem Statement

Supporting engineering design instruction at the K-12 level is necessary in order to meet the nation's increasing demand for science and engineering competency and to increase the diversity of the pipeline of future technology innovators. K-12 engineering education is gaining traction in the U.S. largely because of its inclusion in the *Framework for K-12 Science Education* (NRC, 2012) and in the Next Generation Science Standards (NGSS) (NGSS Lead States, 2013). The *Framework* characterizes engineering design as involving problem definition, specifying criteria and constraints for acceptable solutions, generating and evaluating multiple solutions, and testing and optimizing solutions. It also involves the application of science disciplinary knowledge to the solution of engineering problems.

Gorin and Mislevy (2013) and Scalise (2014a, 2014b) have noted multiple challenges for next-generation science assessment design and validation, including complex domain definitions and complex student performances. Assessing the NGSS engineering design performance expectations (PEs) present particular challenges. Although students typically engage in engineering in the context of one or more of the natural science disciplines (Earth, life, or physical science), assessment of the NGSS engineering design PEs is possible in a natural science context (*science-bundled*) or independent of the natural science discipline (*science-independent*). This paper describes our principled approach for developing NGSS-aligned engineering assessment tasks in either a science-bundled or science-independent way. The science bundled tasks integrate science disciplinary content in Earth science for the upper elementary grade band. We present preliminary empirical validity evidence for an assessment that included three tasks, based on a classroom pilot study.

Background

As the foundation for the NGSS, the *Framework* identifies engineering design as a discipline distinct from the earth, life, and physical sciences. Accordingly, engineering design has a unique set of disciplinary core ideas (DCIs) addressing (1) the definition and delimiting of engineering problems relative to criteria for success and design constraints, (2) the generation and evaluation of multiple design solutions to determine which best meet problem criteria and constraints, and (3) the optimization of design solutions via systematic testing, refining, and making trade-offs (NRC, 2012). At the upper elementary school grade band, the NGSS instantiates these engineering DCIs in three PEs (3-5-ETS1-1, 3-5-ETS1-2, 3-5-ETS1-3) that integrate engineering DCIs with engineering practices. These PEs do not explicitly include natural science content and, at the elementary level, do not include crosscutting concepts.

In addition to identifying engineering as a discipline distinct from science, the *Framework* clearly articulates a vision of integrating science and engineering, enabling students to "explore the practical use of science" and providing "a context in which students can test their own developing scientific knowledge and apply it to practical problems" (NRC, 2012, p. 12). As such, NGSS-aligned engineering instruction typically occurs in the context of one of the natural science disciplines and will frequently involve students using disciplinary science knowledge (e.g. science DCIs) as the basis for engineering design decisions. NGSS-aligned engineering assessments therefore will often require an associated natural science context to measure how well students are able to apply science disciplinary knowledge to the solution of engineering problems.

Alternatively, however, stakeholders such as curriculum designers or high stakes assessment developers may seek evidence for how well students achieve proficiency with the

NGSS engineering PEs more broadly. For instance, engineering instruction should ideally promote students' engineering proficiency in a way that can be generalized beyond the specific disciplinary context of the curriculum, such as in everyday engineering design contexts.

The process of evidence-centered design (ECD) (Mislevy & Haertel, 2006) provides a conceptual framing for analyzing complex domains to be assessed. In this work, we use it to specify the essential components of the NGSS engineering PEs. ECD helps structure the connections among the targeted student proficiencies, task design features, and observable evidence. Assessment designers can use ECD to demonstrate how particular student performances provide evidence for students' proficiency on particular constructs. This up-front specification of the design framework promotes systematicity in task design and scoring, promoting coherence among the rubrics, task features, and evidence of student proficiency. In this work we build on an approach for designing NGSS-aligned assessment tasks in physical and life science (Harris, Krajcik, Pellegrino, & McElhaney, 2016).

Assessment task and rubric design

Design Approach. We designed engineering assessment tasks to align with three upper elementary science and engineering PEs, 3-5-ETS1-1, 3-5-ETS1-2, and 3-5-ETS1-3. The tasks are designed to be used in a pretest-posttest assessment of students who studied a science and engineering curricular unit on urban water runoff. In the unit, students investigate the science phenomenon of urban water runoff caused by the use of impermeable building materials and develop and test engineering solutions to reduce these impacts.

We illustrate our design approach in Figure 1. After identifying the NGSS PEs to be assessed, we first unpack and structure the upper elementary engineering domain. On the basis of the unpacking, we articulate assessable performance statements for PEs that are too broad to be assessed with a single assessment task scenario. We then articulate evidence statements for each performance statement and determine task design features that are likely to elicit this evidence from students. Finally, we develop the tasks and rubrics that include the design features.

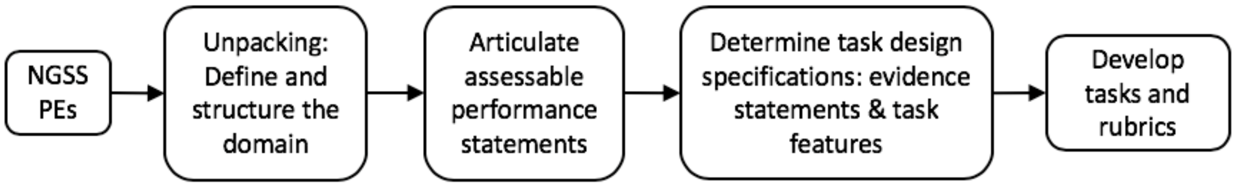

Figure 1. Design process for aligning assessment task and rubric design to NGSS engineering PEs.

Unpacking. We first unpacked our target NGSS upper elementary engineering PEs. In ECD, unpacking entails articulating the boundaries and structure of the domain to be assessed, including identifying key aspects of the DCIs and practices, elaborating and documenting the meaning of key terms, defining expectations for understanding at the upper elementary school level, determining boundaries for content knowledge, and identifying background knowledge and skill that is expected of students to develop a grade-level-appropriate proficiency. Table 1 summarizes the key aspects of our target engineering PEs identified by our domain analysis.

Table 1. Summary of domain analysis of NGSS PEs 3-5-ETS1-1, 3-5-ETS1-2, and 3-5-ETS1-3.

Engineering PE	Engineering DCI	Engineering practice	Aspect	Elaboration
3-5-ETS1-1	ETS1.A: Defining and Delimiting Engineering Problems	Practice 1: Defining problems	Statement of problem or need	Problem responds to need or desire of a person, community, or society.
			Identify constraints and criteria for success	Constraints include materials, resources, tools, safety, ethics, legal. Success criteria include ways to evaluate the solution based on performance
3-5-ETS1-2	ETS1.B: Developing Possible Solutions	Practice 6: Designing solutions	Generating solutions	Determine design features that address the problem and various design constraints or criteria
			Comparing solutions	Compare the merits of one design solution to another, relative to criteria and constraints of the problem
3-5ETS1-3	ETS1.B: Developing Possible Solutions; ETS1.C: Optimizing the design solution	Practice 3: Planning and carrying out investigations	Planning fair tests	Determine a procedure that controls variables in order to improve a design
			Identify failure points	Based on test data, identify design features that lead to design failure

Articulating assessable performance statements. Following the unpacking, we articulated integrated performance statements for the PEs that we deemed were too broad to assess using a single assessment task scenario. We designed one of the tasks to align with the PE 3-5-ETS1-1 as it appears in the NGSS, but decomposed each of the PEs 3-5-ETS1-2 and 3-5-ETS1-3 into two separate performance statements, each addressed by a separate task scenario. These performance statements are listed in Table 2 alongside the NGSS engineering PEs from which they were derived. Our assessment tasks are aligned with these performance statements and reflect the key aspects of the engineering domain as identified in our unpacking.

Table 2. Target NGSS Engineering PEs and integrated performance statements derived from those PEs.

NGSS Engineering PE	Integrated assessable performance statement
3-5-ETS1-1. Define a simple design problem reflecting a need or a want that includes specified criteria for success and constraints on materials, time, or cost.	• Same as the PE
3-5-ETS1-2. Generate and compare multiple possible solutions to a problem based on how well each is likely to meet the criteria and constraints of the problem.	• Generate a solution to a problem given criteria and constraints. • Compare multiple possible solutions to a problem based on how well each is likely to meet the criteria and constraints of the problem
3-5ETS1-3. Plan and carry out fair tests in which variables are controlled and failure points are considered to identify aspects of a model or prototype that can be improved.	• Plan fair tests in which variables are controlled in a way that can identify aspects of a design that can be improved. • Based on data from fair tests, identify failure points that can inform improvements to a design.

Determining task design specifications. For each performance statement (corresponding to a task scenario), we determined task design specifications that guide the design of aligned assessment tasks and rubrics. We focused on three types of design specifications that provide the basis for aligning task and rubric design to a performance statement: (1) evidence statements, which describe features of student responses that constitute evidence of proficiency with the performance statement, (2) characteristic task features, which must be included in a particular type of task to ensure that the task can elicit the target proficiency, and (3) variable task features, which can shift the difficulty or focus of a task. Table 3 illustrates example design specifications for assessment tasks aligned with the NGSS PE 3-5-ETS1-2.

In these design specifications, the science disciplinary context constitutes a variable feature of the task. A disciplinary core idea from the science discipline can be integrated with the two dimensions of the engineering performance expectation to produce a three-dimensional assessment task. To ensure appropriate integration of science and engineering in the task, the task design must require the disciplinary core idea to be used in order to make engineering design decisions required in the task.

Table 3. Examples of design specifications for assessment tasks aligned with the NGSS PE 3-5-ETS1-2 (comparing solutions).

Performance Statement	Compare multiple possible solutions to a problem based on how well each is likely to meet the criteria and constraints of the problem.
Evidence Statements	• Student identifies which of multiple solutions better meets given criteria and/or constraints • Student provides correct rationale for their solution choice
Characteristic Features	• Provides design problem context familiar to students of target grade band • Presents 2 or more design constraints and/or criteria • Presents 2 or more potential solutions to the problem • Prompt to choose the most appropriate solution • Prompt to explain the rationale for choice of solution
Variable features	• Design problem context (science or everyday context) • Number and types of design variables involved • Number and types of constraints presented

Task development. Based on the design specifications, we developed both science-bundled and science-independent versions of the tasks to be used as part of classroom studies of the aforementioned elementary curriculum unit. Figure 2 illustrates the science-independent comparing solutions task aligned with the NGSS PE 3-5-ETS1-2. The task is designed to be used to support claims about students' proficiency with the NGSS PE in an everyday problem context. The task includes the characteristic features listed in Table 3 and employs a science-independent problem context accessible to elementary students that does not require science disciplinary content knowledge.

Figure 3 illustrates an example science-bundled task. The task is designed to be used in conjunction with a science and engineering curriculum unit on the impacts of urban water runoff, to support claims about how well students are able to integrate engineering proficiency with a specific science disciplinary context addressed in their science instruction. The task integrates the DCI ESS3.C: Human Impacts on Earth Systems with the engineering DCI and practice of the NGSS PE 3-5-ETS1-2, resulting in a three-dimensional task.

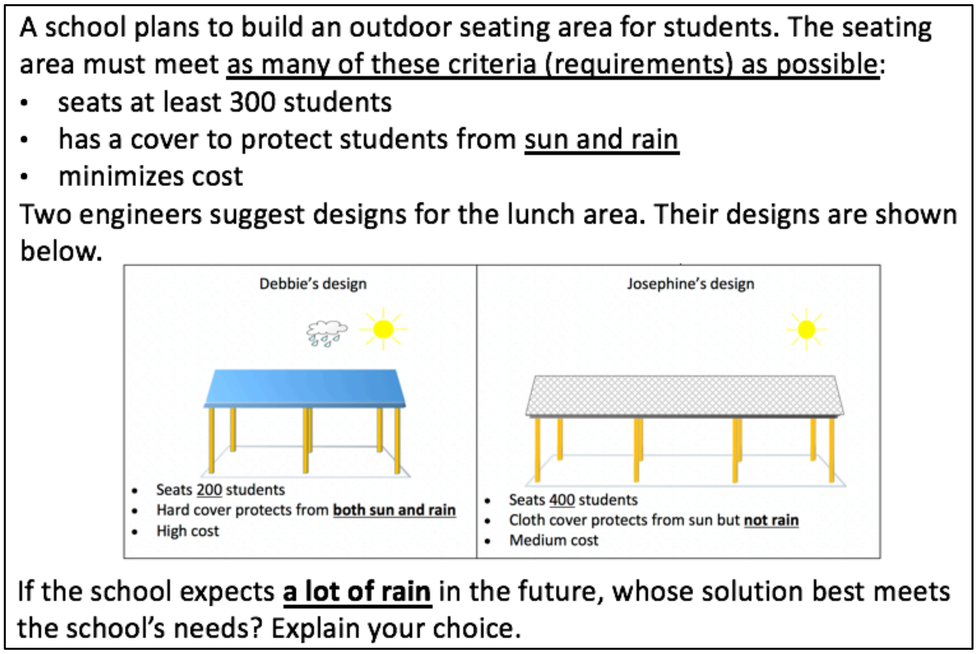

Figure 2. Science-independent assessment task for comparing multiple solutions.

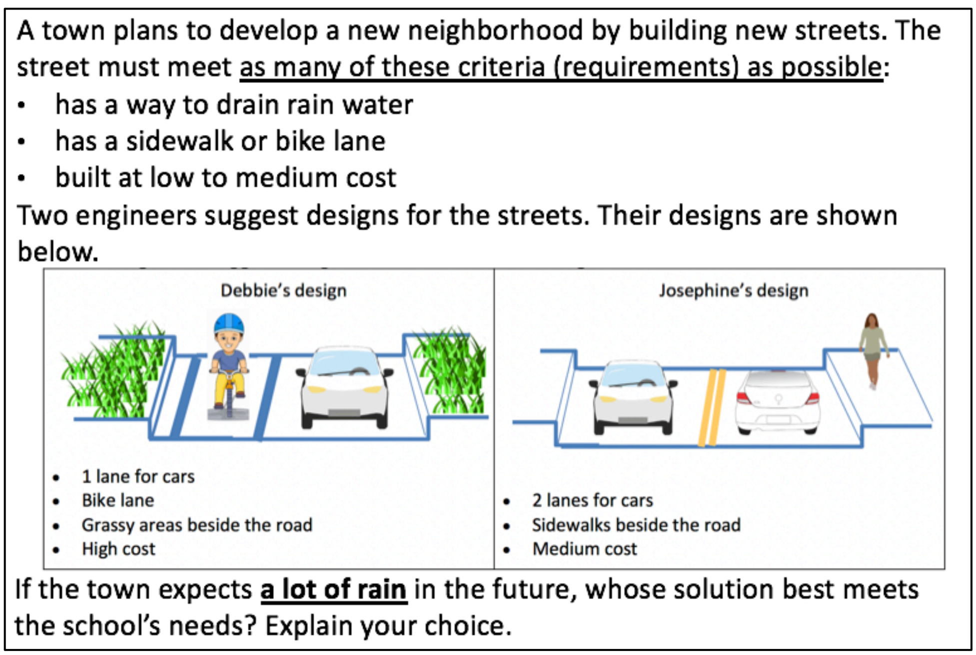

Figure 3. Three-dimensional, science-bundled assessment task for comparing multiple solutions.

The two tasks illustrated in Figures 2 and 3 are similarly constructed, as they arise from the same set of assessment task design specifications. The tasks differ only in the variable task features concerning the problem context and the number and types of design variables and constraints presented to students.

Rubric design. We designed rubrics to align with evidence statements articulated for a particular performance statement. For example, in the comparing solutions tasks shown in Figures 2 and 3, rubrics reward the extent to which students can justify their choice of design by appropriately describing both (1) how a design meets given design criteria and (2) acknowledging how a design does not meet those criteria. For the science-bundled tasks, rubrics include a score level indicating whether students appropriately applied science content knowledge about water runoff to their engineering decisions. For example, in response to the task in Figure 3, a student must indicate that Debbie's design has grassy areas that help drain water in order to achieve the maximum score on the task.

Preliminary empirical validity evidence

We developed two parallel versions (A and B) of a paper-and-pencil assessment, each consisting of three task scenarios (aligned with 3-5-ETS1-1 and 3-5-ETS1-2). Each task aligned with one of the integrated performance statements in Table 2. The assessment combined multiple choice and constructed response sub-tasks. Altogether, the three tasks contained a total of 9 individual sub-tasks (4 for Task 1, 1 for Task 2, and 4 for Task 3).

Classroom pilot study. We administered the assessments as both pretest and posttest with students of one 5th grade teacher and two 6th grade teachers who used the curriculum unit. In all, 123 5th grade students and 274 6th grade students took the assessments. 51% of the participating district's students belong to underrepresented cultural groups in STEM.

Scoring. We scored responses of students whose teachers were randomly assigned version A of the assessment as both the pretest and the posttest (N = 107). Multiple scorers were trained in the application of the rubrics. Scorers achieved at least 85% agreement for each sub-task on a 20% subset of the student responses. Discrepancies were resolved by discussion.

Analysis. We used students' posttest scores for the current analysis because of its expected wider range of student responses compared to the pretest. In this way, we can understand properties of the assessment with an appropriate range of student ability. Seven students from our sample took the pretest only, leaving us with an analysis sample of N = 100. We computed descriptive statistics and conducted initial reliability analyses at the task and sub-task levels. We did not conduct a factor analysis given the sample size relative to the number of tasks and codes.

Overall, students' posttest scores ranged between 0 and 13 points, with a mean score of 6.60 (SD = 3.56) out of a maximum possible score of 18. This finding indicates that the assessment was somewhat difficult for students, though it was able to detect statistically significant pretest to posttest learning gains, as described in a parallel conference paper (Chiu et al., 2019). This level of difficulty of the assessment was expected because students had only 2-3 weeks of instruction on the topic and engineering has been relatively recently introduced in elementary science curricula. Students' scores were distributed somewhat evenly between 0 and 13, indicating that the assessment shows promise at distinguishing multiple levels of integrated science and engineering proficiency. We also examined the distribution of students' scores at the task level (Table 4). These results further indicate that the individual tasks also show promise at effectively distinguishing levels of student proficiency. Moreover, Task 2 is easier than the other two tasks, providing variation in difficulty level across different tasks and potentially enabling the assessment to discriminate among students at the low end of proficiency.

Table 4: Distribution of Task Scores

Task Score	% Students w/ Score		
	Task 1	Task 2	Task 3
8.0	–	–	0%
7.0-7.5	–	–	0%
6.0-6.5	0%	–	2%
5.0-5.5	1%	–	15%
4.0-4.5	9%	24%	9%
3.0-3.5	24%	19%	30%
2.0-2.5	27%	20%	16%
1.0-1.5	23%	11%	2%
0.0-0.5	16%	26%	26%

Cronbach's alpha for all nine scored subtasks is 0.71, which is considered acceptable for an assessment designed to measure a single latent construct—in this case, integrated science-engineering proficiency. We expected alpha to lie on the low end of the established acceptability range (0.7-0.8), and we conjecture that it reflects each task measuring a different aspect of students' science-engineering proficiency (e.g., defining problems, generating solutions, comparing solutions). Further analysis such as factor analysis is required to best understand the internal structure of the assessment.

Implications and Contributions

The *Framework* and NGSS articulate a vision of science proficiency that includes engineering as a discipline alongside physical, life, and Earth Science. In this vision, students apply disciplinary science content toward the solution of engineering problems. The science education research community requires reliable, validated assessments that are aligned with engineering design PEs and that can integrate natural science disciplinary core ideas. Assessments of the NGSS engineering PEs may vary in the extent to which a specific science discipline is included. Depending on the hypotheses stakeholders may wish to test, assessments may need to be either science-bundled or science-independent. This work is of high interest to the NARST community because of its potential to inform the design of high-quality NGSS-aligned engineering assessments. Summative assessments are needed to measure student learning outcomes from NGSS-aligned engineering instruction and evaluate NGSS-aligned engineering curriculum interventions. Teachers require classroom-based, instructionally-supportive assessments that help them gauge students' progress toward the engineering PEs. We illustrate an approach for designing assessments that are appropriate for making either discipline-specific or discipline-general claims about students' NGSS-aligned engineering proficiency.

Moreover, this work illustrates the use of a principled assessment design approach that ensures the alignment of tasks to specific NGSS PEs. The use of a systematic design process constitutes part of a strong validity argument for assessments, especially those addressing complex domains and requiring complex student performances (NRC, 2014). Our future work will entail gathering of additional validity evidence and parallel-forms reliability evidence, as well as creating assessment design specifications that can be used to develop new assessments that integrate engineering PEs with other disciplinary concepts.

References

Chiu, J.C., McElhaney, K.W., Zhang, N., Biswas, G., Fried, R., Basu, S., & Alozie, N. (2019, April). *A Principled Approach to NGSS-aligned Curriculum Development Integrating Science, Engineering, and Computation: A Pilot Study.* Paper presented at the NARST Annual International Conference, Baltimore, MD.

Gorin, J.S., & Mislevy, R.J. (2013). *Inherent measurement challenges in the next generation science standards for both formative and summative assessment.* Commissioned paper presented at the K-12 Center at ETS Invitational Research Symposium Invitational Research Symposium on Science Assessment, Washington, DC.

Harris, C. J., Krajcik, J. S., Pellegrino, J. W., & McElhaney, K.W. (2016). *Constructing assessment tasks that blend disciplinary core ideas, crosscutting concepts, and science practices for classroom formative applications*. Menlo Park, CA: SRI International.

Mislevy, R., & Haertel, G. (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*. Washington, DC: National Academies Press.

NGSS Lead States (2013). *Next generation science standards: For states, by states*. Washington, DC: National Academies Press.

Scalise, K. (2014a) mIRT-bayes as Hybrid Measurement Model for Technology-Enhanced Assessments, National Council on Measurement in Education Annual Meeting, Philadelphia, PA.

Scalise, K. (2014b) *Assessment System Design Options for the Next Generation Science Standards (NGSS) Reflections on Some Possible Design Approaches.* International Research Symposium on Science Assessment. The Center for K–12 Assessment & Performance Management at Educational Testing Service.