Board 159: Development and Initial Findings of a Summative Assessment for High School Engineering Course (Evaluation, Work in Progress)

Dr. Jeanette Chipps, The Johns Hopkins University

Jeannie Chipps is a research assistant at the IDEALS institute at Johns Hopkins. Her interests are in supporting teachers in their efforts to create learning environments that support diverse learners in STEM.

Dr. Stacy S Klein-Gardner, Vanderbilt University

Dr. Stacy Klein-Gardner serves as an Adjunct Professor of Biomedical Engineering at Vanderbilt University. She is the co-PI and co-Director of the NSF-funded Engineering for US All (e4usa) project and Executive Director of e4usa, the non-profit. Dr. Klein-Gardner is a Fellow of ASEE.

Dr. Cathy P. Lachapelle, STEM Education Insights

Cathy is particularly interested in how collaborative interaction and scaffolded experiences with disciplinary practices help children learn science, math, and engineering. Her work on STEM education research projects includes design, evaluation, and effi

Sabina Anne Schill, Colorado School of Mines

Sabina is a Research Associate at Colorado School of Mines working with Get the Facts Out and Teach@Mines. She got her PhD in Environmental Engineering from the University of Colorado Boulder and her BS in Physics from Westminster College in Salt Lake City. Sabina is interested in teaching, engineering education research, and K-12 STEM education.

Development and initial findings of a summative assessment for high school engineering course (Evaluation)

Abstract: The e4usa is a high school engineering course aimed at broadening participation in engineering offered at over XXX schools across the country. This curriculum emphasizes the role of teams in engineering as well as enabling students to engage in authentic engineering design. In the 2021-2022 school year, our team began to develop and offer an end-of-course assessment to explore and communicate student learning in a summative manner. The assessment structure includes a portfolio and a scenario based exam. The portfolio is a team-produced document that describes the design process the team embark in over a period of time. A scenario- based exam is used to assess individual understanding of the engineering design. The scenario-based exam has been piloted several times and we present the initial findings as well as steps for future consideration of summative course assessment in high school engineering courses. This scenario included an authentic engineering problem related to urban design and planning in the suburbs of Nashville, TN. The exam included engineering artifacts designed by members of the team that worked on the actual project, as well as data related to the problem from relevant government agencies and media. Students were given the exam online, in their own classrooms. Teachers were given the opportunity to determine whether to administer the exam in a single long session or multiple shorter sessions. After administering the exam, teachers and students were able to provide feedback. We present the initial findings from the 2022-2023 academic school year. Additionally, we will describe the findings of the design team and feedback provided by teachers. Performance in earlier portions of the test was higher, aligning with teacher reports of students not finishing the exam. Students specifically scored high when asked to identify the problem within the scenario, describe how the engineering design process could be used to address the scenario, identify stakeholders and experts in the scenario, consider the value to society in solving the problem identified, and in choosing appropriate design solutions given a set of design requirements. The initial findings indicated that the exam needed to be shorter, and more aligned in scope to the type of design projects that students engage in during the curriculum. Next steps will be to create a new scenario and implement this with additional instructional scaffolds to tie into the classroom experience.

Introduction

The e4usa curriculum is designed as a single year high school course of study that is offered to students throughout grades 9-12 across the United States. The course curriculum engages students to consider how engineering is human-centered, responsive, intentional, personal, and reflective. The course contains 8 total units, with several potential pathways to teach these units across the course of either a single semester or two semesters. In the first introductory units, students engage in multiple engineering challenges that are supported by their teacher and address specific skills and mindsets that form a basis for future design work that is done more independently to address problems in their own community. Within the curriculum, there are four threads: Discover Engineering, Engineering in Society, Engineering Professional Skills, and Engineering Design.

Past attempts at AP engineering

In 2003, a group of engineering education leaders led by Dr. Leigh Abts approached the College Board with the idea of an AP Engineering course that culminated in the submission of an engineering design portfolio, similar in style to the AP Art Studio course. The College Board correctly told our group that we did not have a standard means of assessing engineering design portfolio work. This led to over twenty years of work to develop the MyDesign Scoring Rubric[1].

Assessment development

e4usa sought to make an end-of-course assessment that was authentic to the practices of engineering. Assessments that are multiple choice or short answer did not meet this goal. While we recognize that there are such multiple choice engineering exams, particularly at higher levels, they do not align well with the curriculum and learning objectives and would therefore not be suitable[2-3]. We decided to continue to follow in the footsteps of the College Board's AP Art Studio course model by asking students to submit an engineering design portfolio. Because engineering design is done by teams and not individuals, the submitted portfolios would necessarily represent the work of the entire design team. This portfolio is scored with the MyDesign Scoring Rubric and submitted through the MyDesign® Learning Management System. We then sought to combine this authentic, community-based engineering work done by a team with a measure that was a unique reflection of an individual student's engineering design expertise. To that end, we designed a test that would be completed in 2.5 hours by each individual student. This exam is scenario-based building upon a real-world problem. Students are provided with chunks of information and asked questions that walk them through an engineering design process. Once those questions are submitted, the responses cannot be changed. The student is then given additional information that builds the scenario out further and asks additional questions. The exam is aligned with the green- and yellow-threads of the e4usa course that focus on the engineering design process and engineering and society, respectively. We did not assess our red-thread that focuses on connecting with engineering because it would necessarily involve assessment at multiple time points. We also did not assess our blue thread that focuses on engineering professional skills including teamwork, project management, and communication because of the need for multiple time points and classroom observation.

Development of the rubric and grading

Building on the rubric designed for the portfolio, we utilized a 0-5 point scoring system for the scenario-based exam[1]. We listed each question along with its associated content objective to retain focus on the targeted knowledge that the student should be able to demonstrate upon successful completion of the question. Team members collaborated on writing the rubric items in an iterative manner; we wrote drafts individually, discussed the items as a team, and then revised the rubric before repeating the process. To assist in this process, we also took the exam ourselves to better understand how students might approach the problems and what an "expert" answer that completely attended to the questions' content objectives might include. We chose one team member's exam to grade and discussed our scores in a negotiated process to refine the rubric

item(s) for each question (e.g., scope of rubric item(s), difference between scores, etc.). We developed a new component to add to each rubric item: Information for Scorers. This contained details and examples from exemplary responses to the exam that would serve to help the rubric be more accessible to a grader unfamiliar with the scenario-based exam.

To further develop the rubric, we enlisted members of the larger e4usa team and some high school students known to the team and informally completed think-alouds with them as they completed the exam. These results provided a more robust view on how engineering experts and students might answer the exam questions and were used internally to refine the rubric language. We expanded our pool of potential student answers by distributing the exam to a sample of e4usa classes. As our team practiced grading these student responses, we continued to update the rubric by adding more Information for Scorers and reformatting the items to best align with the format of the exam and improve accessibility.

We continued to refine the rubric while training graders and examining scores from the initial distribution of the exams. We received feedback from scorers on the usability of the rubric and the ease of understanding its content. We also utilized scores from graders who graded the same exams to perform inter-rater reliability (IRR) testing from the 2022 exams. The IRR analysis shed light on questions that were not consistently scored. We prioritized these questions for revision.

The iterative process of developing the rubric focused on the rubric content (e.g., what should be in an answer for a student to get a 5, a 4, etc.) and the rubric's language and format. We endeavored to create a clear rubric that would lead to consistent scores and performed updates such as reducing language that might lead to subjective scoring and minimizing jargon. For example, Question 1 was initially one item on the rubric where a student could score a 5 if, "Problem identification, definition, and justification is very clear, with considerable depth and consistently objective detail." We separated this single rubric item into three items to better focus on each aspect of the question, like students' ability to identify a problem: "The specificity of the problem identified is written very clearly and with consistently related to information provided in the problem, which includes a reference to at least 2 issues related to transportation/pollution/Nashville population."

The grading of these exams necessitated a large number of graders, including those who serve on the e4usa team as well as experts with knowledge of education and engineering that were hired and trained externally. The training for graders included reading through both the exam and rubric, attending training sessions, and scoring a practice exam. While the final grading is still in process, inter-rater reliability measures will be gathered to explore reliability of this training process.

Creating a summative score

To create a summative exam score, each question within the exam was aligned to the student learning objectives (SLO) identified within the curriculum. In cases where multiple exam items aligned to a single SLO, the performance on these items was averaged to create a score for each SLO. The SLOs are further categorized according to the threads of the curriculum, and

therefore each student received an overall average score for each thread of the curriculum addressed within the exam. While students have not yet been provided with an overall exam during the pilot phase, our goal will be to combine these exam scores with scores from the design portfolio they will submit to generate a final summative score. This summative score description is not within the scope of the current work, but is also in development.

Results

The first small-scale pilot of this scenario-based exam occurred in the spring of 2022. During this pilot, specific teachers were asked to provide the test to their students to gether initial feedback on the design of the exam. The feedback gathered from this exam largely served to inform content changes to the exam, with a small number of exams going through the full grading process. During the spring of 2023, approximately 700 students took part in the pilot of the exam. Feedback on the exam was provided by both students and teachers in several formats, including questions within the exam itself, surveys, and presentations following the exam. Table 1 presents results from the feedback of students immediately following exam completion. This was an optional component of the exam, and only 370 students completed at least some questions on this survey. Overall, approximately 43% of students found the test interesting and only 42% found the test to be an appropriate length. The highest percentage of agreement for these questions was in regards to whether the exam measured well what was learned within the course, with 59% of students affirming that the exam did measure what was learned in the course.

Table 1. Responses to post-exam survey questions

Question	Yes	No
Did you find this exam interesting?	163	216
Do you think it's an appropriate length and difficulty for a final exam?	157	217
Do you think it measures well what you learned in your engineering class this year?	221	154

The qualitative feedback provided by teachers also indicated that the exam was shorter than the initial pilot but still too long for completion within 1.5 hours by most of their students. Qualtrics was used as the exam platform, and records indicated that on average the exam did take students over an hour. During the 2022-2023 school year, we gathered self-report student demographic data which is summarized in table 2. The students were asked to describe their gender, and provided the opportunity to clarify with an open-ended response to “other” and therefore in providing these results we have used the students self-reported terms where appropriate. The average age of students who responded to these questions was 16.

Table 2

	Number of students
Gender	
Boy	160
Girl	186
Other	6
Non-binary	13
Demigirl	1
Locale	
Suburban	191
Small Town	72
Urban/City	71
Rural	37
Ethnicity	
Hispanic or Latino	
No	301
Yes	74
Race	
White	150
Black or African American	85
Asian	51
Other	38
American Indian or Alaska Native	2
Native Hawaiian or Other Pacific Islander	2
Multiple races reported:	41

In our initial set of analyses, we explored whether there were any relationships between student-reported demographic information and responses to the 3 questions regarding the exam itself. Chi squared analyses did not indicate a significant relationship between the student demographic data with either of the 3 questions regarding the exam itself. Students were also asked to explain their responses about whether they were interested in the exam, difficulty/timing of the exam, and whether the exam measured the knowledge they gained from the course. A set of samples of the positive and negative feedback from these questions are presented in table 3. Coding for these open-ended questions is still in progress.

Table 3 Qualitative responses feedback

Question	Positive feedback	Negative feedback
Interest in exam	*I enjoyed the challenge of building a new train system. It was a lot more fun and engaging than a long quiz asking me a bunch of questions. I also liked that there was a drawing portion.*	*I found the design problem confusing and it jumped from different issues really quickly. I couldn't understand what issue the question was asking me about.*
Connection between exam and course learning	*I do believe this measured it well, this is what we learned in class all year and I was able to correctly apply it to this test.*	*This exam was thoroughly boring and provided too much unnecessary information. As well, some of the instructions were hard to follow and the design pictures did not indicate what was actually the case.*

Conclusions and Next steps

Further analysis of the data set will include utilizing models to determine to what extent the student demographic information predicts either performance on the exam, or responses to the questions about the exam. We also aim to administer a new exam in the 2023-2024 school year which responds to several key findings discussed here. First, we have shortened both the exam and the rubric by removing questions students identified as repetitive. We also utilized a smaller scale problem within the scenario that is developed around an authentic engineering task that was encountered by one of the e4usa instructors in previous years. We will repeat similar analyses of the data from both this new exam to continue improving these course-wide summative assessments. Lastly, we will explore the correlation in performance on specific student learning objectives that are measured by both the portfolio and the end-of-course assessment.

References

1. Author (2022, August). The Engineering Design Process Portfolio Scoring Rubric (EDPPSR)–Initial Validity and Reliability (Fundamental). In *2022 ASEE Annual Conference & Exposition*.
2. FE Exam (http://ncees.org/exams/fe-exam/). National Council of Examiners for Engineering and Surveying (NCEES), Seneca, SC, 2015.
3. Fridley, K. J., & Back, W. E., & Williamson, D. G. (2016, June), *The ASCE BOK, ABET Accreditation Criteria, and NCEES FE Exam - Are They Appropriately Aligned?* Paper presented at 2016 ASEE Annual Conference & Exposition, New Orleans, Louisiana. 10.18260/p.26089