2023 Annual Conference & Exposition
Baltimore Convention Center, MD | June 25 - 28, 2023
The Harbor of Engineering Education for 130 Years
Paper ID #39888

Board 170: PADS – The Performance Assessment of Design Skills (Work in Progress)

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 efficacy research. She also teaches the engineering of design for learning (Learning Engineering!) at Boston College.

Ms. Elizabeth Parry, STEM Education Insights

Elizabeth (Liz) Parry is a partner in STEM Education Insights, a woman owned business developing programs, performing external evaluation and consulting on research and practice in P12 Engineering Education.

A graduate of the Missouri Institute of Science and Technology (Missouri S&T), Elizabeth Parry is a partner in STEM Education Insights, a woman owned consulting company specializing in external evaluation, grant writing, curriculum development, engineering coaching and professional learning for P12 teachers and research.

Previously, Liz held an appointment in various roles in the Dean's Office at the College of Engineering at North Carolina State University. For the past twenty five years, she has worked extensively with students from preschool to graduate school, parents, preservice and in- service teachers and administrators to implement engineering and STEM in meaningful and measurable ways. As the Co-PI and project director of a National Science Foundation GK-12 grant, Parry developed a highly effective tiered mentoring model for graduate and undergraduate engineering and education teams as well as a popular Family STEM event offering for both elementary and middle school communities.

Parry has been a co-Pi on two NSF DR-K12 Projects: the Exploring the Efficacy of Elementary Engineering Project led by the Museum of Science Boston studying the efficacy of two elementary curricular programs and Engineering For All, a middle school project led by Hofstra University.

In 2014, Liz was appointed by the ASEE Board of Directors to found and chair the inaugural Board Committee on P12 Engineering Education, now the P12 Engineering Commission. She is a Past Chair of the ASEE K-12 and Precollege Division; served as the Vice President of the executive board of the Triangle Coalition for STEM Education, has been a board member of the STEM Consortium and is a member of the K-12 Advisory Committee for the American Society of Mechanical Engineering. She has authored or co-authored over 80 papers, articles and book chapters on issues relating to P20 integrated STEM, including "Perspectives on Failure in the Classroom by Elementary Teachers New to Teaching Engineering," (co-author with Dr. Pamela Lottero-Perdue of Towson University) which was awarded best Division (K-12 and Precollege), Best PIC (IV) and Best Overall Conference paper for ASEE in 2014. Liz is a frequent invited keynote speaker both nationally and internationally. Prior to joining NCSU, Liz worked in engineering and management positions at IBM Corporation for ten years. A longtime mentor, in 2015 Liz received the Presidential Award for Excellence in Science, Mathematics and Engineering Mentoring from President Barack Obama.

PADS: The Performance Assessment of Design Skills (Work in Progress)

Introduction

Performance assessments are a form of contextual assessment where students engage in tasks within a context that affords the use of practices of interest to the assessor. There are many advantages to performance assessment, including face validity, the emphasis on skills and the ability to deal with complexity and relevance. Performance assessment tasks should meet several criteria: they should (1) elicit observable performances, (2) use a standard set of tasks, (3) have high fidelity to "real life" performances, (4) measure a variety of levels of performance, (5) and afford improvement with practice. In engineering in the P-12 setting in particular, there is need for assessment focusing on engineering design performance—not just knowledge about it. Familiarity with the process of designing a product is the aspect of engineering that is most frequently called for and researched [1].

The Performance Assessment of Design Skills (PADS) assesses how a youth approaches an engineering design problem. It is designed to be used with youth ages 9-14 either in school or in out-of-school settings. The PADS presents a narrative about an engaging challenge and asks youths, step by step, to describe their ideas and process to solve it.

PADS is designed to assess several engineering epistemic practices, adapted from those described by Cunningham & Kelly [2]. Engineering epistemic practices are disciplinary practices particular to the practice and culture of engineering, and co-constructed by engineers working in context. The particular engineering epistemic practices that the PADS is designed to assess are:

- *Consider a problem in context.* Engineering problems are defined by clients and by real-world situations.
- *Use a systematic problem-solving process* to support addressing a problem.
- *Investigate properties and uses of materials* to decide which to use.
- *Consider constraints and criteria* that require trade-offs.
- *Envision multiple solutions* and consider which would best serve the problem at hand.
- *Apply science and math knowledge* to problem solving.
- Evaluate designs and *make evidence-based decisions* for improvements.

The assessment that became the PADS, the Elementary Engineering Performance Assessment (EEPA), was originally developed for an efficacy study of Engineering is Elementary (EiE), an engineering curriculum for use in elementary schools. The study found evidence that stronger engineering design skills correlated with EEPA scores, but uncovered concerns that the amount of writing required was skewing scores towards those skilled with writing [3].

In 2020 we worked to revise the instrument for use with after-school programs. The revised instrument, renamed the PADS, used a scenario tested by the EEPA that challenged youths to think about how to help a veterinarian, Dr. Fox, move heavy pets up onto a table in the pet hospital. This "Pet Hospital" scenario was implemented online. The new version of the PADS was piloted with a few dozen youths and scored using a rubric developed for the purpose.

In 2022, we undertook further development and improvement of the PADS as part of a research project to examine its validity and reliability when used to measure improvements in youths' capabilities to engage in productive problem solving. We added a second scenario, the "Carry Books" scenario, as well as video with actors playing the part of characters in the scenarios. Both versions of the PADS are hosted by the online survey platform Jotform, with the option to speak responses instead of typing them (see linked Figures 1 & 2). The revised 2022 PADS can take between 10-40 minutes to complete, with younger children (age 8-9) tending to take longer to complete the assessment than older youths (age 13-14).

Figure 1: https://tinyurl.com/PADS-CB-Q4
Figure 2: https://tinyurl.com/PADS-PH-Q1

Methodology

In 2022, we undertook two studies of the PADS: Study 1 and Study 2. For each study, we chose 3 camps to participate during the summer of 2022. In addition, a fourth after-school program participated in Study 1 during the autumn of 2022. For both studies, camps administered both a pre-PADS and post-PADS, and a program facilitator working completed a short online survey about their experiences implementing engineering and administering the PADS.

Camps in Study 1 engaged in two additional procedures: (1) 5 youths completed the "Trapped Key" design challenge with an interviewer to compare against their PADS completion; and (2) a researcher conducted a structured observation of two hours of programming to collect data on the quality and characteristics of camp engineering programming, as variables for statistical modeling. For Study 2, facilitators (1) scored youths' pre- and post-PADS, and (2) provided feedback on the scoring process. We are comparing the scores of facilitators and researchers on the same assessments, and using the feedback to improve the rubrics and the scoring process.

Examination of Validity Evidence
An advisory committee of experts working in engineering education reviewed the "Pet Hospital" scenario to provide feedback on the validity of the questions to elicit engineering content and process knowledge. Based on feedback, we made changes to the drafts to address content validity. To evaluate response process validity, we used a think-aloud protocol, interviewing four youths aged 11-13 with drafts of the two PADS scenarios in the spring of 2022, in order to identify and correct difficulties students encountered when completing the tasks. All problems found were minor and involved wording changes to the PADS questions.

For the PADS to be useful for the evaluation of engineering design programs, it must be able to detect improvement in the seven aspects of design skills in a situation where we know that there has been high quality instruction. To detect improvement, we used the two forms of the PADS in counterbalanced order, as pre-post-tests of summer engineering camps for middle school students. A meaningful and significant difference between pre-post scores would indicate that the PADS has potential to measure program efficacy. In order to ensure that participating camps and after-school programs provided high quality instruction, a separate team conducted structured observations of the Study 1 research sites. For each participating site, a trained observer rated two one-hour samples of sessions using the Dimensions of Success (DoS) instrument [4] that is sensitive to good engineering instruction. This team also surveyed program facilitators to determine the overall instructional plan, and administered the Common Instrument Suite (CIS) as a post-test to measure attitudes towards engineering among youths and facilitators [5]. This data was summarized and shared with researchers to contribute to the analysis.

We are also working to evaluate convergent evidence of validity by comparing results of the PADS with another high-quality measure of performance. For the comparison, we adapted the Trapped Key interview, which was developed for use with first grade children; it includes a kit of materials for solving the Trapped Key challenge as well as an interview protocol [6]. We adapted the materials kit and questions to be more challenging for older youths and to have questions more comparable to the PADS assessment. Interrater reliability of the PADS has yet to be completed and will be judged by comparing the scores of two raters working independently.

Data Collection
During the summer data collection, the pre-PADS was given on the first (or an early) day of a camp session and the post-PADS was collected after at least 10 hours of engineering instruction, or on the last day of camp. Youths who were given one form of the PADS before instruction were given the other form after instruction. In order to ensure that pre- and post- surveys were paired for data analysis, while youths remain anonymous to researchers, they were asked to choose an animal and give it a name. Camp facilitators recorded these names so that youths could be sure to use the same names and animals when completing the post-PADS. Data collected from the six camps is recorded in Table 1.

Table 1: https://tinyurl.com/PADS-Table1

Two hundred and twenty matched pre- and post-PADS have thus far been collected and scored from 110 youths in these programs. Of these youths, almost 40% of self-identified as Black, and another 40% of youths as White, while Latinx and Native American youths each accounted for 5.5% of the sample (Figure 3). Most youths indicated that they would be entering grades 3 through 7 in the fall (Figure 4). Sixty-five participating youths indicated they were girls; 43 self-identified as boys, and 2 preferred not to state their gender. Most youths indicated on the pre-PADS that they had done “A few” engineering projects (38.5%) or “One” (25.2%) engineering project; only 15.4% indicated they did not know what engineering is, and only 11.2% said they had done “Lots” of engineering projects (Figure 5). Sixteen percent of youths chose to speak and record their responses instead of typing. These assessments were transcribed prior to scoring.

Figure 3: https://tinyurl.com/PADS-Fig3
Figure 4: https://tinyurl.com/PADS-Fig4
Figure 5: https://tinyurl.com/PADS-Fig5

Scoring of the PADS
Finally, we had Study 2 facilitators score PADS from their camps, to gather feedback on usability and appropriateness of the PADS for camp use. Each PADS was scored independently by two trained scorers. New scorers were trained in the process using the *PADS Scoring Guide*, an online training we developed for Study 2. The guide provides a comprehensive rubric, with a

detailed question by question focus on "look-fors", keywords, requirements, and examples for each point level. All scorers trained on the guide before scoring actual youth PADS.

Trapped Key Interviews
We conducted 11 Trapped Key interviews online after mailing kits of materials to participants; results are to be compared to PADS outcomes for the same youths. The Trapped Key interviews included a kit for creating a tool to retrieve a set of keys from a 24" tall clear tube 2" in diameter which was sent to participants in advance of the interview. There were many wooden pegs crisscrossing the vinyl tube at various heights, serving as "barriers" for the rescue tool to get around, much like a tree root may cross a hole in the ground. Youths were asked to devise a tool using only the materials provided that would enable them to "rescue" the "dropped keys" without putting their hands/arms in the device to help. Youths participating in the Trapped Key interviews had also completed the PADS assessment in their camp. At the end of each interview, we asked youths to compare the two experiences.

Data Analysis
For qualitative analysis, we have transcribed the Trapped Key interviews. We plan to code them and compare with PADS scores for each interviewee. As a first pass, we have averaged the codes given by each coder, but we plan to calculate inter-rater reliability, and then to discuss and resolve coding differences. Thus far we have used statistical tests to examine each variable of interest, but once all data is cleaned and coded, we plan to use multilevel regression models, with two levels, to examine each variable's contribution to explaining variance.

Results

Qualitative Findings (Thus Far)
Most participants enjoyed the Trapped Key interviews. Youths consistently preferred the interview over the online PADS, explaining that they enjoyed the ability to touch the materials they were using to design their tool and were nearly universally engaged throughout the process. One youth who had several failed attempts to rescue the keys before their final success stated "This (Trapped Key) was more fun because you got to actually do stuff." Another said Trapped Key was "super fun." One boy thought the PADS scenario problem solving was "long" and "okay." Youth asked questions in the one-on-one Trapped Key, such as "what are the keys made of?" and remained optimistic throughout the process. In the PADS, about 10% of participants seemed to get stuck or disengaged at a point in the assessment, which resulted in either no response or unrelated glib answers.

PADS took longer for most kids to finish than we had anticipated. According to facilitators, PADS took approximately 20 minutes for middle schoolers to complete, but longer for kids in elementary school, up to 40 minutes. Sixteen percent of participating youths chose to make use of the option to speak their answers instead of typing them. Spoken answers tended to be much longer – even ten times as many words per answer – as answers from youths who chose to type.

Quantitative Findings (Thus Far)
We have completed a preliminary analysis. When examining mean scores on the PADS, keep in mind that the minimum possible score is 9, and the maximum is 45.

With matched assessments from 110 youths, we find a statistically significant difference between pre- and post-PADS. We find no significant differences between scenarios, genders, or grade levels. However, we find that youths from racial/ethnic groups underrepresented in engineering (Black, Native American, Hispanic) had significantly lower scores on both pre- and post-assessments than White and Asian youths. This may be at least partially accounted for by the very different racial/ethnic distributions across the camps (see Table 3); we plan to determine the contribution of camp quality to scores once these variables are coded and added to our final model. An overview of average scores per subgroup is given in Table 2. Table 3 shows the demographic makeup per camp. Table 4 shows statistics and p-values.

Table 2: https://tinyurl.com/PADS-Table2
Table 3: https://tinyurl.com/PADS-Table3
Table 4: https://tinyurl.com/PADS-Table4

Comparison	*Test Used*	*SE*	*statistic*	*p-value*	*Cohen d*
Pre / Post	Paired samples t-test, 2-tailed	.555	-3.15	.002	.30
Scenarios CB/PH	Ind. samples t-test, 2-tailed unequal variances	.752	0.94	.350	.13
PRE CB/PH		1.040	1.45	.150	.28
POST CB/PH		1.061	0.03	.974	.01
Gender Male/Not	Ind. samples t-test, 2-tailed, unequal variances				
PRE Male/Not		1.111	1.08	.284	.22
POST Male/Not		1.071	0.70	.483	.14
Race: Rep/Not	Ind. samples t-test, 2-tailed, unequal variances				
PRE Rep/Und		1.077	2.56	.012	.51
POST Rep/Und		1.059	3.03	.003	.59
Grade Level:	Ind. samples t-test, 2-tailed, unequal variances				
PRE 2-5/6-9		1.056	0.17	.862	.03
POST 2-5/6-9		1.055	0.03	.974	.01

Discussion

We find that we can detect improvement from pre- to post-PADS, as well as differences between racial/ethnic groups that are well represented in STEM and those that are not – differences that may or may not be accounted for demographic differences and quality differences between camps. This suggests that development of the PADS may result in an instrument that could be useful to researchers, schools, and possibly out-of-school programs.

Our analysis thus far informs several aspects of the PADS. For example, in out of school time (OST) programs, the length of time a participant spends on the assessment and their buy in or engagement in an assessment scenario can be critical to the quality of their responses. This presents both challenges and opportunities for use of the PADS. One challenge is that the PADS is lengthy to complete and not as engaging as a one-on-one, hands-on assessment; since youths in OST programs are often there by choice, it will be important to make assessment instruments easy and quick to complete, engaging, or both. One opportunity is that, thus far, we have collected positive evidence of the validity of this approach. Further study will shed more light on the strengths and weaknesses of the PADS for assessing youth learning of engineering practices.