7-2021

Developing a Measure to Capture Middle School Students' Interpretive Understanding of Engineering Design

Jeremiah Pina
Smith College

Glenn W. Ellis
Smith College, gellis@smith.edu

Al Rudnitsky
Smith College, arudnits@smith.edu

Rebecca Mazur
Collaborative for Educational Services

Beth McGinnis-Cavanaugh
Springfield Technical Community College

See next page for additional authors

Authors

Jeremiah Pina, Glenn W. Ellis, Al Rudnitsky, Rebecca Mazur, Beth McGinnis-Cavanaugh, and Isabel Huff

2021 ASEE ANNUAL CONFERENCE
Virtual Meeting | July 26–29, 2021 | Pacific Daylight Time
Paper ID #32337

Developing a Measure to Capture Middle School Students' Interpretive Understanding of Engineering Design

Dr. Jeremiah Pina, Smith College

Jeremiah Pina is a postdoctoral researcher at Smith College. He received a B.S. in Psychology and a Ph.D. in Educational Psychology from the University of Georgia. His current research focuses on developing alternative assessment methods for use in the pre-college STEM education settings.

Dr. Glenn W. Ellis, Smith College

Glenn Ellis is a Professor of Engineering at Smith College who teaches courses in engineering science and methods for teaching science and engineering. He received a B.S. in Civil Engineering from Lehigh University and an M.A. and Ph.D. in Civil Engineering and Operations Research from Princeton University. The winner of numerous teaching and research awards, Dr. Ellis received the 2007 U.S. Professor of the Year Award for Baccalaureate Colleges from the Carnegie Foundation for the Advancement of Teaching and the Council for Advancement and Support of Education. His research focuses on creating K-16 learning environments that support the growth of learners' imaginations and their capacity for engaging in collaborative knowledge work.

Mr. Al Rudnitsky, Smith College

Al Rudnitsky teaches Introduction to the Learning Sciences; Thinking, Knowing and the Design of Learning Environments, How Do We Know What Students are Learning?, and instructional methods in elementary and middle school mathematics and science. He has authored books on curriculum design and teaching children about scientific inquiry. Current research interests focus on creating environments for "good talk" in elementary and middle school classrooms, and also on advancing the use of knowledge building pedagogy in higher education. His most recent article (2013) is entitled "Tasks and Talk: The Relationship Between Teachers' Goals and Student Discourse," in Social Studies Research and Practice.

Dr. Rebecca Mazur, Collaborative for Educational Services
Prof. Beth McGinnis-Cavanaugh, Springfield Technical Community College

Beth McGinnis-Cavanaugh is a professor at Springfield Technical Community College, where she teaches courses in physics, engineering mechanics, and structures. A graduate of the engineering transfer program at STCC, McGinnis-Cavanaugh holds a B.S. and M.S. in Civil Engineering from the University of Massachusetts Amherst. She focuses on developing meaningful educational strategies to recruit and retain a diverse student body in engineering and designs innovative learning environments at all levels of the engineering pipeline. Her work in these areas is particularly focused on full inclusion and equity for community college women in engineering and related STEM fields.

Ms. Isabel Huff, Springfield Technical Community College

Isabel is a curriculum designer and training specialist for Transforming Engineering Education for Middle Schools (TEEMS). She joined the "TEEMS team" in 2010 as a Smith College undergraduate. After spending eight years developing her love of storytelling in education and passion for encouraging women in STEM, Isabel took a one-year leave to get a master's degree in education and technology at Stanford. Since returning, her work has focused more heavily on design and teacher professional development.

Developing a measure to capture middle school students' interpretive understanding of engineering design

Abstract

This research paper describes the development of an assessment instrument for use with middle school students that provides insight into students' interpretive understanding by looking at early indicators of developing expertise in students' responses to solution generation, reflection, and concept demonstration tasks.

We begin by detailing a synthetic assessment model that served as the theoretical basis for assessing specific thinking skills. We then describe our process of developing test items by working with a Teacher Design Team (TDT) of instructors in our partner school system to set guidelines that would better orient the assessment in that context and working within the framework of standards and disciplinary core ideas enumerated in the Next Generation Science Standards (NGSS). We next specify our process of refining the assessment from 17 items across three separate item pools to a final total of three open-response items. We then provide evidence for the validity and reliability of the assessment instrument from the standards of (1) content, (2) meaningfulness, (3) generalizability, and (4) instructional sensitivity.

As part of the discussion from the standards of generalizability and instructional sensitivity, we detail a study carried out in our partner school system in the fall of 2019. The instrument was administered to students in treatment (n= 201) and non-treatment (n = 246) groups, wherein the former participated in a two-to-three-week, NGSS-aligned experimental instructional unit introducing the principles of engineering design that focused on engaging students using the Imaginative Education teaching approach. The latter group were taught using the district's existing engineering design curriculum.

Results from statistical analysis of student responses showed that the interrater reliability of the scoring procedures were good-to-excellent, with intra-class correlation coefficients ranging between .72 and .95. To gauge the instructional sensitivity of the assessment instrument, a series of non-parametric comparative analyses (independent two-group Mann-Whitney tests) were carried out. These found statistically significant differences between treatment and non-treatment student responses related to the outcomes of fluency and elaboration, but not reflection.

Introduction

One of the most timely and pressing goals of promoting early STEM education is to create educational experiences that will both broaden enduring participation in the study of STEM topics [1] and that will establish a long-term learning framework to encourage students to train for important STEM careers [2]. In the field of engineering education, this idea is bound up in metaphors like those of "the pipeline", "the ecosystem", and "the pathways" [3,4], all of which include a multitude of diverse trajectories in the course of study and skill development whereby students can access their opportunities to become expert engineers [5].

There are many ways in which pre-college engineering education can contribute to a larger program of shaping long-term pathways of engineering expertise. Ideally, pre-college engineering education courses can provide students with an introduction to the logic of the engineering design process [6], they can serve as an interdisciplinary venue by which to connect engineering concepts to other STEM topics [7], and they can help foster engineering habits of mind [8].

Many educators and policy makers across the country have made strides towards establishing engineering as a core academic subject and codifying curricular standards such as the engineering standards embedded in the Next Generation Science Standards [9]. These efforts have resulted in widespread agreement about what pre-college engineering students should know and understand about engineering. Still lacking, however, are high-quality methods of assessing learner progress toward mastery of engineering concepts, which is a necessary element of meeting many of our national engineering education goals [10].

Complicating this issue from the assessment standpoint is the fact that measuring understanding and the commensurate development of expertise is challenging and continues to be an area of active research [11]. That said, our best understanding of the development of engineering expertise describes a multifaceted process whereby variant courses of development can result in substantially different outcomes [12]. As such, the more concrete question from the position of practical assessment cannot be reduced solely to the question of *how much development*, but rather the more challenging question of *how much of what kind of development* could be occurring.

Impetus for developing the assessment instrument

This paper describes the process by which we developed an assessment instrument to measure students' *interpretive understanding* of engineering design concepts as nascent indication of developing expertise in the middle school engineering education context. We had multiple interrelated goals for this project. In part, we realized the need to create a new alternative assessment as one of several means by which we might evaluate student learning outcomes related to a narratively-based engineering curriculum aligned to the Next Generation Science Standards (NGSS) [13] that was being co-developed by the project team (this is detailed in sections to follow).

We intended to create assessment items that could elicit a depth of conceptual understanding characterized as the products of *meaningful learning* [14] of a kind that we would expect to result from a curriculum that was aligned to the NGSS and informed by the concepts central to that understanding of engineering design. Taking inspiration from prior assessments like the one described by Atman, Kilgore, and McKenna [15] and building on models of expertise development that are highly specific to the field of engineering education, the assessment instrument is intended to highlight indications of thinking captured from student responses as they are tasked to *generate, reflect*, and *demonstrate* their understanding of the engineering design process.

At the same time, we sought to engage with the theoretical bases of an assessment of interpretive understanding with the kind of research that would bridge basic and applied modes, constituting use-inspired basic research as it might occur in Pasteur's quadrant [16]. Among the theories of the development of interpretive understanding and expertise development we encountered, no previous effort had undertaken to adapt these models into assessment instruments that could be used in the pre-college engineering education context. As such, the synthetic assessment model and the assessment instrument that situates indications of expertise in terms of the NGSS concepts and core ideas (described below) represent more than an effort to collect evidence that can help to determine the efficacy of a curriculum. They also, more generally, represent an attempt to formulate an assessment method that can potentially be used to explore implications of this larger theoretical framework in the specified context.

Theoretical Basis

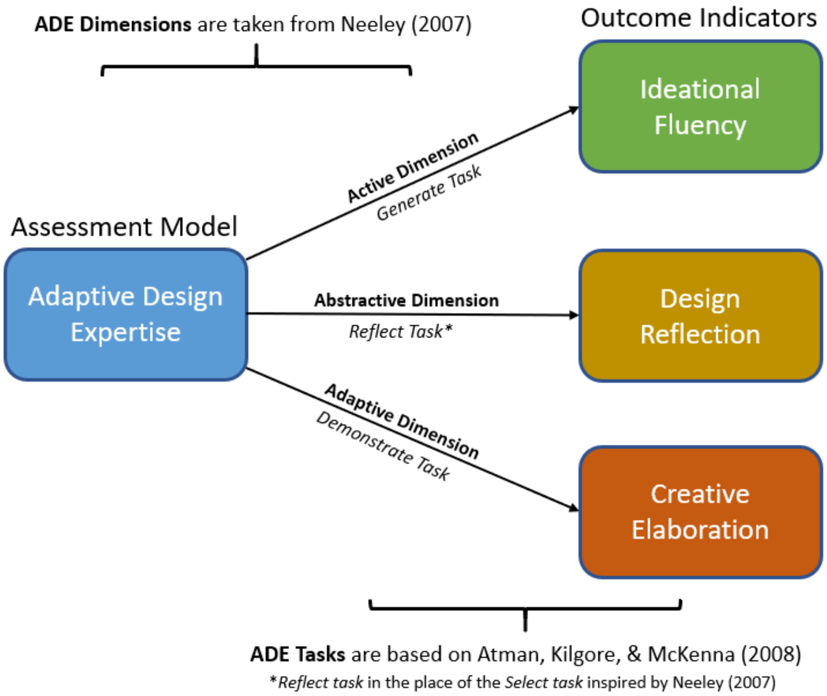
Our definition of *interpretive understanding* was informed by the theories of preparation for future learning (PFL) [17] and Broudy's [18] knowledge typology. In this shared framework, the process of higher-order thinking development is conceptualized as students' increasing ability to make use of their understanding in ways that both incorporate and go beyond replicative and applicative uses of knowledge such that learners can solve new problems in new situations, interpretively "knowing with" the concepts they have internalized [17]. Therefore, assessing interpretive knowledge requires capturing indications of student understanding that take them out of the context of direct application and instead attends to the ability of students to use their current knowledge to facilitate new understanding [19].

This framework owes much to theories of expertise development because PFL frames the overall process of knowledge-rich learning as being fundamentally contributory [19, 20] to the development of adaptive expertise (AE) [21] and also because PFL depends on AE to describe underlying dimensions of student learning, as with efficiency and innovation [19]. On its own, AE is a theory that problematizes variations in courses in the development of expertise, most especially attending to those outcomes that lead to routinized understanding, rote expertise, and those which can promote flexibility and innovation in creating new solutions. This latter, more significant, type of expertise development is what is most strongly emphasized in AE [21]. AE has been described as a necessary form of expertise development in the field of engineering education for the reason that adaptive expert engineers are more capable of organizing their thinking around big ideas of design [22] and are therefore able to meet a wider range of challenges with better problem solving strategies [23].

What these different models hold in common is the notion that expertise development is a multidimensional construct and that the process of expertise development can be separated into the thoughtful cultivation of different types of thinking. Broadly, these theories can be considered variations of meaningful learning models [14], as they disaggregate differences of the kind of thinking that contributes to rote learning–learning that requires only the memorization of definitions and fixed procedures–and the kind of thinking that fuels the more interpretive, transfer-based, meaningful learning.

For our purposes, we sought to simplify these elaborate theoretical relationships by creating a synthetic assessment model (see Figure 1) that mapped the elements of interpretive understanding on to well-defined, testable, indicators of student learning. To do so, we looked both to the dimensions of *active*, *abstractive*, and *adaptive* thinking described in the theory of Adaptive Design Expertise (ADE) [24] and to the to the methods and assessment tasks employed by Atman, Kilgore, and McKenna [15]. As with the unmodified ADE model, we conceptualized low-level understanding of the process of design to be accounted for in the development of active thinking alone. However, deeper, more interpretive, understanding of design concepts can only occur if the higher dimensions of ADE are also engaged [24].

Figure 1: Synthetic ADE assessment model



We articulated student learning outcomes in terms thinking behaviors belonging to each of the three ADE dimensions (see Table 1). Details of the associated thinking behaviors, identified as outcome indicators of ADE development, were adapted from Atman, Kilgore, and McKenna [15] with a few alterations: The indicator behavior of *fluency* was recast from its initial, primarily linguistic, role, and instead was more broadly interpreted through the lens of *ideational fluency* [25]. The engineering design activities [26-28] that served as indicators of the *reflect* task were retained, though they were positioned not as ranked comparisons, but instead as features of reflection [29] that could be represented in student responses.

Table 1: Definitions of ADE thinking behaviors

ADE Dimension	Thinking Behavior	Definition
Active Thinking	Ideational Fluency	Using terms and concepts associated with the engineering design process with ease and fluidity.
Abstractive Thinking	Design Reflection	Reflecting on specific features, processes, and experiences of engineering design activities.
Adaptive Thinking	Creative Elaboration	Producing vivid and sophisticated details of relevant engineering design process understanding.

Finally, again expanding on the ideas present in Atman, Kilgore, and McKenna [15] and using ideas considered in an earlier assessment of this type [28], we delimited the assessment task of "demonstrating ADE" as assessing those thinking behaviors that would contribute most to the process of modulating both novel and appropriate [24]–or innovative and efficient–uses of engineering design. To inform this aspect of our assessment model, we looked to the literature on innovation and creativity development in engineering education [30,31] and incorporated the idea of assessing creative elaboration of student responses [32, 33].

Study Goal

The goal of the study was to develop an assessment instrument based on the synthetic ADE assessment model for use in the pre-college engineering education setting. The following discussion will provide an account of the iterative process by which the component tasks of the instrument were refined into their current form. We follow with a brief discussion of the validity and reliability of the instrument. As part of that discussion, we then provide the results of a comparative analysis of student responses that demonstrates the utility of the assessment instrument in distinguishing early indications of the ADE development in variant instructional contexts.

Development Process

We developed our assessment instrument over the course of five iterations–three for item prompt development and two for the construction and clarification of the scoring protocols. Our process was informed throughout by two factors: Our intention to align our assessment instrument with the fundamental concepts and disciplinary core ideas represented Next Generation Science Standards (NGSS) content standards for sixth, seventh, and eight grade engineering education [13] and our desire to incorporate the perspectives and insights of middle school instructors in our partner school system–an urban school district in the Northeast U.S. that primarily serves students of color.

This latter effort was supported through a collaboration with our Teacher Design Team (TDT). This group was composed of between three and six (depending on availability) 7th and 8th grade instructors who had been teaching in the partner school system for at least two years. No TDT member was actively teaching 6th grade during the development period, as the project team was planning to implement an experimental engineering curriculum for classes at that grade level.

The role of the TDT in the early phases of development was to provide guidance and feedback to optimize our assessment materials to better fit the contextual needs of classrooms in our partner school system. Initial meetings focused on challenges in student understanding of engineering concepts, students' prior access to engineering as a field of study, familiarity with STEM topics, and effective strategies for improving engagement with the assessment instrument. Based on considerations surfaced in our meetings with the TDT, a set of guidelines for assessment item development and implementation were generated (see Table 2) that were incorporated as measures to facilitate student access to the assessment instrument and mitigate perceived challenges in the method employed.

Table 2: Guidelines and implementation plans based on TDT feedback

Guideline	Description	Implementation Plan
Maximize student access to assessment materials	Assessment implementation protocols would promote multiple methods of student participation	Assessment would be offered in a variety of media, including as a pen-and-paper test and as digital items
Maximize the relatability of assessment tasks	Assessment would focus on real-world topics, using well-known situations	Real-world situations and concerns would be actively invoked in assessment prompts when possible
Minimize the complexity of requests made of students	Assessment questions would only make simple requests of students	Assessment would undergo multiple revisions to reduce task complexity
Minimize the use of technical jargon	Assessment prompts would only include STEM language that was taught in recent prior lessons	Assessment items would be focused on utilizing simple vocabulary and a highly controlled use of jargon
Minimize the length of the assessment	Assessment instrument should not prove a time burden on teachers or students	Total length of assessment tasks would be monitored

Building out from the guidelines co-developed in our meetings and communications with the TDT, our first version of the assessment instrument was a broad-spectrum attempt to measure of the concepts embedded in the NGSS content standards and disciplinary core ideas [13] that pertained to the middle school-level study of engineering design concepts.

The NGSS has enumerated four middle school engineering design content standards. In brief, these can be summarized as: MS-ETS1-1: Defining the criteria and constraints of a design problem; MS-ETS1-2: Evaluating design solutions using a systematic process; MS-ETS1-3: Analyzing data to compare design solutions; and MS-ETS1-4: Developing a model for iterative testing of design solutions. These content standards are elaborated through the application of three NGSS disciplinary core ideas (ETS1.A, ETS1,B, and ETS1.C) that relate to the three-step NGSS engineering design process. More information about these topics can be found on their standards summary page [34].

The working draft of the assessment instrument contained at total of 17 items, some of which were supplementary assessment measures and alternate, short form, versions of the ADE items. These consisted of ten selected-response items focused on concepts represented in NGSS standards MS-ETS1 and MS-ETS1-2. We also designed four simple problem-solving items aimed at capturing indications of students’ ability to make use of the engineering design process, touching to elements in both NGSS standards MS-ETS1-3 and MS-ETS1-4, and cross-cutting concepts in the disciplinary core ideas. And finally, we drafted three long form narratively based design scenario problem-solving tasks that were focused more tightly on directly measuring students' understanding of the NGSS disciplinary core ideas (ETS1.A, ETS1.B, and ETS1.C) of engineering design.

Table 3: ADE task, prompt condition, and NGSS standard of the final assessment items

ADE Task	Prompt Condition	NGSS Standard	Assessment Item
Generate	Initial Problem-Solving of a Design Scenario	MS-ETS1-1; ETS1.A	Some of your neighbors (people who live near you) have worked hard to raise money and buy some empty land on Springfield Avenue. Now they want to make it into something better. How can *you* help? You're a middle-school student who knows about engineering design–so you can give your neighbors ideas about how to solve their problem. [*An image of an empty lot is presented*]. Thinking like an engineer, what would you say to your neighbors about their first step in figuring out how to use the land?
Reflect	Retrospective Reflection	ETS1.A	Think about what you have learned about the engineering design process. How has what you learned changed how you think about solving problems?
Demonstrate	Figural Representation of Design Concepts	ETS1.A; ETS1.B; ETS1.C	Using the space on this page, create a picture or diagram that shows your ideas about what the engineering design process is. Include as many details as you can remember.

Ultimately, feedback from the TDT reduced the number of assessment items to only three (see Table 3), all of which could best be characterized as alternative assessment items [35,36] that made use of problem-solving scenarios and creative approaches to the engineering design process. TDT teachers expressed a preference for open-ended assessment items that could allow students to provide authentic answers and an aversion to traditional, selected-response items.

The process of creating scoring protocols drew from the scoring methods presented in multiple prior works contributory to the synthetic ADE assessment model. They were, however, adapted and respecified for use with middle school students as follows:

To score the *generate* task, a simplified rubric was developed to score *uses of engineering design*. These were terms, processes, and ideas that would be novel to a middle school student being introduced to engineering design. This modification aside, the scoring for that task was like the one presented in Atman, Kilgore, & McKenna [15].

The scoring protocol for the *reflect* task was more complex and drew from the methods for capturing the *features of reflection* described in Turns et al. [29]. Over two iterations, a final set of six codes were generated, considerably simplifying the 23 engineering activities described in Hill [26]. These six codes were features of reflection relating to collaborating and community (CO), making a model (MO), assessing and measuring (AS), planning out a design (PL), using a step-wise process (PR), and building a context outside of the problem (BC).

For the *demonstrate* task, a rubric was designed to score the figural elaboration of student representations of the engineering design process. This rubric was a synthesis of the "sketch data sort" method employed Mosborg et al. [28] and the "technical strength" scoring criteria described in Denson, Buelin, Lammi, and D'Amico's [32] engineering design assessment instrument. Using a graduated rubric, student responses scored in two stages. They were first grouped based on the overall features of their drawn response and then sorted based on the elaboration of details in the drawing.

Samples of both the 3-item final assessment instrument and the scoring procedure materials are available in the Appendix I and Appendix II of this paper.

Validity and Reliability

An essential consideration in the development of any new assessment instrument is the degree to which it can be demonstrated to be valid and reliable. Standards of validity and reliability often look to gathering multiple sources of evidence in defense of the quality of an assessment and the interpretability of its scores [37-39]. These sources of evidence tend to be influenced by the context and purpose of the assessment instrument [37] and, as such, arguments for validity, rather than being monolithic and timeless, tend to be ongoing and contingent [38].

For our purposes, and primarily due to the nature of our assessment items (these being open-ended alternative assessment tasks), our validity framework was most strongly informed by a set of validity criteria described in the work of Lane [40], which we viewed as being complementary to the widely adopted *Standards for Educational and Psychological Testing* [39]. In line with these standards, we gathered evidence to support four arguments–those from the standards of (1) content representativeness, (2) meaningfulness, (3) generalizability, and (4) instructional sensitivity.

Evidence for the *content representativeness* of the assessment instrument–addressing "the coherency and representativeness among the assessment tasks, scoring rubrics and procedures, and the target domain" [40]–has already been provided in the Development Process section of this paper, most notably in the discussion of alignment with NGSS content standards and disciplinary core ideas as referenced in Table 3. Likewise, evidence for the *meaningfulness* of the instrument–encompassing the necessity to "measure more directly the types of reasoning and problem-solving skills valued by educators" [40]–is also detailed in the Development Process section of this paper, conveyed in our efforts to work closely with our partner school system to build an assessment to match the circumstances and ideals of our partner instructors. The guidelines and implementation plans co-developed with the TDT (Table 2) and the selection of final assessment items (Table 3), informed as they were by direct TDT feedback, are evidence for strict adherence to this standard.

To gather evidence for validity claims to the latter two standards, a study was carried out in the fall of 2019 in our partner school system. Our methods for composing this evidence were primarily quantitative, drawing on interrater reliability data to address the standard of *generalizability*–a metric that reveals the consistency or reliability of the scoring protocols [41]–and comparative group analysis as method of developing evidence for the standard of *instructional sensitivity*.

Participants

Our partner school system for this study was an urban public P-12 school district in a Northeastern state. The district serves more than 25,000 students across grade levels, and supports 32 elementary schools, 12 middle schools, 3 secondary schools (grades 6 to 12), and 8 alternative schools. A majority of the district's students are Hispanic (67%) or African American (19%). A sizeable majority (83%) of the district's students are designated as "high needs" students. This is a designation that describes factors related to language needs, economic disadvantages, and/or disability status.

Six of the district's 12 middle schools agreed to participate in the study: Four as treatment schools and two as comparison (non-treatment) schools. Teachers in treatment schools were tasked with implementing an experimental engineering design module based on the Imaginative Education [42] teaching approach and aligned with NGSS standards and disciplinary core ideas. Teachers in comparison schools implemented their regular curriculum, a variety of NGSS-informed instructional practices highlighting the engineering design process, not too dissimilar from those described in Chandler, Fontenot, and Tate [43]. Each school had a single 6th grade science/engineering teacher. In total, 724 students were assigned to these classrooms, 410 in treatment classrooms and 314 in comparison classrooms.

Data Collection

The assessment instrument was administered to 6th grade students in our partner school system in October and November of 2019, following a two- to three-week instructional segment which introduced students to concepts, procedures, and terminology related to the engineering design process. Pen-and-paper versions of the assessment instrument were supplied to teachers in both treatment and non-treatment schools as three-page assessment packets. Digital versions of the assessment packets were created and uploaded to a centralized database for ease of access, but all the participating teachers opted for the pen-and-paper versions.

Prior to implementation, instructions for administering the assessment were provided by our project's training specialist. Teachers were provided in advance with the assessment items and scoring procedures and worked with our training specialist to field questions concerning implementation of the assessment instrument. One significant concern surfaced at the time of implementation by multiple participant instructors was the question of language of administration, as some participating schools had students with low English proficiency. In these cases, teachers were encouraged to provide translations of the assessment items and we made it clear that we would accept responses in languages other than English.

In all, 447 assessment packets were returned, 246 from comparison schools and 201 from treatment schools–a response rate of 61%. The rate of response was influenced by a combination of factors, including that some students chose not to complete the assessment (it was not counted toward student grades) and because some students were absent on the day selected by their teacher to administer the assessment. Only one assessment packet was received with answers written in a language other than English. This packet was translated by our project team during the scoring process.

Generalizability

Analysis of students' responses started in December of 2019 and continued through March of 2020. One of our early goals during this process was to utilize the scoring procedures and student responses to determine interrater reliability [44] of the rubrics and codes. We did this to ensure that scores were generalizable across graders and, thereby, as a means to "examine the extent to which scores derived from an assessment can be generalized to the domain of interest" [40].

Scorers were recruited from the undergraduate student body of the project's host university. The groups of scorers consisted of at least two undergraduate students and a member of the research team, though group composition varied somewhat over the course of analysis. Our procedure for analyzing and improving the interrater reliability of the assessment instrument was as follows: scorers were (1) provided with a one-hour training session on the purpose of the assessment and the scoring methods; they were then (2) provided with the scoring materials and, in a group, were guided through the scoring process using example student responses; afterwards, they were (3) individually and separately provided with a randomized sample of 60 student responses to score, their scores were (4) initially analyzed and an intra-class correlation coefficient (ICC), a measure of the agreement among scorers [45], for each outcome was calculated. In the case of low initial reliability–especially pertinent to the reflect codes–scorers were (5) reconvened to discuss disagreements in responses and discontinuities in the scoring rubrics that could be addressed; finally (6), using revised rubrics and discussion notes, a second random set of 60 responses were scored and analyzed to calculate an ICC.

Table 4: Intra-class correlation coefficient (ICC) of the scoring method for each assessment task

Assessment Task	Scoring Method	ICC
Generate	Rubric	0.89
Reflect (CO)	Code	0.95
Reflect (MO)	Code	0.84
Reflect (AS)	Code	0.75
Reflect (PL)	Code	0.72
Reflect (PR)	Code	0.82
Reflect (BC)	Code	0.86
Demonstrate	Rubric	0.94

Final ICCs (see Table 4) were computed using R software [46]. All eight ICCs were modelled as a two-way random-effects model with absolute agreement [47]. All final ICC models were found to be statistically significant at or below the p = .01 level.

As Koo and Li [47] have outlined, interpretations of ICCs can be made for scores ranging between .50 and .90 as incrementally improving from "poor" to "excellent." The eight reported ICCs of our assessment instrument performed at or above the "good" level, with the Reflect (CO) code and the Demonstrate rubric performing at the "excellent" level, and the Reflect (PL) code slightly underperforming, just under the criteria for being considered "good", at the “moderate” level. Overall, these interrater reliability metrics provide strong evidence that the scoring methods are reliable across trained scorers and, therefore, generalizable in the sense that the scores can be produced and used consistently.

Instructional Sensitivity

Another goal of our study was to generate evidence of the instructional sensitivity of our assessment instrument. Instructional sensitivity measures the degree to which assessment tasks are sensitive to changes in instruction [48]. One expression of instructional sensitivity, as described by Lane, is “the extent to which differences in instruction affect performance on the assessment” [40].

Our method of developing evidence in support of the instructional sensitivity of our assessment instrument was to draw statistical comparisons between the treatment and non-treatment groups in our study. The logic of this analysis depends on the position that our experimental curriculum is sufficiently different from the standard teaching practices of middle school engineering that instructors in the treatment group would meaningfully differ from non-treatment teachers in their approach. Three features of our experimental curriculum–as described in Ellis, Piña, Mazur, Rudnitsky, McGinnis-Cavanaugh, Huff, Ellis, Ford, Lytton, and Cormier [49]–supported this presupposition: the transmedia nature of the lessons, the overarching narrative structure of the curriculum, and the use of Imaginative Education thinking tools.

Singly, a transmedia, narrative, or Imaginative Education-driven curriculum would be noteworthy in pre-college engineering education. Transmedia engineering education is a novel approach that relies on technology to build learning experiences that weave between various digital media (images, movie clips, written documents) and classroom experiences in a “nonlinear, participatory” manner [50]. Narrative-based engineering education uses storytelling as a context to link together elements of engineering design thinking so that learners can make better sense of their process [51]. The Imaginative Education teaching approach employs specific thinking tools–extremes of reality, heroism, metaphor, and others–to encourage students and teachers “rethink how they engage with content” [52].

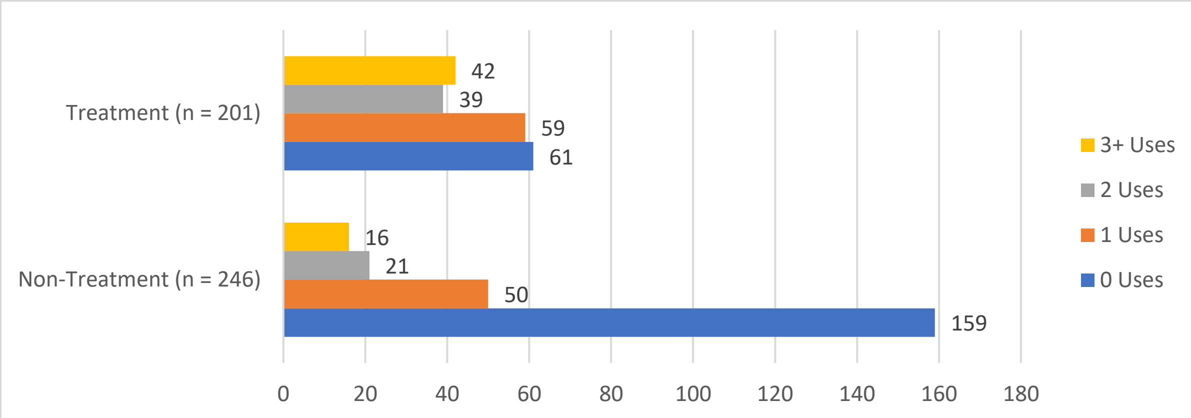
As our experimental curriculum represented a union of these three factors–each a divergence from traditional classroom practice in our partner school system–we were confident that instruction in the treatment classrooms differed enough to aid our attempts to calibrate the instructional sensitivity of our assessment instrument.

However, it is worth highlighting that, due to the manner in which we positioned differences in the curriculum as the origin of variations in instruction, our analysis of student responses was focused on broad group differences (treatment vs. non-treatment) rather than those fine-grained conditions that might occur at the level of individual teachers. This simplification represents a limitation of these analyses that we address in the following section.

Results from our analysis of student responses are organized by task below:

Generate: Scoring of the generate task focused on the fluency of engineering design ideas utilized in responses to a problem-solving scenario. The pattern of student responses to the generate task (see Figure 2) showed marked differences in the ideational fluency of responses. Students in the treatment group were better oriented to the task and were able to produce engineering design-based ideas in their responses at a higher rate than students in the non-treatment condition. Especially noteworthy was the finding that a majority of students in the non-treatment group (65%) were not able to produce a single indication when responding to this task.

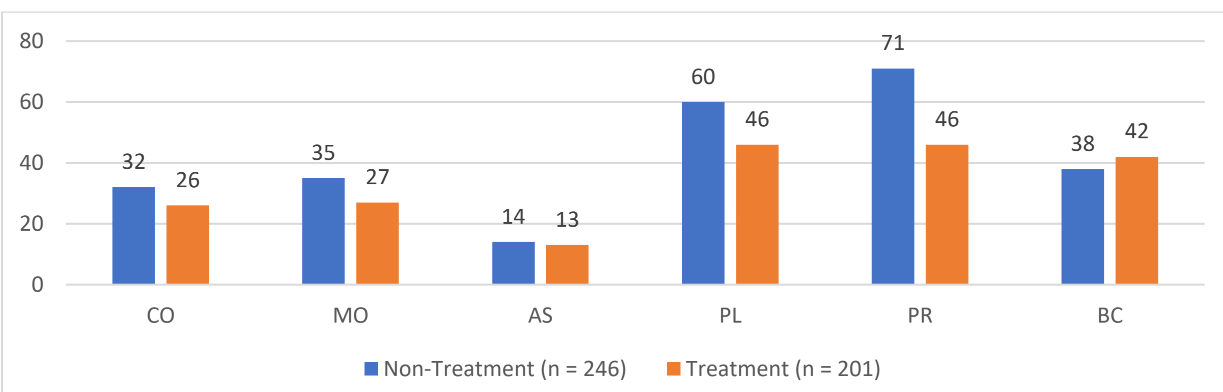
Figure 2: Fluency levels of student responses to the generate task (treatment vs. non-treatment)



A preliminary Shapiro-Wilk's test of normality determined that the fluency data was not normally distributed ($p < .05$). As such, non-parametric comparative analysis of the unranked by-group data (a Mann-Whitney U test) was carried out. This analysis showed a statistically significant difference in the responses of the 201 students in the treatment group (Mdn = 1) and the 246 responses from students in the comparison group (Mdn = 0), $U = 15024$, $p < .001$. This finding supports the inference that the task is sensitive enough to detect differences in the fluency of student responses resulting from group differences in instruction of the engineering design unit.

Reflect: Responses to the reflect task were coded for each of the six representations of engineering design activity that had been derived from prior work [26]. In essence, all six codes were treated as independent outcome variables through which by-group comparative analyses could be undertaken. Furthermore, we had hoped that, when viewed across reflection outcomes, the pattern of scores could demonstrate a signature of responses revealing students' inclinations toward attending to certain features of reflection [29].

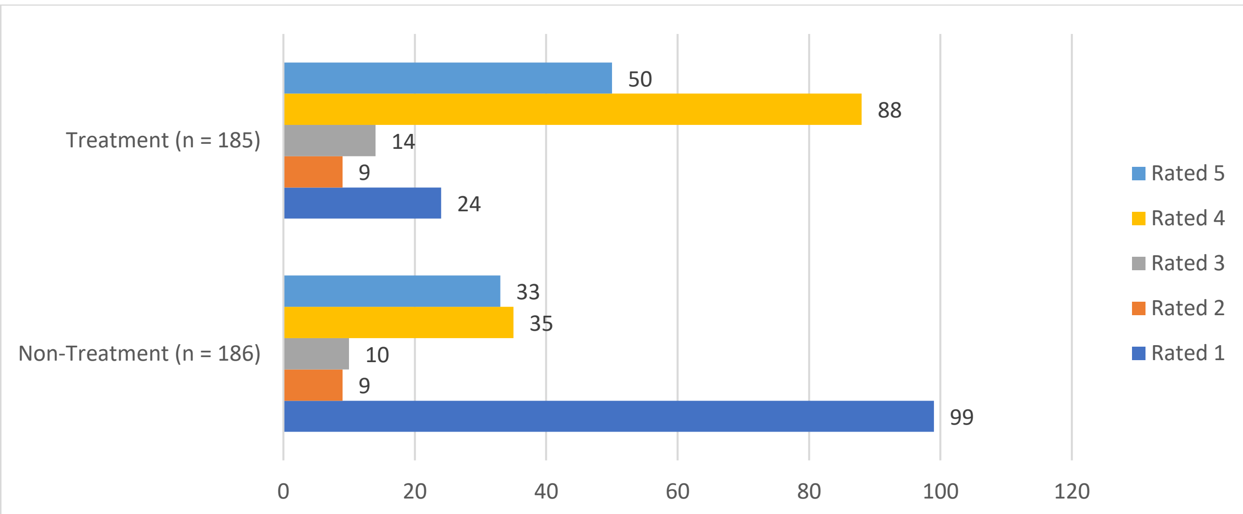
Figure 3: Number of students with one or more responses identified for each reflect code (treatment vs non-treatment)



Progress toward this goal can be seen in counts of the student responses for each of the codes (see Figure 3) which demonstrate an overall pattern of student responses favoring the insight that students were approaching the reflection task using many of the same features in both treatment and non-treatment classes, with some differences in how students reflected on planning their solutions (the PL code) and the appeals they made in their responses to a specific stepwise process (the PR code). However, comparative statistical analysis did not find differences in the performance of students in terms of any of the six outcomes. This constituted a limitation in our understanding of the instructional sensitivity of this task which will we discuss in the following section.

Demonstrate: Student responses to the demonstrate task were scored using graduated rubric that ranked the characteristics of the response first by type and then by elaboration of details. Analysis of those scores compared the responses of treatment and non-treatment groups based on these rankings (see Figure 4). Perhaps due to the nature of the task, a number of packets from both treatment and non-treatment classrooms were returned blank–60 for the non-treatment and 16 for the treatment group. These were excluded from our later analysis.

Figure 4: Count of ranked responses to the demonstrate task (treatment vs. non-treatment)



Group differences in the pattern of student responses to the demonstrate task were clear. Students were better oriented to the task and a majority of students in this group (75%) were able to provide responses that met the criteria for being considered a visual depiction of a stepwise engineering design process (responses Rated 4 or Rated 5). In contrast, a majority of students (53%) in the non-treatment group produced visual representations that were categorized as simple illustrations (Rated 1), and that often did not relate to engineering design.

As with the earlier fluency analysis, a Shapiro-Wilk's test of normality indicated that the distribution of student response to the demonstrate task was non-normal ($p < .05$). For this reason, as earlier, a Mann-Whitney U test was carried out to investigate the statistical significance of group differences in the responses. The analysis demonstrated a statistically significant difference in the responses of the 185 students in the treatment group (Mdn = 4) and the 186 responses from students in the comparison group (Mdn = 1), $U = 10310$, $p < .001$. This finding supports the inference that the task is sensitive enough to detect differences in the creative (figural) elaboration of student responses resulting from group differences in instruction of the engineering design unit.

Limitations and Future Work

While our results provided good evidence of the validity and reliability of our assessment instrument, there were limitations to the study. One of these was related to the deeply contextual nature of the study and the development process of the assessment instrument itself. Working with our partner school system at multiple stages of development was a boon in that the assessment was well-matched to the concerns of teachers and students; however, certain of these concerns are not shared across public schools in every context and so while we chose to create an assessment that would be more meaningful, we did so by balancing against the potential generalizability of the assessment.

Another limitation of the study was in gauging the instructional sensitivity of the assessment tasks. Ideally, the unit of analysis of instructional sensitivity would be the instruction provided by each individual teacher, such that we could "ensure alignment and coherency among curriculum, instruction, and assessment" [40]. However, we could not conceive of a method of validation that would uphold this standard at the level of individual instruction without incorporating a separate study of the *fidelity of implementation* [53] of the curriculum. Without this component of analysis, there would be no direct indication of the degree to which individual teachers were efficacious and effective in implementing the curriculum [54]. Our current analyses sidestepped the issue by simplifying these topics, though still following the logic of analysis described in prior studies of this type–see, for comparison, Neimi, Wang, Steinberg, Baker, and Wang [55]. That said, our intention is to incorporate a measure of fidelity of implementation as part of our future work in investigating the instructional sensitivity of our assessment instrument.

The instructional sensitivity of the reflect codes was yet another limitation of the study. Our analyses did not detect differences for any of the six codes we had created as outcome measures. One possible reason for this is that our specification of the codes was not robust enough to adequately capture the features of reflection that would be salient for students in a pre-college engineering education setting. The literature of reflection in engineering education is sparse [56], and this is certainly the case for uses of reflection as assessment in pre-college engineering education. Our future work in this area will center on an attempt to wrangle the theoretical categories of the details that students can reproduce in their reflections as they attend to engineering design and the meanings they attach to those understandings. This will most likely entail generating an assessment model including a more expansive set of response indications by which to describe the process of reflection.

The final limitation of this study was in the types of validity and reliability evidence collected and examined. Cultivating the validity of an assessment instrument is an on-going process that does not end when one or more types of validity evidence are presented [39]. As such, we recognize that there are other validation strategies to pursue. Lane [40] has outlined several standards that we did not explore in this article; these range from cognitive complexity, to fairness, and, further, to the long-term consequences of the use of the assessment.

In the future, we expect to approach a broader selection of these standards as means of continuing to develop an assessment instrument that is thoughtful, useful, and well-composed. Our further work will also more directly investigate the influence of group differences—most especially those of gender, culture, and socioeconomic background—in expressions of ADE thinking behaviors of middle school engineering students. This topic is an essential part of a related ongoing research project that will be explored in future publications.

Conclusions

We developed our assessment instrument to measure the ideational fluency, design reflection, and creative elaboration of middle school engineering students as they engaged with the *generate*, *reflect*, and *demonstrate* tasks.

Our assessment instrument was strongly informed by the NGSS content standards and disciplinary core ideas for the middle school level of study. It was, furthermore, shaped by thoughtful feedback provided by TDT teachers from our partner school system. Our strict adherence to these bases for creating and framing assessment items is strong evidence of validity from the standards of content representativeness and meaningfulness. The intraclass correlations found in our study provided evidence of good-to-excellent reliability for the scoring procedures, which is, in turn, strong evidence for the generalizability of the assessment instrument. Comparative analyses of student responses between treatment and non-treatment groups provided evidence that the assessment was instructionally sensitive for both the fluency and elaboration tasks.

In all, this body of evidence indicates that the assessment instrument can reveal some of the thinking behaviors that we expect to contribute to the development of adaptive design expertise. While we view this as is an important step, it is worth noting that the path from the middle school engineering classroom to the completion of an engineering degree is an exceedingly long one. This assessment reflects an attempt to measure only one of the many kinds of influence that will be necessary if students–especially young women and students from minority backgrounds–are able to fully participate in the long-term process of becoming an engineer [57].

However, by getting a better sense of students' interpretive understanding at the middle school level through high-quality assessment instruments, we can better inform instruction and create more focused and topical opportunities for students to cognitively engage with engineering design concepts. Moreover, as we develop better measures, we can more usefully aid teachers and students in cultivating a deeper and more fluent early understanding of the principles of engineering. The assessment instrument we developed represents, at best, a modest contribution in the overall project of the shaping developmental trajectories of engineering students, but one that is vital if we want continue the project of improving the effectiveness of our pre-college engineering education programs.

Acknowledgement

This material is based upon work supported by the National Science Foundation under Grant Nos. 1814033 and 1813572.

References

[1] President’s Council of Advisors on Science and Technology. “Prepare and inspire: K-12 education in science, technology, engineering, and math (STEM) education for America’s future,” 2010, Available: http://www.whitehouse.gov/administration/eop/ostp/pcast/docsreports

[2] National Research Council. *Rising above the gathering storm: Energizing and employing America for a brighter economic future.* Washington, DC: National Academies Press, 2007.

[3] J. Cruz & N. Kellam, “Beginning an engineer's journey: A narrative examination of how, when, and why students choose the engineering major,” *Journal of Engineering Education*, vol. 107(4), pp. 556-582, 2018.

[4] W. C. Lee, “Pipelines, pathways, and ecosystems: An argument for participation paradigms,” *Journal of Engineering Education*, vol. 108, pp. 8-12, 2019.

[5] C. Atman, S. Sheppard, L. Fleming, R. Miller, K. Smith, R. Stevens . . . D. Lund, “Moving from pipeline thinking to understanding pathways: Findings from the Academic Pathways Study of Engineering Undergraduates.” *Proceedings of the 2008 American Society for Engineering Education Annual Conference*, Pittsburgh, PA, June 2008, DOI: 10.18260/1-2—3786.

[6] T. Martin, P. Ko, T. Benton, C. Farmer, & D. Allen, “Transfer of adaptive expertise to transform engineering education.” *Proceedings of the 2010 American Society for Engineering Education Annual Conference, Louisville, KY and P-12 Engineering and Design Education Research Summit*, 2010.

[7] C. Chase, L. Malkiewich, & A. Kumar, “Contrasting Cases Enhance Transfer of Physics Knowledge from an Engineering Design Task,” *Proceedings of the 2018 CogSci Conference: The Annual Meeting of the Cognitive Science Society*, 2018, ISBN: 978-0-9911967-8-4

[8] M. Lammi, C. Denson, & P. Asunda, “Search and review of the literature on engineering design challenges in secondary school settings,” *Journal of Pre-College Engineering Education Research (J-PEER)*, vol. 8(2), Article 5, 2018.

[9] T. J. Moore, K. M. Tank, A. W. Glancy, & J. A. Kersten, “NGSS and the landscape of engineering in K-12 state science standards,” *Journal of Research in Science Teaching*, vol. 52(3), pp. 296-318, 2015.

[10] S. Purzer, T. J. Moore, D. Baker, & L. Berland, “Supporting the implementation of NGSS through research: engineering,” *Proceedings from the 2014 NARST Annual International Conference*, Reston, VA, 2014.

[11] R. Hutton, P. Ward, J. Gore, P. Turner, R. Hoffman, A. Leggatt & G. Conway, “Developing adaptive expertise: A synthesis of literature and implications for training,” *Proceedings of 13th International Conference on Naturalistic Decision Making*, pp. 81-86, June 2017.

[12] K. B. Carbonell, R. E. Stalmeijer, K. D. Könings, M. Segers, & J. J. van Merriënboer, "How experts deal with novel situations: A review of adaptive expertise," *Educational Research Review*, vol. 12, pp. 14-29, 2014.

[13] National Research Council. *Next Generation Science Standards: For States, By States.* Washington, DC: The National Academies Press, 2013, Available: https://doi.org/10.17226/18290

[14] R. E. Mayer, "Rote versus meaningful learning," *Theory into practice*, vol. 41(4), pp. 226-232, 2002.

[15] C. J. Atman, D. Kilgore, & A. McKenna, "Characterizing design learning: A mixed-methods study of engineering designers' use of language," *Journal of Engineering Education*, vol. 97(3), pp. 309-326, 2008.

[16] D. E. Stokes, *Pasteur's Quadrant – Basic Science and Technological Innovation*, Washington, DC: Brookings Institution Press, 1997.

[17] J. D. Bransford, & D. L. Schwartz, "Rethinking transfer: A simple proposal with multiple implications," *Review of research in education*, vol. 24(1), pp. 61-100, 1999.

[18] H. S. Broudy, "Types of knowledge and purposes of education," in *Schooling and the acquisition of knowledge*, R.C. Anderson, R.J Spiro, & W.E. Montague, Eds., Hillsdale, NJ: Erlbaum, pp. 1-17, 1977.

[19] D. Schwartz, J. Bransford, & D. Sears, "Efficiency and innovation in transfer." in *Transfer of learning: Research and perspectives*, J. Mestre, Ed., Greenwich, CT: Information Age Publishing, 2005.

[20] D. Opre, "Adaptive Expertise: Efficiency and Innovation*," Cognition, Brain, Behavior,* vol. 19(2), pp. 115-128, 2015.

[21] G. Hatano & K. Inagaki, "Two courses of expertise," in *Child development and education in Japan*, H. A. Stevenson, & K. Hakuta, Eds., New York: Freeman, pp. 262-272, 1986.

[22] M. G. Pandy, A. J. Petrosino, B. A. Austin & R. E. Barr, "Assessing adaptive expertise in undergraduate biomechanics," *Journal of Engineering Education*, vol. 93(3), pp. 211-222, 2004.

[23] F. T. Fisher & P. L. Peterson, "A tool to measure adaptive expertise in biomedical engineering students," *Proceedings of the 2001 American Society for Engineering Education Annual Conference*, Albuquerque, NM, 2001.

[24] L. Neely, *Adaptive design expertise: A theory of design thinking and innovation.* (Doctoral dissertation), Stanford University: Palo Alto, CA, 2007.

[25] T. D. Vannorsdall, D. A. Maroof, B. Gordon & D. J. Schretlen, "Ideational fluency as a domain of human cognition," *Neuropsychology*, vol. 26(3), pp. 400-405, 2012.

[26] R. B. Hill, “The design of an instrument to assess problem solving activities in technology education,” *Journal of Technology Education*, vol. 9(1), pp. 31-46, 1997.

[27] W. C. Newstetter & W. M. McCracken, “Novice conceptions of design: Implications for the design of learning environments,” in *Design knowing and learning: Cognition in design education*, C. M. Eastman, W. M. McCracken, & W. C. Newstetter, Eds., New York: Elsevier, 2001.

[28] S. Mosborg, R. S. Adams, R. Kim, C. J. Atman, J. Turns, & M. E. Cardella, “Conceptions of the engineering design process: An expert study of advanced practicing professionals,” *Proceedings of the 2005 American Society of Engineering Education Annual Conference & Exposition*, Portland, OR, 2005.

[29] J. A. Turns, B. Sattler, K. Yasuhara, J. L. Borgford-Parnell, & C. J. Atman, “Integrating Reflection on Experience into Engineering Education,” *Proceedings of the 2014 American Society for Engineering Education Annual Conference*, Indianapolis, IN, June 2014.

[30] D. H. Cropley, A. J. Cropley, & B. L. Sandwith, “Creativity in the engineering domain,” in *The Cambridge Handbook of Creativity Across Domains*, J. Kaufman, V. Glaveanu, and J. Baer, Eds., New York: Cambridge University Press, pp. 261-275, 2017.

[31] S. R. Daly, E. A. Mosyjowski, & C. M. Seifert, “Teaching creativity in engineering courses,” *Journal of Engineering Education*, vol. *103*(3), pp. 417-449, 2014.

[32] C. D. Denson, J. K. Buelin, M. D. Lammi, & S. D'Amico, “Developing Instrumentation for Assessing Creativity in Engineering Design,” *Journal of Technology Education*, vol. 27(1), pp. 23-40, 2015.

[33] B. T. Christensen & L. J., “Fluctuating Epistemic Uncertainty in a Design Team as a Metacognitive Driver for Creative Cognitive Processes,” *CoDesign*, vol. 14(2). pp. 133-152, 2017, Available: http://dx.doi.org/10.1080/15710882.2017.1402060

[34] National Research Council. *Middle School Engineering Design Standards*, 2013. Available: https://www.nextgenscience.org/pe/ms-ets1-1-engineering-design

[35] J. L. Herman, P. R. Aschbacher, & L. Winters, *A practical guide to alternative assessment*, Alexandria, VA: ASCD Publishing, 1992.

[36] F. G. Nasab, “Alternative versus Traditional Assessment,” *Journal of Applied Linguistics and Language Research*, vol. 2(6), pp. 165-178, 2015, Available: http://www.jallr.ir/

[37] K. A. Douglas & S. Purzer, “Validity: Meaning and Relevancy in Assessment for Engineering Education Research,” *Journal of Engineering Education*, vol. 104(2), pp. 108-118, 2015.

[38] M. T. Kane, “Validation,” in B. Brennan, Ed., *Educational Measurement*, American Council on Education & Praeger: Westport, CT, 2006.

[39] American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (Eds.). *Standards for educational and psychological testing. American Educational Research Association*, 2014.

[40] S. Lane, *Performance assessment: The state of the art (SCOPE Student Performance Assessment Series*), Stanford, CA: Stanford University Press, 2010.

[41] S. Lane & C. A. Stone, "Performance assessments," in *Educational Measurement*, B. Brennan, Ed., American Council on Education & Praeger: Westport, CT, 2006.

[42] K. Egan, *An Imaginative Approach to Teaching*, San Francisco: Jossey-Bass, 2005.

[43] J. Chandler, A. D. Fontenot, & D. Tate, "Problems associated with a lack of cohesive policy in K-12 pre-college engineering," *Journal of Pre-College Engineering Education Research (J-PEER)*, vol. 1(1), pp. 40–48, 2011, Available at: http://docs.lib.purdue.edu/jpeer/vol1/iss1/5

[44] N. Gisev, J. S. Bell, & T. F. Chen, "Interrater agreement and interrater reliability: key concepts, approaches, and applications," *Research in Social and Administrative Pharmacy*, vol. *9*(3), pp. 330-338, 2013.

[45] P. E. Shrout & J. L. Fleiss, "Intraclass correlations: uses in assessing rater reliability," *Psychological bulletin*, vol. *86*(2), pp. 420-428, 1979.

[46] R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2020. Available at: https://www.R-project.org/

[47] T. K. Koo & M. Y. Li, "A guideline of selecting and reporting intraclass correlation coefficients for reliability research," *Journal of chiropractic medicine*, vol. 15(2), pp. 155-163, 2016.

[48] W. J. Popham, "Living (or dying) with your NCLB tests," *School Administrator*, vol. 60(11), pp. 10-14, 2003.

[49] G. W. Ellis, J. Piña, R. Mazur, A. Rudnitsky, B. McGinnis-Cavanaugh, I. Huff … K. C. Cormier, "Developing Transmedia Engineering Curricula using Cognitive Tools to Impact Learning and the Development of STEM Identity," *Proceedings of the 2020 American Society for Engineering Education Virtual Conference: At Home with Engineering Education*, June 2020.

[50] A. Stansell, B. Quintanilla, E. Zimmerman & T. Tyler-Wood, "Teaching engineering concepts through a middle school transmedia book," *TechTrends*, vol. 59(2), pp. 27-31, 2015.

[51] T. C. Dusold, *Applications of narrative to the engineering decision making process and the pedagogy of engineering education* (Master's thesis), Iowa State University: Ames, IA, 2008.

[52] G. W. Ellis, A. Rudnitsky, I. Huff, B. McGinnis-Cavanaugh & S. K. Ellis, "Engaging children in design thinking through transmedia narrative," *Proceedings of the 2018 American Society for Engineering Education Annual Conference and Exposition*, Salt Lake City, Utah, June 2018.

[53] B. Harn, D. Parisi & M. Stoolmiller, "Balancing fidelity with flexibility and fit: What do we really know about fidelity of implementation in schools?" *Exceptional Children*, vol. 79(2), pp. 181-193, 2013.

[54] C. L. O'Donnell, "Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K–12 curriculum intervention research," *Review of Educational Research*, vol. 78(1), pp. 33-84, 2008.

[55] D. Niemi, J. Wang, D. H. Steinberg, E. L. Baker & H. Wang, "Instructional sensitivity of a complex language arts performance assessment," *Educational Assessment*, vol.12(3-4), pp. 215-237, 2007.

[56] K. Csavina, C. Nethken & A. R. Carberry, "Assessing student understanding of reflection in engineering education," *Proceedings of the 2016 American Society for Engineering Education Annual Conference & Exposition*, June 2016.

[57] National Science Foundation, "Broadening participation in engineering (BPE)," *Directorate for Engineering*, 2017, Available: https://www.nsf.gov/funding/pgm_summ.jsp?pims_id=504870

Appendix I: Assessment Items

1. Some of your neighbors (people who live near you) have worked hard to raise money and buy some empty land on Springfield Avenue. Now they want to make it into something better. How can *you* help? You're a middle-school student who knows about engineering design – so you can give your neighbors ideas about how to solve their problem.

Thinking like an engineer, what would you say to your neighbors about their first step in figuring out how to use the land?

2. Think about what you have learned about the engineering design process. How has what you learned changed how you think about solving problems?

3. Using the space on this page, create a picture or diagram that shows your ideas about what the engineering design process is. Include as many details as you can remember.

Appendix II: Samples of Scoring Materials

Generate Task – Codes and Definitions			
New Codes	**Original Codes**	**Definition**	**Response Indicators**
Measuring	Evaluation Feasibility	Analyzing parameters of the problem by taking measurements and evaluating resources (costs, materials, etc.).	“find the area of the land”, “measure the land”, “figure out the budget”, “get your materials “do research”
Modeling	Modeling	Creating physical representations of potential solutions – most often by sketching, prototyping, or making blueprints.	“make a prototype”, “draw a blueprint”, “make a sketch”, “come up with a model”
Strategizing	Problem definition Idea Generation Decision Making	Seeking to better understand the problem, considering the value of various solutions, and setting goals.	“find the problem”, “define the problem”, “make a plan”, “think a solution”, “set a goal”, “buy a book on the topic”, “study the problem”, “think of the benefits”, “think of impacts”, “ask about side-effects”
Collaborating	Information gathering Communication	Speaking to and working with others, seeking feedback, and gathering information from expert sources.	“ask for my neighbors for ideas”, “see what problems they have with this land”, “ask ‘what do want to do?’”, “share ideas with neighbors”, “vote on an idea”, “talk with experts”, “talk with people who know”, “gather info from others”

Reflection Task - Codes and Definitions	
Collaborating and Community (CO) **Original Codes:** Brainstorming, Communicating, Seeking Information **Indicators:** Asking questions of others, working with others, getting feedback, seeking out experts **Definition:** Engaging with others openly in the engineering design process through asking questions and seeking to accumulate knowledge resources	**Planning Out a Design (PL)** **Original Codes:** Goal Setting, Planning, Understanding the Problem **Indicators:** Thinking first, not rushing in, understanding the problem, being organized, making a plan **Definition:** Using the engineering design process as the basis for promoting specific ways of idea-centered thinking, such as planning, strategizing, goal setting, and organization.
Making a Model (MO) **Original Codes:** Modeling, Prototyping, Sketching, Visualizing **Indicators:** Blueprints, drawing, sketching, making a model, making a prototype **Definition:** Creating representations of a problem to be solved, including sketches, physical models, drawings, blueprints, and prototypes	**Using a Process (PR)** **Original Codes:** Generating Alternatives, Iterating, Making Decisions **Indicators:** Going step-by-step, improving/fixing things, making things better, using the design process **Definition:** Relying on the engineering design as a set of activities and practices (such as coming up with alternatives and iterating) that help to solve problems in a concrete manner
Assessing and Measuring (AS) **Original Codes:** Evaluating, Identifying Constraints, Testing **Indicators:** Measuring, gathering evidence, gathering data **Definition:** Expressing assessment practices – taking measurements, gathering data, etc. – as being part of the engineering design process	**Building a Context (BC)** **Original Codes:** Imagining, Using creativity **Indicators:** helping others in the real world, seeking a career in engineering, learning more about math/science **Definitions:** Making connections with the engineering design process that take it outside the engineering classroom

Drawing Task Rubric					
Group A		**Group B**		**Group C**	
An Illustration (Does not appear to be related to engineering; demonstrates *misunderstanding*).		**A Blueprint** (Makes use of *engineering language*, but does not demonstrate *engineering design*).		**A Process** (Demonstrates a *clear sequence of steps* constituting an engineering design process).	
0	**1**	**2**	**3**	**4**	**5**
Poor Orientation	**Drawing or Statement**	**Basic Schematic**	**Advanced Schematic**	**Basic EDP Diagram**	**Advanced EDP Diagram**
No image or statement.	A written description or an image that does not relate to engineering or engineering design. **Examples Include:** A drawing of a house, A drawing of a flower, a written statement saying "IDK."	A response that includes very simple examples of engineering language use and imagery. This can include a student's attempts at making simple lists or taking notes. **Examples Include:** A drawing of a house labelled "building process", a list of the words "thinking, problem solving", a drawing of a playground with the label "skateboard diagram."	A response that demonstrates a student's attempt to draw a blueprint while also defining parameters. This includes labelling components, showing measurements, and/or listing materials. **Examples Include:** A skateboard with measurements on the wheels, a house with components listed in a chart, a drawing of a pen with its materials labelled.	A diagram that shows a sequence of at least two engineering design steps, but does so in a simplistic or rote manner, often described as "lines and labels." Diagrams may also contain extra engineering terms and descriptions of the steps of the engineering design cycle. **Examples Include:** A simple drawing of a 3-step EDP, a list that says "define, develop, optimize."	A diagram that adds context or sophistication to a student's EDP drawing. This may include extra illustrations that provide a "walk-through" of the EDP, that highlight an example from the student's experience, or that embellish the student's EDP. **Examples Include:** A fully illustrated EDP, a comic that shows the EDP process, an EDP with doodles next to some steps.