ASEE 2022 ANNUAL CONFERENCE
Excellence Through Diversity
MINNEAPOLIS, MINNESOTA, JUNE 26TH-29TH, 2022 ASEE
Paper ID #36824

A Comparison of Two Scenario-Based Assessments of Systems Thinking

Siddhant Sanjay Joshi (Graduate Research Assistant)

Siddhant, from Pune, India is a doctoral student pursuing his Ph.D. in the School of Engineering Education at Purdue University. Prior to starting his Ph.D., Siddhant completed his M.S. in Aeronautics and Astronautics from Purdue University and a B.E. in Mechanical Engineering from MIT World Peace University. To complement his academic experience, Siddhant has a year-long industry experience working as a Lean and Operational Excellence trainee at Sandvik Asia. At Purdue University, Siddhant is also an instructor for two courses at the Gifted Education Research and Resource Institute Summer Residential program and has recently introduced a new course for aspiring engineering students. Apart from academics, Siddhant currently serves as the Treasurer of the American Society of Engineering Education - Purdue Chapter and is a member of the Graduate Student Advisory Council with focus on ensuring a better engineering experience for undergraduate and graduate students.

Kirsten Davis

Kirsten Davis is an assistant professor in the School of Engineering Education at Purdue University. Her research explores the intentional design and assessment of global engineering programs, student development through experiential learning, and approaches for teaching and assessing systems thinking skills. Kirsten holds a B.S. in Engineering & Management from Clarkson University and an M.A.Ed. in Higher Education, M.S. in Systems Engineering, and Ph.D. in Engineering Education, all from Virginia Tech.

Lori Czerwionka

Elisa Camps Troncoso

Francisco J Montalvo

A Comparison of Two Scenario-Based Assessments of Systems Thinking

Abstract

Engineers face complex and multidisciplinary problems in the modern work environment. To understand and solve these complex problems, engineers require systems thinking skills that allow them to consider the interconnected technical and contextual factors. Therefore, it is important to provide engineering students with opportunities to develop these skills during their education. A part of this process is developing assessment approaches that can help instructors measure students' systems thinking ability. A variety of approaches have been used in the literature to assess the development of systems thinking, including surveys, interviews, design projects, and scenario-based instruments. Scenario-based assessments can offer a more in-depth view of student learning than typical surveys while also being faster to analyze than open-ended data such as interviews. However, a range of scenario-based assessments that are available claim to assess similar skills, making it challenging to identify which fits the needs of a particular educational context. To help address this challenge, we compared two scenario-based assessments: the Village of Abeesee scenario [1] and the Energy Conversion Playground (ECP) design task [2], to understand concepts of systems thinking emphasized by each instrument and how students' scores on the assessments are related. The participants in this study were 19 undergraduate engineering students enrolled in an interdisciplinary humanities-based engineering course in Spring 2021. We administered both scenario-based assessments at the start and end of the semester to examine the change in students' scores over time. We then compared the assessment results from each instrument by examining average scores for each of the systems thinking dimensions and also individual total scores on each assessment. Lastly, we compared the experience of scoring the assessments from the perspective of the instructor or researcher using the assessment. Based on our findings, we make recommendations about when an instructor might choose to use one assessment or the other. Our results can inform future research and assessment projects that aim to assess students' systems thinking skills by comparing both student outcomes and instructor experience for these scenario-based assessments.

Introduction

As the world has become a more connected and globalized space, problems that engineers address have become more complex and multi-dimensional [1]. Engineers are required to understand multiple contexts and academic disciplines to successfully solve modern engineering problems. They must conceive of problems and solutions within local and global systems, considering both contextual and technical aspects. This critical engineering skill is called systems thinking — a decision-making process that accounts for multiple dimensions of technical and contextual factors, their interactions with each other, and their interactions over time [3]. Therefore, there is a need to develop engineers who have awareness and knowledge of interdisciplinary domains and interaction among subsystems to solve complex problems [1].

To address the need for systems thinkers, engineering programs have developed courses that help students improve their systems thinking skills [4]–[6]. These courses help students account for multiple contextual elements beyond the traditionally emphasized technical aspects of a problem [2]. However, to determine if students have developed a systems thinking approach to solving problems, it is essential to measure their systems thinking skills. Therefore, in this paper, we compare two scenario-based assessment instruments to understand the concepts of systems thinking measured by each instrument and also determine how student performance compares across the two instruments. Finally, we evaluate the usability of each instrument from the perspective of the instructor or research team.

Literature Review

The real-world problems that engineers face today are complex and multi-dimensional. These problems are better addressed when decision-makers understand the various dimensions and their interdependencies [7]. Traditionally, engineers have prioritized the technical dimensions while solving problems over other considerations [8], [9]. However, with the increasing complexity of problems, the lines between engineering, humanities, and social sciences have blurred [1], [10]. In response to these shifts, government agencies have emphasized that graduates must be able to address complex problems, collaborate across disciplinary boundaries, and utilize critical thinking skills [11], [12]. To prepare the engineering workforce to solve complex engineering problems, engineering research and pedagogy can benefit from a systems thinking approach [13].

Systems thinking is the ability to see the world as a complex interconnected system where different dimensions influence and interact with each other [14], [15]. To solve complex real-world problems, engineers require knowledge of the technical as well as contextual dimensions [2]. These dimensions include technology, long-term and short-term perspectives, stakeholders, communities, cultures, politics, economics, and the environment, among others [5], [6], [16]–[20]. To assess if engineers understand interrelationships amongst different dimensions and employ systems thinking to solve problems, it is important to measure systems thinking abilities.

Various measurement and assessment strategies have been proposed to understand the development of systems thinking skills in engineering graduates and students. A few of these include course-based systems thinking instruction followed by assessment using surveys, field reports, reflections, open-ended examinations, quizzes, etc. [4]–[6], [20], [21]. Other strategies include cognitive mapping and concept mapping to explore how visualization techniques can aid the consideration of multiple dimensions of systems thinking [3], [22]. Another approach, which will be the focus of this paper, uses scenario-based assessments of systems thinking skills [1], [2], [19], [23]. Some advantages of using scenario-based assessments over self-reporting, observational, simulation-based, and other assessment methods are: (a) Ability to identify salient technical and contextual dimensions (b) Help evaluate the potential behaviors of participants in realistic situations (c) Less time consuming than some other approaches (d) Can be administered to a large number of students simultaneously (e) Include both open-ended and closed-ended questions (f) Easier to score due to a well-defined scoring rubric [2]. Moreover, scenario-based assessments can indicate growth in dimensions of systems thinking over a short period [16].

Multiple scenario-based assessments are available that claim to assess similar skills, making it challenging to identify which assessment fits the needs of a particular educational context. Furthermore, there are few comparisons of existing scenario-based assessments of systems thinking [3]. Therefore, in this paper we will address this gap by comparing two scenario-based assessment instruments: the Village of Abeesee scenario [1] and the Energy Conversion Playground (ECP) design task [2]. We address the following research questions:

1. Which concepts of systems thinking are assessed by each instrument (Abeesee and ECP)? How do they align and differ from each other?
2. Do the instruments assess students' systems thinking skills in different ways, as evidenced by a comparative analysis of class average scores and individual students' scores?
3. How does the experience of using the Abeesee and ECP scenario-based assessments compare?

Methods

In this study, we compared two scenario-based assessments of systems thinking: the Village of Abeesee [1] and the Energy Conversation Playground [2]. We implemented both scenario-based assessments in a pre/post test study design within a semester-long humanities-informed engineering course. To compare these assessments, we examined the theoretical concepts each instrument assesses, considered the scores that students received on the assessments, and reflected on the experience of scoring the assessments. The Purdue University IRB approved this investigation.

Participants

In Spring 2021, 19 students enrolled in a one-credit course called Humanities-Informed Engineering Projects (for more information on the course, refer to [16]). The students were multilingual and had various engineering majors. Nine of the 19 students were women.

Data Collection

Two scenario-based assessments were administered to all the students enrolled in the course. Data were collected from the participants using an online questionnaire at the start and end of the semester. In the questionnaire, students were asked to respond to two scenarios: Village of Abeesee [1] and Energy Conversion Playground [2]. Each instrument assesses systems thinking along multiple dimensions that explore both the technical and contextual elements of defining and solving a problem.

Village of Abeesee Scenario (Abeesee)

The Abeesee scenario problem statement (Figure 1) was developed by Grohs et al. based on the Dimensions of Systems Thinking Framework [1]. It describes the challenges faced by residents of a fictitious village during a five-month winter period. For the Abeesee scenario, the participants were prompted to answer eight open-ended questions with two to five sentences each such that it showcased their analysis of the problem statement (See Appendix A for list of questions). The students were asked to consider various aspects of problem-solving such as problem definition, constraints, implementation and decision-making challenges, and probable solutions [1]. The scoring of the responses is based on a series of rubrics developed by Grohs et al. based on the seven dimensions as shown in Table 1. For each dimension, the rubrics provide a guide for how to score students' responses on a three-point scale (0–3). A score of three (3) on the Abeesee scale requires mentioning both the contextual and technical aspects of a problem along with the interaction between them while a zero (0) constitutes no response to the given prompt. For more details see the original paper [1] by Grohs et al..

Figure 1

Village of Abeesee Scenario (reproduced from [1])

Problem statement for the village of Abeesee:

The Village of Abeesee has about 50,000 people. Its harsh winters and remote location make heating a living space very expensive. The rising price of fossil fuels has been reflected in the heating expenses of Abeesee residents. In fact, many residents are unable to afford heat for the entire winter (5 months). A University of Abeesee study shows that 38% of village residents have gone without heat for at least 30 winter days in the last 24 months. Last year, 27 Abeesee deaths were attributed to unheated homes. Most died from hypothermia/exposure (21), and the remainder died in fires or from carbon monoxide poisoning that resulted from improper use of alternative heat sources (e.g., burning trash in an unventilated space).

Table 1

List of Systems Thinking Dimensions for Abeesee scenario [1]

Dimension	Definition
Problem Identification	Refers to a respondent's ability to describe perceptions of the problems and/or issues facing Abeesee
Information Needs	Refers to a respondent's ability to identify additional context/information beyond the details provided in the scenario that is needed to address the problem identified
Stakeholder Awareness	Refers to a respondent's ability to identify and include relevant stakeholders and the role that they will play in the problem and solution identification, planning, and implementation process
Goals	Refers to a respondent's ability to identify short- and long-term goals towards addressing the problems and/or issues of the scenario
Implementation Challenges	Refers to a respondent's ability to identify expected barriers to their crafted response to the Abeesee scenario

Unintended Consequences	Refers to a respondent's ability to demonstrate flexibility in being self-critical and identifying possible blind spots of an attempted solution and the degree to which a respondent explored possible limitations and unintended consequences
Alignment	Refers to the degree to which a respondent incorporates aspects of the problem identified in responses to goals and plans

Energy Conversion Playground Scenario (ECP)

The ECP scenario problem statement relates to an energy production problem in a developing country and is shown in Figure 2. The students were asked to list five considerations to account for when solving the problem and describe the most important consideration. The ECP design scenario is meant to capture students' sociotechnical considerations to solve the problem [2]. Although the authors do not describe the assessment using the term "systems thinking", their definition of sociotechnical thinking aligns with the systems thinking definition we are using for this paper and used by Grohs et al. [1] .Thus, we believe that it is reasonable to compare these assessments as they focused on similar constructs despite some variation in terminology. This comparison will advance the understanding of systems thinking scenario-based assessments.

The scoring of the responses for the ECP scenario is based on the rubric developed by Mazzurco et al. [2] and consists of three dimensions: *Technology*, *People*, and *Broader Context*, as shown in Table 2. For each dimension, students' responses were scored on scale of 0–3 (See [2] for instructions). For instance, participants were assigned a score of three (3) for the *Technology* dimension if they mentioned the breadth of considerations that focused on all the four technical categories, or a score of zero (0) if no mention of any of the technical category was provided. A similar scoring strategy was followed for the dimensions of *People* and *Broader Context* as well.

Figure 2

Energy Conversion Playground Scenario (reproduced from [19]*)*

> In developing countries, energy production is one of the most critical problems. Resources and technologies to produce energy are not often available. Thus, human power conservation systems might be used to power small appliances. Imagine you and your team are assigned to a design project in partnership with a Non-Governmental Organization of a developing country. The NGO needs a low-cost power system that can generate enough energy for the lights of a primary school. One of the members of your team suggests using a merry go round, seesaw, and swing to produce energy that can be converted to electricity for the lights.
>
> **Question**: What considerations do you need to take into account to solve the problem described in the scenario? List and describe all the constraints and justify their inclusion

Table 2

Dimensions for ECP scenario (modified from [2])

Score	Key Characteristics of the Domain
Technology	Considerations focused on four technical categories: • *Inputs or constraints to the technology:* Power requirements, time of operation, cost, materials, safety, climate, people as a source of energy input, and so forth. • *Functionality:* Efficiency, feasibility, ease of operation, maximum power generated, friction, storage of energy, functioning of components, alternative techs to meet same goals, ability to generate the needed energy output, and so forth. • *Long-term technological considerations:* Maintenance, repairs, spare parts, upgrades, and so forth • *Additional considerations added during the current study:* Durability, Focus on system safety/equipment safety; people as part of the larger system; funding, budget, cost of maintenance and operation, etc.
People	• Considerations focused on stakeholders' needs, desires, expertise, and degree of participation in the design process (e.g., listening to the community, hearing their voices, collaborating with them in the design process, etc.). • *Additional considerations added during the current study:* Focus on the safety of people; the willingness of people participation, the influence of people on the playground system
Broader Context	Considerations focused on four contextual categories: • *Local norms:* Social norms, culture, gender/ethnic/power dynamics, religious views, and so forth. • *Ethics and law:* Regulations, standards, laws, moral and ethical issues. • *Other socio-material contexts:* Built environment, impact on the natural environment, local economy, education system • *Additional considerations added during the current study:* Political aspects (under ethics and law), Profitability, Ability to own or produce the technology in a financial sense.

Note: More information on the *Additional Considerations added during the current study* can be found in results section Research Question 3.

Data Analysis

The first step in the data analysis process was to score both scenarios using the rubrics provided by their original developers. The participant information was deidentified before scoring responses. For both scenarios, two members of the research team scored each response. They worked to identify agreed-upon definitions for the terms in the rubrics, adding additional clarification where needed (see *Additional Considerations* in Table 2; more details provided in the Results). After scoring, they met to discuss the ratings and reach a final agreement.

To address Research Question 1, we analyzed the concepts of systems thinking present within each of the scenario-based assessments. For Research Question 2 about how the instruments measure student performance, we compared the class averages for each of the instruments' dimensions on the pre- and post-test administrations. We also analyzed individual students' overall scores for each scenario-based assessment by summing the scores across all the dimensions within each assessment. We responded to Research Question 3 by reflecting on and analyzing our own experiences with the instruments.

Results

The purpose of this paper is to compare two scenario-based assessments based on the concepts they assess, students' performance on the assessments, and the experience of using the assessments as an educator or researcher. We have organized this section around the three research questions addressed in this study.

Research Question 1: Which concepts of systems thinking are assessed by each instrument (Abeesee and ECP)? How do they align and differ from each other?

The Abeesee and ECP instruments assess seven and three dimensions of systems thinking, respectively; however, these dimensions overlap in many ways. The clearest example of conceptual alignment between the two assessments is the overlap between the *Stakeholder Awareness* dimension of the Abeesee instrument and the *People* dimension of the ECP instrument. The *Stakeholder Awareness* dimension refers to the community members who are end-users of the services as well as the experts who can give advice [1]. Similarly, the *People* dimension includes considerations that focus on stakeholders of the project, their needs, desires, and the expertise of individuals involved in the scenario [2]. Thus, it seems that the two dimensions measure similar concepts of systems thinking.

There are additional conceptual alignments between other dimensions of these instruments, although these are not as completely aligned. For example, the Abeesee *Information Needs* dimension aligns with the ECP *Broader Context* because both dimensions explore the range of contextual aspects associated with a problem, such as social, cultural, environmental, or legal considerations [1], [2]. However, one difference between these dimensions is that the Abeesee *Information Needs* dimension also assesses technical aspects of the problem, while *Technology* is a separate dimension in the ECP assessment. Similarly, the Abeesee *Goals* dimension aligns with *Long-term technical considerations* from the ECP *Technology* dimension because both dimensions address the long-term goals of the solution. However, the Abeesee *Goals* dimension also includes contextual aspects of both the long-term and short-term goals, whereas those aspects are considered in separate dimensions in the ECP assessment. Finally, the Abeesee *Implementation Challenges* dimension and the *Constraints* aspect of the ECP *Technology* dimension both address challenges, yet the view in the ECP is limited to technological challenges only, whereas the Abeesee instrument assesses the integration of technical and contextual challenges. The remaining dimensions of the Abeesee and ECP instruments have a smaller overlap as compared to the above-mentioned dimensions.

In summary, we found that many of the dimensions measured by the Abeesee instrument are also measured by the ECP instrument to a certain degree. The comparison also indicates that whereas a main goal of the Abeesee instrument is to intertwine technological and contextual thinking, the

ECP evaluates them separately by including *Technology* as its own dimension. The Abeesee instrument also includes a more nuanced reference to time (for e.g., short- and long-term, interconnection among different points in time of the project).

Research Question 2: Do the instruments assess students' systems thinking skills in different ways, as evidenced by a comparative analysis of class average scores and individual students' scores?

Pre-and Post Test Class Averages Across Instrument Dimensions

The class's average scores increased for three of the systems thinking dimensions in the Abeesee scenario: *Information Needs*, *Goals*, and *Alignment*, and for two dimensions in the ECP scenario: *People* and *Broader Context* (see Table 3 and Table 4). In contrast, the average scores for the Abeesee *Stakeholder Awareness*, Abeesee *Implementation Challenges*, and ECP *Technology* dimensions declined between the start and end of the semester. Paired *t*-tests comparing the pre- and post-test scores indicate that the change was positive and statistically significant ($p < .05$) for three dimensions (shown in green) in the Abeesee scenario and two dimensions (shown in green) for the ECP scenario. The results are summarized in Table 3 and Table 4.

Table 3

Comparing Pre- and Post-Test Scores for the Abeesee Scenario (Scale of 0–3)

Abeesee Dimension		Pre-Test		Post-Test					
	df	*Mean*	*SD*	*Mean*	*SD*	Diff	*t*	*p*	Effect Size*
Problem Identification	18	1.53	.68	1.53	.68	0.00	0.00	.500	0.00
Information Needs	18	1.74	.71	2.11	.43	**0.37**	1.79	.045*	0.53
Stakeholder Awareness	18	1.58	.88	1.16	.92	*-0.42*	-1.80	.044*	-0.45
Goals	18	1.95	.51	2.21	.41	**0.26**	1.42	.086	0.56
Implementation Challenges	18	2.05	.69	1.68	.65	*-0.37*	-1.79	.045*	-0.54
Alignment	18	1.68	.73	1.95	.89	**0.26**	1.16	.131	0.32

Note: Green & bold cells = increase from pre to post test. Orange & italic cells = decrease from pre- to post-test.
*Effect Size measured using Hedges's *g*

Table 4

Comparing Pre- and Post-Test Scenario Scores for the ECP scenario (Scale of 0–3)

ECP Dimension		Pre-Test		Post-Test					
	df	*Mean*	*SD*	*Mean*	*SD*	Diff	*t*	*p*	Effect Size*
Technology	18	2.53	.60	2.21	.69	*-0.32*	-1.68	.055	-0.48
People	18	.47	.75	1.11	.91	**0.63**	2.27	.018*	0.74
Broader Context	18	.16	.49	.74	.78	**0.58**	3.64	.001*	0.87

Note: Green & bold cells = increase from pre to post test. Orange & italic cells = decrease from pre- to post-test.
*Effect Size measured using Hedges's *g*.

Examining these results, we observe that the class scores increased for at least one-third of the systems thinking dimensions for the two instruments, indicating that both assessments identified some development in systems thinking. Considering the overlapping dimensions between assessments (i.e., results of RQ1), the class average results point to different findings across Abeesee *Stakeholder Awareness* and ECP *People* and also Abeesee *Goals* and ECP *Technology*. This indicates that while overlapping conceptually, these dimensions do not seem to measure the same concept in the data. The Abeesee *Implementation Challenges* dimension and ECP *Technology* dimension (which includes the concept of constraints) both decreased over time, yet because *Technology* includes many other aspects beyond the constraints, it is unclear to what degree these findings indicate overlap across dimensions. Finally, the class average scores for Abeesee *Information Needs* and ECP *Broader Context* both show development over the course and a significant change over time, which supports the notion that these dimensions may measure similar concepts across assessments. The increase in the Abeesee *Information Needs* dimension represents a shift in students thinking, from thinking about information needs within either a technological or contextual perspective to thinking about the integration between the two types of information needs. The increase in the ECP *Broader Context* score indicates that more of students' top five considerations were related to contextual aspects. Putting these results together, it is likely that students increased their consideration of the context as it relates to information needs or main considerations over the course of the semester, a finding that was measured by both assessments.

More generally, we noticed greater growth in class average scores for the ECP instrument when compared with the Abeesee instrument. The increase in class average scores for the ECP scenario could be attributed to the method through which the quality of the score was determined for each assessment. To receive a full score on the dimensions of the Abeesee scenario, students had to address the breadth and depth (complexity) of both the technical and contextual factors across multiple aspects for each dimension, thus making it harder to score full points. In contrast, it was relatively easier to achieve a full score on the ECP scenario because only the dimension of *Broader Context* was scored for the breadth and depth of both the technical and contextual considerations [2]. Thus, it may be because the Abeesee rubrics required more complex thinking for a top score that students showed more growth on the ECP assessment generally. Additionally, the scoring system of the Abeesee instrument assessed students' ability to intertwine their analyses of various types of considerations, whereas the ECP assessed the three dimensions in an isolated fashion for the most part. The greater increases in the ECP scores may also reflect the course's goals to increase students' awareness of people and context and/or a lack of focus in the course on considering the interaction of different contextual and technical factors in a situation.

Individual Overall Scores Across Instruments

As a second approach to Research Question 2, we examined the overall scores for individual students on each instrument for the pre- and post-test administrations. It is important to note that neither the Abeesee nor ECP instruments were originally designed with the intent of calculating an overall score (i.e., by summing the scores from the individual dimensions). We are not advocating for the use of these overall scores in assessing systems thinking but chose to calculate this value as a way to explore whether individual students scored similarly across the two scenario-based assessments. Specifically, we wanted to know a) whether the same students had high scores on both instruments and b) whether students' scores on both instruments moved in the same direction between the pre-and post-test.

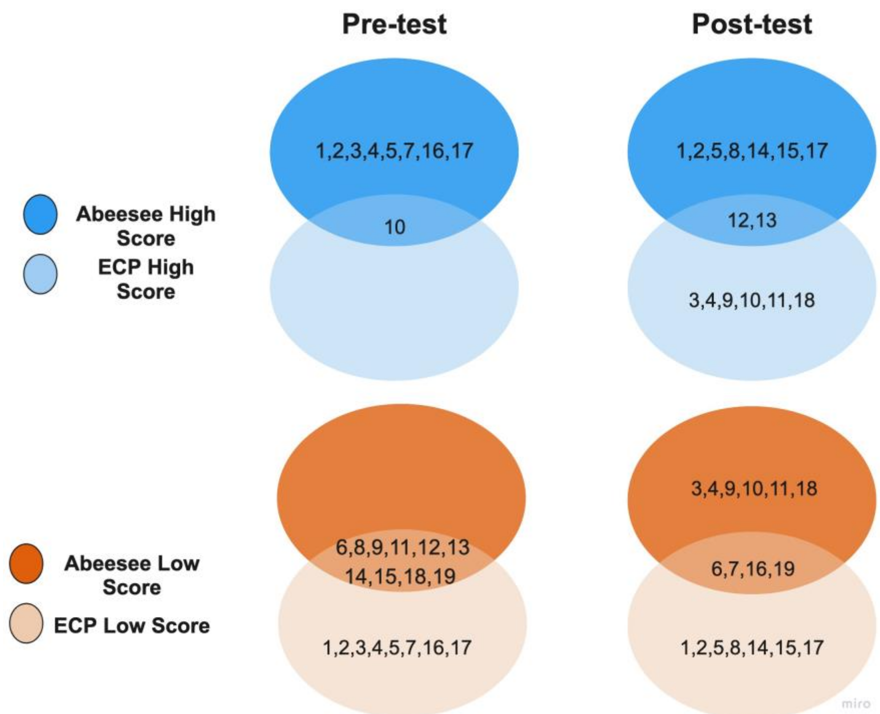
To explore these questions, we divided the students into "high" and "low" scoring groups for both instruments for both the pre- and post-test. The Abeesee overall scenario scores (i.e., the sum of all the individual dimension scores) ranged from 0 to 18 and the ECP scores ranged from 0 to 9. Any student who scored greater than the middle score on the Abeesee instrument and ECP instrument was placed in the "high" category. Students who scored less than or equal to the middle score were placed in the "low" category.

The results of our individual student score comparison are shown in Figure 3. For the pre-test scores, the ECP scenario had 18 low score and the Abeesee scenario had 10 low scores. Many of the same students scored low on both instruments, as shown by the large number of participants in the middle section of the bottom left Venn diagram. In contrast, only one participant scored high on both the instruments, while the remaining participants who scored high on the Abeesee scenario did not score high on the ECP scenario (shown in the top left Venn diagram).

In the post-test scores, we noticed that both instruments have nine students in the high-scoring category. For the Abeesee instrument, the number of participants in the high-scoring category remained unchanged. Five of the participants who scored low on the pre-test shifted to high scores on their post-test for the Abeesee instrument. Even though the number of participants in the high scoring category for the ECP instrument increased to eight in the post-test, only two participants were in the high scoring category for both the Abeesee and ECP scenario (as shown in the top right Venn diagram). We also see less overlap in the participants who scored low on these instruments (bottom right Venn diagram). Therefore, the results indicate that even though the dimensions of systems thinking measured by the two instruments have some commonality in assessing students' performance, they are not identical to each other.

Figure 3

Pre- and Post-Test Classification of Students into a High and Low Scoring Groups for Abeesee and ECP Scenarios



From our analysis of individual students' scores, we observe that student scores on the two assessments did not necessarily move in the same direction at the individual level. Moreover, an overall increase in the number of students with high scores on the ECP instrument suggests that humanities-integrated engineering course was better aligned with the dimensions addressed in the ECP scenario like *People* and *Broader Context*.

Research Question 3: How does the experience of using the Abeesee and ECP scenarios compare?

As our team scored the student responses on both scenarios, we observed several differences between the two that might influence which scenario an instructor or researcher would choose to use in a specific context. In this section, we will discuss four such differences between the scenarios: the depth of students' responses, the clarity of the rubrics for the two instruments, the time required to score students' responses, and the experience of scoring the assessments from the perspectives of researchers from different disciplines.

The team of researchers who were scoring both the Abeesee and ECP scenarios noticed that student responses for the Abeesee scenario were more elaborate as compared to the ECP scenario. The typical length of responses for the Abeesee instrument prompts were two to three lines as compared to a single line or one-word response for the ECP instrument. A possible reason for this could be that students focused more on listing their five considerations rather than providing a justification for their inclusion. This limited the depth to which student responses could be understood by the rater and thus were scored accordingly following the guidelines of the rubric.

When considering the usability of the rubrics for these scenario assessments, the raters found that the terminology used in the Abeesee instrument rubric was easier to interpret relative to the ECP rubric. The terminology in the ECP rubric overlapped amongst its dimensions, and therefore the coders had to redefine some of the terms to ensure consistency of scoring. For instance, the term ‘safety’ is defined in the ECP instrument rubric as a constraint under the *Technology* dimension (see Table 2 for ECP dimensions). However, student responses mentioned safety in the context of equipment as well as in the context of people. To help interpret this different usage, we chose to classify the safety of equipment and the safety of people as independent characteristics of safety. Safety in the context of equipment meant the ability of the people to use the equipment safely without damaging it, whereas safety of the people considered whether the playground or equipment was safe for people to operate. Therefore, the responses about the safety of equipment were scored as the *Technology* dimension and the safety of people in the *People* dimension. Many such examples were identified by the researchers while initially scoring the responses and thus were defined separately for each dimension (see Table 2, “Additional considerations added during current study” for specific examples).

Finally, the raters belonged to different disciplines (engineering and the humanities). For the ECP scenario, the raters noticed that the context of a student response can be interpreted differently based on the discipline of the rater. For instance, one participant wrote the following as a response to the ECP prompt:

> Do the children have enough time to spend playing on the playground for it to be feasible? (If the area is poor, maybe the children do not have time to play on the playground because they must help their family.)

The rater from the engineering background interpreted this as a part of the *time of operation constraint* (*Technology* dimension) while the rater from a non-engineering background interpreted it as the *People* dimension. However, once the terminology was more clearly defined and agreed upon by both the raters, it was easier and quicker to score the ECP scenario as compared to the Abeesee scenario. The additional scoring time was necessary for the Abeesee scenario because:

1. The raters had to score the Abeesee scenario for seven dimensions as compared to three for the ECP scenario
2. The Abeesee scenario scoring criteria had greater complexity relative to the ECP scenario
3. The typical response length for the Abeesee scenario was longer than for the ECP scenario

In addition, some Abeesee dimensions such as *Stakeholder Awareness*, *Goals, Unintended Consequences,* and *Alignment* took longer to score because they required the rater to interpret information from responses to multiple prompts. However, even though it was harder to score the Abeesee scenario, this instrument allows for the assessment of greater complexity of thinking within each of the systems thinking dimensions, thereby providing a more detailed assessment of systems thinking skills.

Discussion

In this paper, we have compared two scenario-based assessment instruments, the Abeesee scenario [1] and the ECP scenario [2]. In this comparison, we sought to understand which concepts associated with systems thinking were emphasized by each instrument, how the students' scores compared across the two instruments, and the experience of using these instruments as a researcher or instructor. We administered the two scenario-based assessments at the start and end of the semester for a *Humanities-Informed Engineering Projects* course developed at Purdue University during Spring 2021. In comparing the two instruments conceptually, we found that there is overlap between a few of the dimensions of systems thinking that they assess. In particular, we observed that the dimensions of systems thinking assessed by the ECP instrument were a subset of the dimensions assessed by the Abeesee instrument. We also found that the Abeesee instrument assessed greater complexity of systems thinking by requiring integration of technical and contextual aspects to achieve high scores within most dimensions. In contrast, the ECP instrument divided these components into separate dimensions so that integrative thinking was not required to achieve high scores.

Our comparison of the pre- and post-test scores indicated that students improved in multiple dimensions across both instruments. However, the growth in scores was more considerable for the ECP scenario. We believe that the greater increase in ECP scores could be because it is relatively easier to score highly on the instrument due to the less complicated scoring criteria (i.e., students are not required to demonstrate integrative thinking). In addition, because the ECP instrument addresses fewer dimensions of systems thinking, it is more likely that a single-semester course could impact these dimensions. The Abeesee scenario addresses a wider range of skills, and it may not be reasonable to expect students to develop significantly along several dimensions within the context of a single course.

To examine if the same students demonstrated an increase in their systems thinking scores across the two instruments, we classified the students into high and low categories based on their scores in the pre- and post-test of both the instruments. We found that even though the number of students with high scores increased in the post-test, only two students were in the high scoring category for *both* the Abeesee and ECP instruments. Our results suggest that even though the dimensions of systems thinking assessed by these instruments have some conceptual overlap, they may not identically measure these concepts. It is possible that the differences in the level of complexity assessed between the two instruments may account for some of this gap, but further research would help to better understand the differences between how these instruments are assessing similar-sounding constructs.

Lastly, we also compared the experience of scoring the responses for these two instruments. We observed that it was generally quicker to score the ECP instrument responses. However, due to some overlap between its dimensions, the terms in the ECP rubric were not always clearly defined and thus, it was necessary to redefine some additional terms. Although it was faster for us to score the ECP scenario, it was also clear that the Abeesee instrument provides a more detailed assessment of systems thinking skills as compared to the ECP instrument, both in terms of a greater number of dimensions and a greater focus on integrated thinking within each dimension. Based on our experiences, we recommend that instructors or researchers determine whether speed or level of detail is more important for your assessment needs when deciding what type of scenario-based assessment to use in your context. It would also help to consider what learning outcomes are the focus of your course or research study. The ECP scenario may fit better for assessing a shorter-term program or a single course, whereas the Abeesee scenario may be appropriate for a more advanced course or longitudinal assessment over the course of a program of study.

Acknowledgments

We acknowledge the National Endowment for the Humanities for the partial funding of this project (AKA-270141-20), and we express our gratitude to the students who participated.

References

[1] J. R. Grohs, G. R. Kirk, M. M. Soledad, and D. B. Knight, “Assessing systems thinking: A tool to measure complex reasoning through ill-structured problems,” *Think. Ski. Creat.*, vol. 28, pp. 110–130, Jun. 2018, doi: 10.1016/j.tsc.2018.03.003.

[2] A. Mazzurco and S. Daniel, “Socio-technical thinking of students and practitioners in the context of humanitarian engineering,” *J. Eng. Educ.*, vol. 109, no. 2, pp. 243–261, 2020, doi: 10.1002/jee.20307.

[3] K. E. Dugan, E. A. Mosyjowski, S. R. Daly, and L. R. Lattuca, “Systems Thinking Assessments: Approaches That Examine Engagement in Systems Thinking,” presented at the 2021 ASEE Virtual Annual Conference Content Access, Jul. 2021. Accessed: Aug. 30, 2021. [Online]. Available: https://peer.asee.org/systems-thinking-assessments-approaches-that-examine-engagement-in-systems-thinking

[4] M. D. Bedillion, K. H. Muci-Kuchler, C. M. Birrenkott, M. Lovett, and L. O. Pottmeyer, “A Combined Online Learning / In-Class Activity Approach to Teach Systems Thinking and Systems Engineering Skills to Freshman Engineering Students,” presented at the 2021 ASEE Virtual Annual Conference Content Access, Jul. 2021. Accessed: Sep. 08, 2021. [Online]. Available: https://peer.asee.org/a-combined-online-learning-in-class-activity-approach-to-teach-systems-thinking-and-systems-engineering-skills-to-freshman-engineering-students

[5] J. C. Evans, “Exploring the Human Dimension of Engineering Through the Built Environment,” presented at the 2018 ASEE Annual Conference & Exposition, Jun. 2018. Accessed: Jun. 23, 2021. [Online]. Available: https://peer.asee.org/exploring-the-human-dimension-of-engineering-through-the-built-environment

[6] J. A. Leydens, K. Johnson, S. Claussen, J. Blacklock, B. M. Moskal, and O. Cordova, “Measuring Change over Time in Sociotechnical Thinking: A Survey/validation Model for Sociotechnical Habits of Mind,” presented at the 2018 ASEE Annual Conference & Exposition, Jun. 2018. Accessed: Jun. 23, 2021. [Online]. Available: https://peer.asee.org/measuring-change-over-time-in-sociotechnical-thinking-a-survey-validation-model-for-sociotechnical-habits-of-mind

[7] W. Currie and B. Galliers, *Rethinking Management Information Systems: An Interdisciplinary Perspective*. OUP Oxford, 1999.

[8] W. Faulkner, “Dualisms, Hierarchies and Gender in Engineering,” *Soc. Stud. Sci.*, vol. 30, no. 5, pp. 759–792, Oct. 2000, doi: 10.1177/030631200030005005.

[9] A. L. Pawley, “Universalized Narratives: Patterns in How Faculty Members Define ‘Engineering,’” *J. Eng. Educ.*, vol. 98, no. 4, pp. 309–319, 2009, doi: 10.1002/j.2168-9830.2009.tb01029.x.

[10] M. Hynes and J. Swenson, “The Humanistic Side of Engineering: Considering Social Science and Humanities Dimensions of Engineering in Education and Research,” *J. Pre-Coll. Eng. Educ. Res. J-PEER*, vol. 3, no. 2, Oct. 2013, doi: 10.7771/2157-9288.1070.

[11] National Academy of Engineering, *The Engineer of 2020: Visions of Engineering in the New Century*. Washington, DC: The National Academies Press, 2004. doi: 10.17226/10999.

[12] National Research Council, *Education for Life and Work: Developing Transferable Knowledge and Skills in the 21st Century*. Washington, DC: The National Academies Press, 2012. doi: 10.17226/13398.

[13] M. C. Davis, R. Challenger, D. N. W. Jayewardene, and C. W. Clegg, “Advancing socio-technical systems thinking: A call for bravery,” *Appl. Ergon.*, vol. 45, no. 2, Part A, pp. 171–180, Mar. 2014, doi: 10.1016/j.apergo.2013.02.009.
[14] P. M. Senge, *The fifth discipline: the art and practice of the learning organization*. New York: Doubleday/Currency, 2006.
[15] L. B. Sweeney and J. D. Sterman, “Bathtub dynamics: initial results of a systems thinking inventory,” *Syst. Dyn. Rev.*, vol. 16, no. 4, pp. 249–286, 2000, doi: 10.1002/sdr.198.
[16] K. A. Davis *et al.*, “Integrating the humanities with engineering through a global case study course,” *J. Int. Eng. Educ.*, vol. 3, no. 1, p. Article 4, 2021.
[17] J. Handley and E. B. Moje, ““What the Problem Really Was…": A Preliminary Exploration of Youth Problem Definition in Everyday Contexts,” presented at the 2018 CoNECD - The Collaborative Network for Engineering and Computing Diversity Conference, Apr. 2018. Accessed: Jun. 30, 2021. [Online]. Available: https://peer.asee.org/what-the-problem-really-was-a-preliminary-exploration-of-youth-problem-definition-in-everyday-contexts
[18] D. H. Jonassen, “Toward a design theory of problem solving,” *Educ. Technol. Res. Dev.*, vol. 48, no. 4, pp. 63–85, Dec. 2000, doi: 10.1007/BF02300500.
[19] A. Mazzurco, J. Huff, and B. Jesiek, “The Energy Conversion Playground (ECP) Design Task: Assessing How Students Think About Technical and Non-technical Considerations in Sustainable Community Development,” *Int. J. Serv. Learn. Eng.*, vol. 9, no. 2, pp. 64–84, Sep. 2014, doi: 10.24908/ijsle.v9i2.5585.
[20] C. Skokan, M. Simoes, and J. Crocker, “Designing Humanitarian Engineering Classes,” *undefined*, 2006, Accessed: Jun. 23, 2021. [Online]. Available: https://www.semanticscholar.org/paper/Designing-Humanitarian-Engineering-Classes-Skokan-Simoes/5b2ce9ed882b6ab48cba47b486b13181b75eb2ad
[21] E. Barsky, A. Berndt, A. Greenwood, and C. Paterson, “What information sources do engineering students use to address authentic sociotechnical problems,” presented at the 2011 ASEE Annual Conference & Exposition, Vancouver, Canada, Jun. 2011.
[22] S. Gray *et al.*, “Assessing (Social-Ecological) Systems Thinking by Evaluating Cognitive Maps,” *Sustainability*, vol. 11, no. 20, Art. no. 20, Jan. 2019, doi: 10.3390/su11205753.
[23] K. Davis *et al.*, “The Lake Urmia vignette: a tool to assess understanding of complexity in socio-environmental systems,” *Syst. Dyn. Rev.*, vol. 36, no. 2, pp. 191–222, 2020, doi: 10.1002/sdr.1659.

Appendix A

List of questions that students respond to as part of the Village of Abeesee scenario assessment, taken from [1].

1. Given what you know from the scenario, please write a statement describing your perception of the problems and/or issues facing Abeesee.
2. What additional information do you need before you could begin to develop a response in Abeesee? Consider both detail and context of the problems/issues you identified.
3. What groups or stakeholders would you involve in planning a response to the problems/issues in Abeesee?
4. Please briefly describe the process you would use to plan a response to the problems/issues in Abeesee.
5. What would you expect a successful plan to accomplish?
6. Given what you know and a budget of $50,000, develop a plan that would address the Abeesee situation maximizing the impact of your $50,000. Use a numbered, step-by-step guide, recipe-style to explain.
7. What challenges do you see to implementing your plan? What are the limitations of your approach?