2021 ASEE ANNUAL CONFERENCE
Virtual Meeting | July 26–29, 2021 | Pacific Daylight Time
Paper ID #32549

Systems Thinking Assessments: Approaches That Examine Engagement in Systems Thinking

Ms. Kelley E. Dugan, University of Michigan

Kelley Dugan is a Ph.D. candidate in the Department of Mechanical Engineering at the University of Michigan. She has a B.S. in Mechanical Engineering from The Ohio State University. Before pursuing her Ph.D., Kelley worked in the consumer appliance industry for two years. Her current work focuses on the development and assessment of systems thinking skills. Research interests also include front-end design practices, socially engaged design, and sustainable design.

Dr. Erika A. Mosyjowski, University of Michigan

Erika Mosyjowski is a research fellow and lecturer focusing on engineering education at the University of Michigan. She earned a B.A. in sociology and psychology from Case Western Reserve University and an M.A. and Ph.D. in Higher Education from the University of Michigan. Her research interests include cultural beliefs about what engineers do and who they are, students' career thinking and trajectories, and ways to effectively facilitate more diverse, inclusive, and equitable engineering environments.

Dr. Shanna R. Daly, University of Michigan

Shanna Daly is an Associate Professor in Mechanical Engineering at the University of Michigan. She has a B.E. in Chemical Engineering from the University of Dayton and a Ph.D. in Engineering Education from Purdue University. Her research characterizes front-end design practices across the student to practitioner continuum and studies the impact of developed front-end design tools on design success.

Dr. Lisa R. Lattuca, University of Michigan

Lisa Lattuca, Professor of Higher Education and member of the Core Faculty in the Engineering Education Research Program at the University of Michigan. She studies curriculum, teaching, and learning in college and university settings, particularly how faculty attitudes, beliefs, and cultures influence curricular and instructional practices and how these in turn affect student learning.

Systems Thinking Assessments:

Approaches That Examine Engagement in Systems Thinking

While systems engineers rely on systems thinking skills in their work [1], given the increasing complexity of modern engineering problems, engineers across disciplines need to be able to engage in systems thinking [2], [3], including what we term *comprehensive* systems thinking. *Comprehensiv*e systems thinking is a holistic problem-solving approach wherein connections and interactions between constituent parts of the problem and the immediate work, stakeholder needs, broader contextual aspects (e.g., social and political) and possible impacts over time are identified and included in decision making [4], [5]. Due to the inherent complexity of systems thinking, and more specifically comprehensive systems thinking, it is not easy to know how well students (and practitioners) are learning and leveraging systems thinking approaches. Engineering managers and educators can benefit from systems thinking assessments. Research has established that assessments aid in understanding abilities and student progress [6], [7] and support the development of curriculum and training materials [8]. In addition, organizations use systems thinking assessments to inform job placement based on capability and/or fit [9], [10]. More generally, assessments provide critical information that can be used to advance the field of engineering [11].

Building from an understanding that engineers need systems thinking skills to address complex engineering problems, our research is aligned with best practices in curriculum and training material development. Once a desired result is identified, in this case the goal is to develop engineers who are able to use *comprehensive* systems thinking knowledge and skills to address complex problems, the next step is to determine how the achievement of that goal will be assessed [8]. Such assessment(s) then guide the development of learning activities and experiences, e.g., methods for teaching systems thinking [8]. Our analysis sought to understand the ways in which existing systems thinking assessments relevant in an engineering context attend to various dimensions relevant to *comprehensive* systems thinking.

A variety of systems thinking assessments exist that are relevant to engineers, including some focused on the demonstration of systems thinking knowledge or skills and others measuring attitudes, interests, or values related to systems thinking. However, many of the assessments that operationalize broader conceptions of systems thinking—e.g., *comprehensive* systems thinking—focus on attitudes or preferences rather than the application of systems thinking skills [12]. Given the importance of understanding engineers' actual engagement in and skill development related to systems thinking, we focus on characterizing existing behavior-based assessments—those related to the examination of demonstrated knowledge or skills—of systems thinking using a *comprehensive* systems thinking lens. Starting with a collection of systems thinking assessments from a systematic literature review [12] conducted by our team, we analyzed in-depth those behavior-based assessments that included the creation of a visual representation and were open-ended, i.e., it did not presuppose or provide answers. Across behavior-based assessments the creation of a visualization was common and including the open-ended requirement made the collection of assessments more amenable to comparisons.

The findings from this in-depth analysis of systems thinking behavior-based assessments identified 1) six visualization types that were leveraged, 2) dimensions of systems thinking that were assessed and 3) tensions between the affordances of different assessments. In addition, we consider the ways assessments can be used. For example, using assessments to provide feedback to students or using assessments to determine which students are meeting defined learning goals. We draw on our findings to highlight opportunities for future *comprehensive* systems thinking behavior-based assessment development.

Background

While there are many definitions of systems thinking in use, frequently, research on systems thinking in engineering focuses on the aspects of the proximal problem context (e.g., [13]–[15]). For example, Frank and Elata [14, p. 191] describe an engineer who has a capacity for systems thinking as someone who "understands the whole system beyond its single components (part, box, card, element) and understands how the single component functions as part of the entire system or assembly." If you imagine systems thinking as taking of holistic view of only the proximal, most immediate aspects of a problem, you may visualize that as "drawing a box" around those proximal aspects. For example, if an engineer is working on a generator, she might draw the box around all the physical components that work together to deliver the generator's function: produce electricity.

At the same time, systems thinking research in engineering often underplays the importance of incorporating contextual aspects of a problem or system. One exception is Grohs et al.'s [16] work that includes constructs such as awareness of stakeholders as well as contextual aspects such as legal, political, and cultural. However, support for regular use of a more comprehensive view of systems thinking in engineering is evident from several areas. For example, some fields tend to think of systems thinking as integrating connections between both technical and contextual factors into their decision-making processes (e.g. [17], [18]). In addition, several engineering studies recognize the importance of integrating context into engineering solutions [19]–[21]. Some of these studies focus specifically on contextual competence or "an engineer's ability to anticipate and understand the constraints and impacts of social, cultural, and environmental, political, and other contexts on engineering solutions" [20, Sec. Introduction]. Although contextual competence has not been specifically tied to systems thinking, the recognition of its importance for engineers aligns with reasons to include broader contextual considerations as a facet of the engineers' systems thinking skills.

The way we think about systems thinking emphasizes consideration of both the proximal problem context and broader contextual factors such as political or social factors. Thus, *comprehensive* systems thinking means drawing a much bigger box that encompasses more than a problem's proximal aspects. Continuing with the generator example, drawing a bigger box means that in addition to all the physical components that deliver the generator's function, the engineer now includes factors such as the environmental impacts of different fuel sources, who has access and is able to use the generator, what regulations or laws apply to the production and use of the generator, and so forth, in her analyses and problem-solving process. *Comprehensive* systems thinking pushes for a broadening of what is conceptualized as part of a system and a recognition that understanding constituent elements of the immediate context and interactions between those elements is necessary but not sufficient.

Research Methods

Study Goals. The goal of this study was to characterize existing, behavior-based assessments of systems thinking by focusing on the benefits and challenges afforded by different assessments. For this analysis, we focused on a subset of behavior-based assessments that included the creation of a visual representation and did not presuppose or provide answers. In other words, these were open-ended assessments where the expected structure of the analysis may be specified (e.g., which visualization should be used) but the content of the analysis is not provided (e.g., no specification of which concepts to include in a concept map).

The following research questions guided our analysis of the assessments:

RQ1: What visualization forms are used in behavior-based systems thinking assessments?
RQ2: What dimensions of systems thinking are included in behavior-based systems thinking assessments?
RQ3: What are the affordances of different approaches to assessing systems thinking?

Data Collection. The six assessments analyzed in the current paper were identified from our team's larger systematic literature review of systems thinking assessments relevant to engineering contexts [12]. Each assessment has a single source paper that we drew on for analysis of the assessment. One additional paper by Watson et al. [22] is not included as an assessment source paper but is referenced throughout our analyses as one assessment drew from the scoring methods outlined by Watson et al. Our literature review identified 27 unique systems thinking assessments, six of which met this study's inclusion criteria: (1) examined demonstrated knowledge or skill, (2) included the creation of at least one visual representation where the type of visualization is not specific to the problem, and (3) were open-ended. We aimed to capture the breadth of existing systems thinking assessments, and thus assessments that used and/or referenced the terms "*systems* thinking" or "*system* thinking," were included in what will be referred to in this paper as "systems thinking assessments." We summarize both the assessments and an overview of the study contexts in which the assessments were used as follows:

1. Lavi et al.'s [23] assessment required students to create conceptual models using object-process methodology (OPM). In one study using this assessment [23], Lavi et al. investigated the creation of a rubric for assessing the systems thinking of engineering students and the use of the rubric for students' conceptual models of systems. One hundred and forty-two undergraduate engineering students majoring in Industrial Engineering and Management or Information Systems Engineering in a course on information systems were split into teams of three to six people, forming 32 teams. These teams created two conceptual models of a selected "freely available, consumer focused Web-based information system," [23, p. 40] where the first one was submitted in the middle of the semester and the second one was submitted at the end of the semester. Systems thinking attributes were classified in system function, structure, or behavior and each attribute was scored from "0 (no expression of attribute understanding)" to "3 (full expression of attribute understanding)" [23, p. 42]. While the detailed scoring guidelines

were not provided in the source paper, we included this assessment because descriptions of each attribute were provided.

2. Gray et al.'s [24] assessment asked for the creation of "fuzzy cognitive maps," the generation of scenarios based on the fuzzy cognitive maps, and short essays describing leverage points. In the study that used this assessment, Gray et al. demonstrated guidelines for teaching and measuring systems thinking skills. The study had 40 student participants who were in an introductory sustainability science class. Most of the students were freshman or sophomore undergraduates and approximately one-third were STEM majors. Our analysis focused on the fuzzy cognitive mapping of "scientific and popular articles" [24, p. 6] that were created using a free online software, www.mentalmodeler.org, because it included a visual representation of a system's structure and informed the creation of a graphical representations of system changes. While no formal guiding criteria was provided to the ten university faculty who independently ranked the cognitive maps, we inferred which dimensions of systems thinking were assessed based on patterns in the map rankings.

3. Hu and Shealy's [25] assessment entailed the completion of two paper-based concept maps while wearing a functional near-infrared spectroscopy (fNIRS) system. In the study that used this assessment, Hu and Shealy compared multiple methods for measuring systems thinking including three concept map scoring methods, a self-report survey, and cognitive activation. Twenty-eight undergraduate engineering students, who were a mixture of freshmen and seniors, participated in the study. Our analyses focused on the creation of concept maps related to sustainability topics such as renewable energy and water availability used because it was the only behavior-based assessment. Hu and Shealy [25] drew on existing concept map scoring methods reported by Watson et al. [22]. Watson et al.'s work is not considered a systems thinking assessment on its own. Watson et al.'s [22] traditional, holistic, and categorical scoring methods are only discussed in the context of Hu and Shealy's [25] Assessment.

4. The assessment developed by Brandstädter, Harms, and Großschedl's [26] was focused on a concept-mapping task. This assessment was part of Brandstädter, Harms, and Großschedl's examination of how the medium and directedness of concept-mapping practices affect concept-mapping performance and the validity of using concept-mapping for system thinking assessment. The authors looked at combinations of computer-based versus paper-based and highly directed versus nondirected concept maps, and used a questionnaire of open-ended and multiple-choice questions to assess procedural and structural system thinking ability. Study participants consisted of 154 fourth graders and 93 eighth graders from German primary and secondary schools. In our analyses we focused only on the nondirected, paper-based version because it was the only combination that met both our visualization and open-ended inclusion criteria. Although the rubric for scoring propositions in the concept maps focused specifically on proposition accuracy, we included this as a systems thinking assessment because the product moment correlations were calculated between proposition accuracy scores and the procedural and structural system thinking scores from the questionnaire.

5. Rehmann et al.'s [27] assessment was a group assessment, in which the groups completed rich pictures, causal loop diagrams, and behavior over time graphs for topics such as renewable energy and clean water. This assessment was part of a study where Rehmann et al. evaluated the success of introducing systems thinking to students in an engineering scholars program during freshman-seminar (n=21) and reinforcing it during a sophomore-seminar (n=20) by using qualitative observations and quantitative assessments. Our analyses focused on the sophomore seminar because a rubric was provided for scoring rich pictures, causal loop diagrams, and behavior over time graphs.

6. Vanasupa, Rogers, and Chen's [28] assessment was also a group assessment, in which groups were tasked to draw rich pictures. This assessment was part of a preliminary study on teaching and measuring systems thinking skills. The pilot study included a mixture of 15 undergraduate materials engineering majors from all four academic levels (freshman through senior). Each of the two groups, one group of 7 females and one group of 8 males, drew rich pictures of "a 'successful' engineering student and an 'unsuccessful' engineering student" [28, p. 2].

Data Analysis. The six assessments were thematically analyzed according to our three research questions. For RQ1 the themes were readily identified as the names for the different visualizations used. Data was compiled in a word document that organized visualizations' definitions, uses, and examples. For RQ2 and RQ3 an inductive analysis approach was used, where dimensions of systems thinking and affordances were identified, grouped based on commonalities, and iteratively refined to the descriptions provided below. The initial categories for dimensions of systems thinking and affordances were developed based on multiple research team reviews and discussions.

Findings

In this section we present descriptions of the different visualizations used, salient dimensions of systems thinking examined, and affordances identified across the six assessments, respectively addressing RQ1, RQ2, and RQ3. The majority of assessments analyzed here (5 of 6) included the creation of only one type of visualization. One assessment by Rehmann et al. [27] used three different types of visualizations: rich pictures, causal loop diagrams, and behavior over time graphs.

RQ1: Types of Visualizations

There were six types of visualizations used across the six assessments: 1) rich pictures, 2) concept maps, 3) causal loop diagrams, 4) fuzzy cognitive maps, 5) conceptual maps, and 6) behavior over time graphs. The first five types described below can all be thought of as variations of concept maps, in that they all display multiple elements or concepts and represent connections between those elements, and are presented in order of increasing structure and rules. The last visualization, behavior over time graphs, is different in that it focuses on the time-varying behavior of a single system variable, rather than how that variable connects to or interacts with other aspects of a system.

Rich Picture. A rich picture is a visualization tool for showing the numerous elements in a real-world situation [29]–[31]. There are no formal modeling symbols that people need to use, though in in some spaces the use of standard symbols has become accepted [30]. Rich pictures were the least structured visualization among the five variations of concept maps, and used in two of the six assessments—Vanasupa's and Rehmann's. This lack of structure aligns with how Checkland discouraged the use of pre-defined visuals in rich pictures because individuals need to find the "ways which are as natural as possible for them as individuals [to develop 'rich pictures']," [29, p. 22]. Rich pictures rely on the rationale that "pictures are a better medium than linear prose for expressing relationships. Pictures can be taken in as a whole and help to encourage holistic rather than reductionist thinking about a situation," [29, p. 22]. The example rich picture provided by Vanasupa et al. [28] was a hand drawing that included sketches, symbols (arrows, plus sign "+"), and (some) text, see Figure 1. Rehmann et al. [27] provided two examples of rich pictures, one that was created digitally and another than was hand drawn. The digital rich picture included digital art, a graph, symbols (arrows, money sign) and text. Figure 1 shows a rich picture that was made to represent "an 'unsuccessful' engineering student" [28, Fig. 1].

While rich pictures are characteristic of Soft Systems Methodology (SSM) and their use has been helpful during exploratory discussions regarding the investigation of a situation, its stakeholders, and issues [29], Rehmann et al.'s [27] paper did not relate their work to SSM. Vanasupa et al.'s study included an overview of SSM and introduced rich pictures through "Checkland's specific concept of information gathering" [28, p. 2].

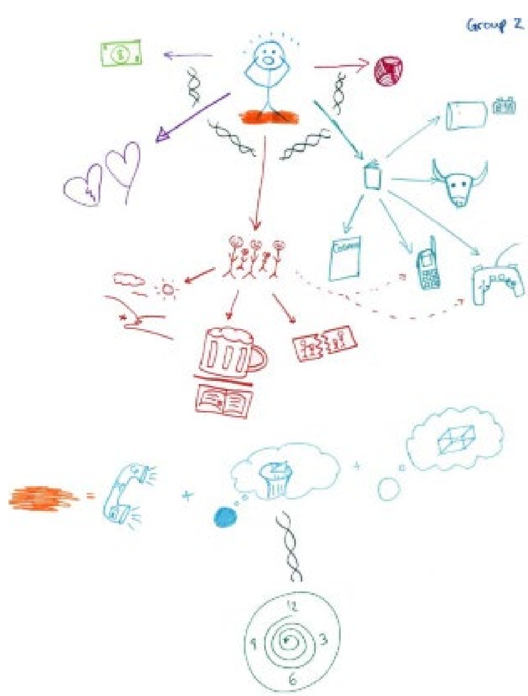

Figure 1: Rich Picture [28]

Concept Map. Concept maps are graphical tools that do not have to conform to a specific structure (i.e., they can have a hierarchical versus nonhierarchical arrangement) [22]. A key aspect of concept maps are their propositions, which consist of two concepts connected by a linking line or arrow and a description of the connection [22], [26]. Two of the six assessments in our study leveraged concept maps. Hu and Shealy described the action of concept mapping as beginning with “a main idea and then branches out to show how that main idea can be broken down into specific topics and drawing links between concepts at various hierarchical levels within the map” [25, Sec. Introduction]. While one of the scoring methods used by Hu and Shealy [25]—the traditional scoring method [22]—inherently values a hierarchical structure, Brandstädter, Harms, and Großschedl acknowledged that the concept map’s net structure can be hierarchical or nonhierarchical [26], [32]. Figure 2 shows a fuzzy cognitive map [26, Fig. 4] that was made to describe the “development, enemies, living and feeding of eggs, larvae, young and adult blue mussels” [26, p. 2151].

Figure 2: Concept Map [26]

Causal Loop Diagram. Causal loop diagrams “describe basic causal mechanisms hypothesized to generate the reference mode of behavior of the system over time” [33, p. 37]. Causal loop diagrams are similar to concept maps in that both show how one concept is connected to another, but they are different in that causal loop diagrams “depict how changes in one concept are linked to changes in another,” [28, p. 1]. In their study, Rehmann et al. [27] examined causal loop diagrams that built upon the rich pictures created by student teams. They used “the notation “s” and “o” to indicate whether concepts connected with an arrow change in the *same* or *opposite* direction,” (emphasis original) [27, Sec. Introduction]. Figure 3 shows a causal loop diagram that was made by students working on levels as protection against flooding; delays in relationships are represented by double lines on the connecting arrows [27, Fig. 5].

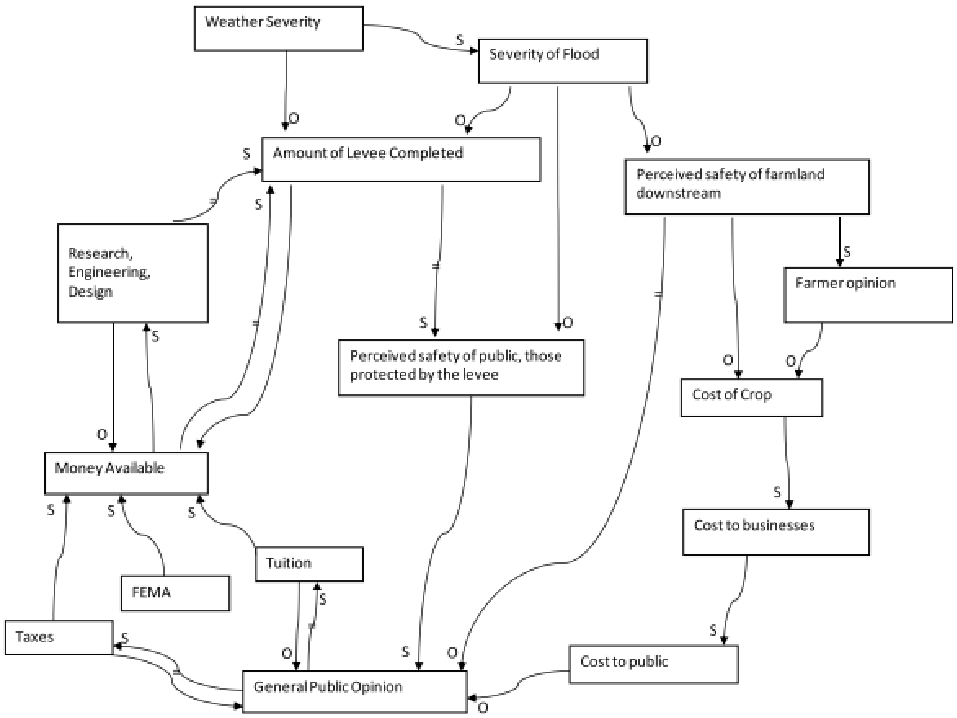

Figure 3: Causal Loop Diagram [27]

Fuzzy Cognitive Map. A fuzzy cognitive map is a semi-quantitative cognitive mapping technique that represents "systems as directed and weighted graphs, where the nodes of the graph qualitatively represent elements of the system (i.e., concepts), and the edges [links] between the nodes quantitatively represent the direction and strength of causal relationships between concepts" [24, p. 7]. A fuzzy cognitive map is similar to a causal loop diagram in that the concepts can increase or decrease in quality and quantity [24], [27]. However, there is a notational difference in fuzzy cognitive maps in that the relationships (links) between concepts represent positive or negative influences that act in the direction of the arrow [24]. The strength of this influence is quantified and this quantification enables analysis of a system's function when using the scenario setting at www.mentalmodeler.org. One of the six assessments in our study used fuzzy cognitive maps—Gray's. Figure 4 shows a fuzzy cognitive map that was made to represent students understanding of the structural links between "climate change, natural resource availability, and terrorism" [24, Fig. 2].

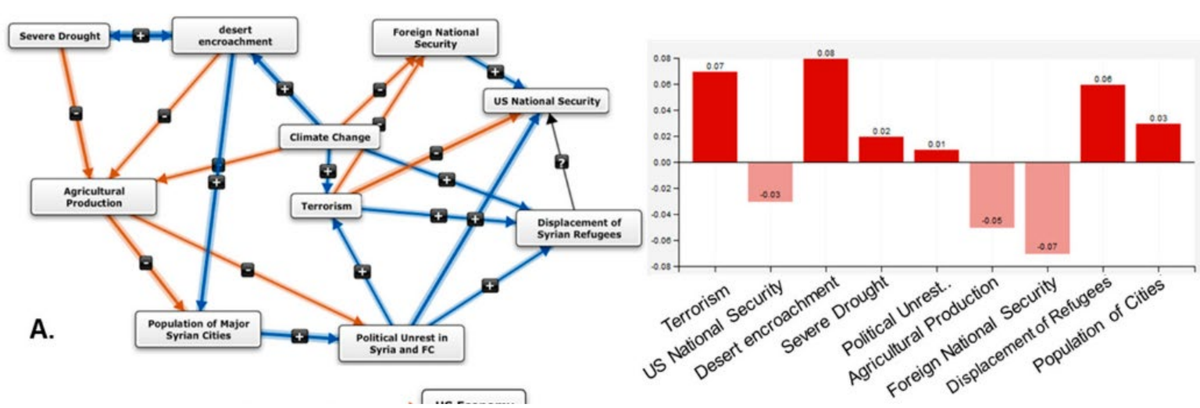

Figure 4: Fuzzy Cognitive Map (left) and a Scenario Output (right) [24]

Conceptual Model. "Conceptual models are products of the system representation process in model-based systems engineering (MSBE)" [23, p. 40]. Conceptual models created using object-process methodology (OPM) are more expressive than concept maps because different types of concepts (nodes)—objects or processes—and relationships (links)—structural or procedural relations—are clearly distinguished [23], [34]. OPM is a formal language and methodology that "builds on a minimal set of concepts: stateful objects—things that exist, and processes—things that happen and transform objects by creating or consuming them or by changing their states," [35, p. v]. In addition, interpretation of conceptual model meaning is straightforward as OPM is ISO 19450, which means that there are "two semantically equivalent modalities of representation for the same model: graphical and textual," [36]. Conceptual models were the most structured visualization among the five variations of concept maps and were used in one of the six assessments in our study—Lavi's. Figure 5 shows the second detail level of a conceptual model of Dropbox[1], specifically looking at File Sharing, created using OPM [23, Fig. 4].

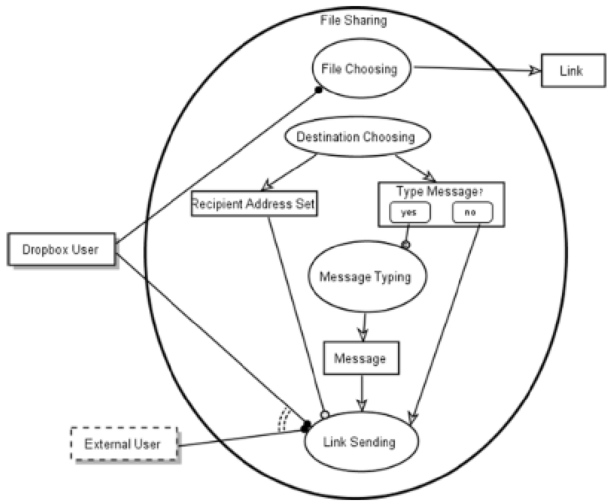

Figure 5: Conceptual Model Created Using OPM [23]

Behavior over time graphs. Behavior over time graphs are "usually schematic depictions of how an important variable behaves over time, although they can also include real data" [28, p. 1]. In their study, Rehmann et al. [27] examined behavior over time graphs and noted that a person's understanding of dynamic behavior can be supported by tracing a loop in a causal loop diagram. For example, consider heat conducted from inside a warm building to the outside environment assuming the heat transfer rate can be calculated using $\dot{Q} = \frac{k\Delta T}{L}\sum_{i=1}^{6} A_i$ [37, pp. 491–492]. In this case,

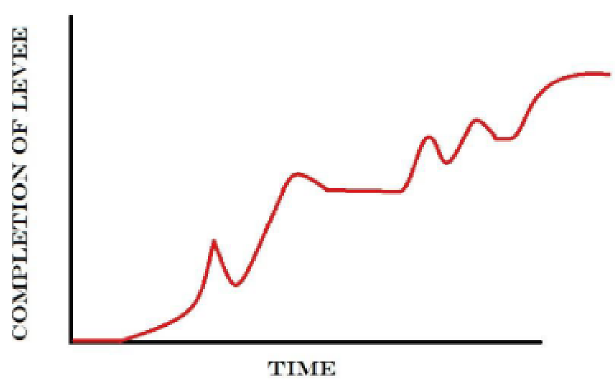

Figure 6: Behavior Over Time Graph [27]

[1] www.dropbox.com

where the temperature difference, ΔT, is not zero, an increase in the rate of heat transferred, $\dot{Q}$, from inside the building to the outside environment leads to an increase in energy used to heat the building, which leads to an increase in utility and social costs, which might cause a building manager to invest in new insulation with a lower thermal conductivity, k. The heat loss would decrease and the investment in new insulation would continue to increase until the utility and social costs are balanced with the cost of adding new insulation. Figure 6 shows a behavior over time graph that was made by students working on levels as protection against flooding to represent completion of the levee over time [27].

RQ2: Dimensions of Systems Thinking

Across all six assessments we saw attention to three dimensions of systems thinking, including: 1) identifying individual elements or aspects of a problem, 2) identifying connections or relationships between these elements, and 3) accounting for time and change over time. These three dimensions were further broken down to show a number of ways in which problem complexity—"the number of issues, functions, or variables involved in the problem; the degree of connectivity among those properties; the type of functional relationships among those properties; and the stability among the properties of the problem over time" [38] as quoted in [39, pp. 67–68]—was recognized and analyzed.

Identifying Individual Elements. One dimension of systems thinking that was evaluated was the identification of individual aspects of a problem or system. Assessments referred to these different elements in various ways, in that students had to identify the "things" [23], "concepts" [25], [26], "nodes" [26], and "elements" [27] that represent and describe a problem or system. For example, when using the traditional scoring method [22] in Hu & Shealy's [25] assessment, the number of concepts was combined with other metrics into a single score. Lavi et al.'s assessment scored multiple attributes from 0 to 3 related to identifying function, structure and behavior, including: "A_2 – Main function", "A_4 – Main object and its sub-objects", and "A_8 – Temporary objects and decision nodes," [23, p. 42] which was enabled by how object-process methodology (OPM) distinguishes between elements that are objects and elements that are processes. Gray et al.'s [24] assessment implicitly valued identification of elements: cognitive maps with higher systems thinking assessments generally had more concepts.

In addition to identifying individual elements, some assessments evaluated the depth with which individual elements were considered. For example, Hu and Shealy's [25] assessment demonstrated explicit consideration of depth when the holistic scoring method was used as it included evaluating whether there was "sufficient detail" [22, p. 129] in the various dimensions (types of issues) included in the concept map. Hu and Shealy's [25] assessment also demonstrated explicit consideration of depth when the traditional scoring method [22] was used as it included the highest level of hierarchy—the number of concepts in the longest path down from a proposition that includes the topic of the concept map—in the equation for the total score. Lavi et al.'s assessment evaluated depth through attribute "A_3 – Complexity levels," which is the "number of levels of detail," [23, p. 42] under the assumption that modeling deeper levels of a system enables the expression of more complex system behaviors. In addition, consideration of depth was implicit across all the attributes in Lavi et al.'s assessment in that each attribute was scored on a scale from zero, or "no expression of attribute understanding", to three, that was "full expression of attribute understanding" [23, p. 42].

A couple assessments examined the breadth in terms of types of issues considered across all the identified elements. In Rehmann et al.'s assessment the evaluation of rich pictures included whether there were elements from at least five of the following seven types of issues: "engineering, social, ethical, cultural, environmental, business, and political issues" [27, Sec. Introduction]. Another example was Hu and Shealy's [25] assessment when Watson et al.'s [22] holistic scoring method was used to evaluate how many different dimensions were included in a concept map. These dimensions included "economic, environmental/natural resources, social, and temporal (requires inclusion of present and future considerations)" and "values, spatial imbalances, technology, education, actors and stakeholders" [22, p. 129], [40]. Also, Hu and Shealy's [25] assessment showed consideration of breath when the categorical scoring method was used as it calculated a concept distribution value for each of the following categories: "environment, resource (scarcity), social impact, values, future (temporal), spatial unbalances, education, stakeholders, technology, and economy" [22, p. 130], [41].

Identifying Relationships. Another dimension of systems thinking attended to was the identification of relationships or connections between elements. In Brandstädter, Harms, and Großschedl's [26] assessment, the assessment of individual elements was inextricably tied into the assessment of relationships as concept maps were evaluated using the relational scoring method [42], which scores concept maps based on the correctness of propositions—which consist of two concepts connected by a linking line or arrow and a description of the connection [22], [26]. Gray et al.'s [24] assessment implicitly valued identification of relationships as cognitive maps with higher systems thinking assessments generally had more connections between concepts.

There were a number of ways assessments went deeper than the identification of relationships. One way was to evaluate the depth with which relationships were considered. For example, Rehmann et al.'s assessment implicitly showed consideration of depth in that the rubric included requirements such as "connections drawn suggest careful thought and contemplation" and "argue convincingly for the relationships" [27, Sec. Observations] across rich picture and causal loop diagram score descriptions, respectively.

In addition to valuing the identification of relationships, some assessments valued the different types of relationships or specific kinds of relationships. For example, distinguishing between relationships that show structure and relationships that show behavior or specifically looking for feedback loops. Lavi et al.'s assessment scored multiple attributes from 0 to 3 related to relationships, including: "A_5 – Structural relations: Links between objects and links between processes" and "A_6 – Procedural relations: Links between objects and processes" [23, p. 42]. While Hu and Shealy's [25] assessment demonstrated explicit consideration of feedback loops when the holistic scoring method was used as it evaluated "sophisticated branch structure and connectivity" [22, p. 129] by the existence of cross-links and feedback loops. While checking for the existence of feedback loops was not explicitly included in the rubric for Rehmann et al.'s [27] assessment, it was implicitly valued in the authors' discussions of the presence or absence of feedback in rich picture responses and in their use of causal loop diagrams, as causal loop diagram "represent the feedback loop systems diagrammatically" [33, p. 37].

Valuing connectedness refers to when assessments valued the complexity of the overall network of relationships. For example, Hu and Shealy's [25] assessment evaluated connectedness when the traditional scoring method was used by looking at the number of cross-links— "descriptive linking lines that create propositions by joining two concepts from different map hierarchies" [43], [44] as quoted in [22, p. 121]. Gray et al.'s [24] assessment implicitly valued connectedness as cognitive maps with higher ratings for the level of systems thinking generally had a higher value for the ratio of the number of connections to the number of concepts. One assessment combined considerations of connectedness with considerations of breadth—in terms of the type of issues covered. When Watson et al.'s [22] categorical scoring method was used, Hu and Shealy's [25] assessment evaluated complexity—equated with "overall coverage of and connectedness between the categories" [22, p. 124]—where connectedness between categories was determined by looking at the number of interlinks, which are "connections between concepts from different categories" [22, p. 124].

Accounting for Time and Change Over Time. Accounting for time refers to considering contextual factors that may change with time, e.g., the consideration of a systems impact on future generations [40]. While accounting for change over time includes identifying and analyzing the dynamic, or time-varying behavior of a system of interest [27]. At least one type of temporal consideration— "accounting for time" or "change over time" —was a part of each of the six assessments studied. Accounting for time may look like including elements in a concept map related to how a problem affects present versus future generations in different ways [22], [25], while accounting for change over time may look like including temporary objects in a conceptual model [23], or sketching out the behavior of a key variable over time [27], or including feedback loops in a concept map [22], [25]. Hu and Shealy's [25] assessment accounted for time and demonstrated "present and future considerations" [22, p. 129] contribute to a comprehensive understanding of a subject area as such considerations were one of the dimensions in the holistic scoring method's consideration of how many different dimensions were included in a concept map. In addition, Hu and Shealy's [25] assessment showed attention to both time and change over time when Watson et al.'s categorial scoring method was used as a concept distribution value (percentage of the number of concepts in a specific category to the total number of concepts) was calculated for the "future (temporal)" category which included "future/past generations, scenario analysis, forecasting" [22, p. 130].

One way in which Lavi et al.'s assessment accounted for change over time was by explicitly evaluating understanding of dynamic behavior through scoring the attribute "A8 – Temporary objects and decision nodes" [23, p. 42]. Rehmann et al.'s [27] assessment explicitly evaluated consideration of dynamic behavior by scoring scenarios shown in behavior over time graphs. In the "Identifying Relationships" section, above, we discussed how Lavi et al.'s [23] assessment analyzed structure and behavior by using OPM's ability to distinguish between objects and processes as well as structural and procedural relations. Gray et al.'s [24] assessment takes another approach to analyzing structure and behavior. Gray et al.'s assessment implicitly valued element identification, relationship identification, and connectedness and therefore implicitly valued understanding system structure. Their assessment provided students feedback on their understanding of a system's dynamic behavior by looking at the scenario output of a fuzzy cognitive map (see www.mentalmodeler.org) [24]. Hu and Shealy's [25] assessment showed explicit attention to change over time when using the holistic scoring method [22] that looked for the existence of feedback loops. In addition, feedback and dynamic behavior was implicitly valued in Rehmann et al.'s [27] assessment as demonstrated by their use of causal loop diagrams—the structure of which enables identification of "the principal feedback loops of the systems" [33, p. 181]. This is also supported by Rehmann et al.'s [27] note that the creation of behavior over time graphs can be supported by tracing a loop in a causal loop diagram.

RQ3: Affordances of Different Approaches to Assessing Systems Thinking

Two main themes were identified that represent tensions between different types of affordances seen across assessments. First was the tension between (a) having less structure so there are fewer constraints on the students' visualization and (b) the potential for ambiguity when evaluating a visualization. Another was the tension between (a) the time it takes to administer an assessment and (b) how much an evaluator can learn about a student's or a group of students' understanding of a problem. These two tensions overlap. For example, assessments that had more structure generally required students to learn the specific format and rules, which likely means more time in administering the assessment.

Assessment Structure and Interpretation. The tension between assessment structure and interpretation of responses was visible across several assessments. For example, Lavi et al.'s [23] assessment had the most structured visualization with the use of OPM, that is ISO 19450, and therefore interpretation of the conceptual models were unambiguous given that there is a textual equivalent of the graphical representations used in the conceptual models [36].

On the opposite end of the spectrum for assessment structure was Vanasupa, Rogers, and Chen's [28] assessment that used rich pictures, as well as the use of rich pictures in Rehmann et al.'s [27] assessment. As one of the least structured visualizations, rich pictures afford students the ability to create a visualization that works for them [29] and thus inherently introduces ambiguity in interpretation.

The structure of the assessment was affected by the scoring technique in addition to the visualization used. For example, Hu and Shealy's [25] assessment inherently assumed a hierarchically structured concept map when the traditional scoring method [22] was used. If those assumptions are communicated to students it acts as a constraint on their creation of a visualization and if such assumptions are not made clear to students it can lead to misinterpretations of non-hierarchically organized concept maps.

Assessment Administration Time and Details Provided. Lavi et al.'s [23] assessment had one of longest administration times with teams submitting their first version of their concept models mid-semester and a second version at the end of a semester. With more time, came more levels of detail as the first conceptual model had to have at least three levels of detail and the second model had to have four or five levels of detail (see "complexity levels" in "Identifying Individual Elements" above). Rehmann et al.'s [27] assessment took place over seven weeks, or half of one semester, with instructors providing feedback to students on one part of their projects each week. With more time, came more details about a system. In the case of Rehmann et al.'s assessment this included a variety of visualizations: rich pictures, causal loop diagrams, and behavior over time graphs, which lend themselves to showing elements, connections, feedback loops, and dynamic behavior as described above.

In contrast, Hu and Shealy's [25] assessment allotted ten minutes per each of the two concept mapping tasks and Brandstädter, Harms, and Großschedl's [26] assessment allotted 20 minutes for concept map creation following a 15-minute introduction. In comparison to Rehmann et al.'s [27] assessment, these concept maps only show a subset of what a combination of rich pictures, causal loop diagrams, and behavior over time graphs could convey, and in comparison to Lavi et al.'s [23] assessment there is more opportunity for ambiguity when trying to interpret a participant's understanding.

Discussion

The research revealed six types of visualizations present in our subset of six assessments that (1) examined demonstrated knowledge or skill, (2) included the creation of at least one visual representation, and (3) were open-ended. Most (5 of 6) of these visualizations required participants to identify multiple aspects of a problem or system and relationships between aspects, while one visualization—behavior over time graphs—focused on the time-varying behavior of a single system variable. The visualizations varied in terms of the dimensions of systems thinking they enable examination of and how structured they were.

In addition to being used for assessment, visualizations or graphical representations can also be considered tools or methods that support systems thinking [28], [45]. For instance, in Gray et al.'s [24] assessment the use of fuzzy cognitive mapping and scenario outputs act as the scaffolding for students' identification of leverage points, where the leverage points were described in writing. When looking across systems approaches more broadly there are a variety of visualizations in use, with different methods e.g., causal loop diagrams (CLDs) and rich pictures, associated with different methodologies e.g., system dynamics and soft system methodology, respectively [45].

There were three prominent dimensions of systems thinking evident across the assessments: identifying individual elements or aspects of a problem, identifying connections or relationships between these elements, and accounting for time and change over time. These dimensions are consistent with a number of definitions of systems thinking such as Senge's [15] definition of systems thinking as "a discipline for seeing wholes. It is a framework for seeing interrelationships rather than things, for seeing patterns of change rather than static 'snapshots'" and Jaradat et al.'s definition where systems thinking "focuses on how the constituent parts of a system pertain to the whole system and the way in systems work within larger systems over time" [46, Sec. Introduction].

Based on the six systems thinking assessments analyzed here, there are a number of opportunities to more explicitly evaluate participants' or teams' analysis of a problem, in addition to their identification of various aspects of the problem. For instance, while some assessments, e.g., [24], [25], did value having a higher number of concepts and having a higher ratio of the number of relationships to the number of identified elements, assessments that evaluated understanding elements in depth (e.g. [25] using scoring methods described in [22]) and relationships in depth (e.g. [27]) were limited in terms of describing what qualifies as "depth." In other words what exactly does it means to have "sufficient detail" [22, p. 129] or to "argue convincingly" [27, Sec. Observations].

Another opportunity relates to breadth in terms of the range of issues considered, e.g., environmental, political, social, economic, and stakeholder considerations. The five visualizations that can be considered various forms of concept maps—rich pictures, concept maps, causal loop diagrams, fuzzy cognitive maps, and conceptual models—all provide the opportunity to evaluate breadth when evaluating the elements identified, but only two the assessments explicitly considered breadth [25], [27] and both these assessments were contextualized using topics related to sustainability. Rehmann et al. [27] focused on seven types of issues listed as "engineering, social, ethical, cultural, environmental, business, and political issues," [27, Sec. Introduction]. By having an "engineering" category many types of issues are positioned as "non-engineering" considerations. The definition of *comprehensive* systems thinking that we advance would define all of these types of issues or aspects of a problem as integral to engineering problem-solving. In line with our advancement of *comprehensive* systems thinking, we argue that a broad understanding of the problem is applicable across all types of engineering—not only systems engineering or sustainability engineering—and by not explicitly evaluating breadth some assessments may be implicitly devaluing such considerations. For example, by not evaluating breadth systems thinking assessments may implicitly signal that the scope of a problem is limited to a single artifact or process and does not account the political context in which that artifact will be embedded or the stakeholders that participate in a process. For the same reasons, it is also important for researchers to be explicit about what constitutes breadth. For example, does "breadth" indicate that multiple aspects of a problem were identified or does it explicitly examine whether participant responses attended to often over-look dimensions such as cultural and political factors.

Finally, we identified tensions between numerous affordances of the different assessments, namely the tension between (a) having less structure so there are fewer constraints on the students' visualization and (b) the potential for ambiguity when evaluating a visualization and the tension (a) the time it takes to administer an assessment and (b) how much an evaluator can learn about a student's or a group of students' understanding of a problem.

An example of where unambiguous interpretation was afforded by more structured visualizations was Lavi et al.'s [23] Assessment. This assessment used the most structured visualization—conceptual models created using OPM—and while the assessment could be formative or summative assessment purposes, its affordance of unambiguous interpretation would be particularly useful in gaining a comprehensive understanding of student progress, i.e., summative assessment [6]. Lavi et al.'s Assessment lends itself to summative assessment because when there is less potential for ambiguity in rater interpretation, there is more potential for consistency across raters, which is particularly important in the context of assigning grades and understanding "the extent to which students have achieved the intended learning outcomes of the project, course or program," [6, pp. 165–166]. On the other hand, the freedom of expression afforded by less structured visualization, e.g., rich pictures, combined with the potential for ambiguity in interpretation, suggests that assessments using such visualizations are better suited for early formative assessments rather than summative assessments. Further discussions on the use of concept maps, specifically in formative and summative assessments, can be found in Watson et al. [22].

A couple assessments demonstrated how more information about an individual's or group's learning was afforded at the cost of longer administration times. Assessments with longer administration times, e.g. [23], [27], included visualizations with more levels of detail [23] or more visualization types [27]. Since the different visualization types included in Rehmann et al.'s [27] assessment enabled the examination of different dimensions of systems thinking, the longer administration times correlated with developing more detailed understandings of systems and thus lend themselves to be used for summative assessments of a course. In contrast the short administration times of some assessments, e.g. [25], [26], combined with relatively short scoring times [22], [26], afford such assessments to be used to provide feedback to students and identify misconceptions they may have about a problem, i.e., to be used as a formative assessment [6].

In addition to these tensions, another factor related to assessment selection or development is how easy it would be for someone to select one the six assessments analyzed here and deploy in on their own. In addition to thinking about rater consistency, administration time, and how much can be learned from the assessment, considerations include the availability of a rubric or measurement tool for scoring and how much scoring relies on internal judgement. For example, Lavi et al.'s assessment's source paper [23] provided descriptions of each attribute but did not provide detailed scoring guidelines, while Brandstädter, Harms, and Großschedl's assessment's source paper [26] included the scoring protocol used. Watson et al. [22] discuss how many judges are needed based the interrater reliability and subjectivity involved in the three concept scoring methods used by Hu and Shealy [25]. Similar to discussions of formative and summative assessments above, if there is more ambiguity in visualization interpretation there may be more internal judgement required during scoring, which in turn can make assessment adoption more difficult.

Related to both what visualizations are used in behavior-based systems thinking assessments (RQ1) and what the affordances are of different approaches to assessing systems thinking (RQ3)— particularly with respect to formative versus summative assessment and the consideration of problem scope—is the extent to which various visualizations support early versus more detailed problem analysis. Rehmann et al.'s [27] assessment scaffolded rich pictures, followed by causal loop diagrams and then behavior over time graphs, which aligned with Vanasupa, Rogers, and Chen's statement that completing a rich picture "enable[s] the next step of constructing an accurate causal loop diagram" [28, p. 2].

Limitations. One limitation of our study is that it may not include all systems thinking assessments in engineering that meet our inclusion criteria of (1) examined demonstrated knowledge or skill, (2) included the creation of at least one visual representation where the type of visualization is not specific to the problem, and (3) were open-ended, because of our decision to select a subset of assessments from those we found in our systematic literature review and the strategic choices embedded in that review [12]. Another limitation is that our findings focus on the visualization aspects of the six assessments, though the visualizations were not necessarily standalone evidence for evaluating demonstrated systems thinking knowledge or skill. Verbal or written explanations of a participant's or team's reasoning may be used in combination with visualizations, but we did not analyze those here because of the variation in the descriptions of what such verbal or written out explanations entailed. For example, Rehmann et al.'s assessment indicates that in order for a team's casual loop diagrams to receive a score of three, the team must "give plausible arguments for the relationships" [27, Sec. Observations]. Finally, we limited out scope to attending to the various forms of visualizations, the dimensions of systems thinking included in the assessments, and the affordances of different systems thinking assessment approaches. Thus, we did not address other aspects that would impact the selection of an assessment such as embedded ontological assumptions, learning goals, and demonstrated reliability.

Implications. Our findings have several implications for systems thinking assessment development. First, the findings demonstrated an opportunity for such assessments to make explicit the value of considering a breadth of issues, for example environmental, economic, and political considerations. In particular, there are opportunities for explicitly valuing the identification and consideration of stakeholders, which were only mentioned in Hu and Shealy's [25] assessment when Watson's [22] holistic and categorical scoring methods were used. A second opportunity exists around establishing clearer guidelines for examining the depth with which individual elements and relationships are understood. Even when assessments examined the depth [23], [25], [27] with which individual elements or relationships were considered, the descriptions of how depth was examined were often vague. Finally, our findings suggest that future development of systems thinking assessments that examine demonstrated knowledge or skill and include the creation of open-ended visualizations should consider: the use case (e.g., formative versus summative assessment), context (e.g., how much time is available to administer the assessment), and ease of adoption (e.g., how easy it for someone to learn how to administer and score the assessment) when deciding if and what visualizations to include in an assessment.

Conclusions

This study analyzed six systems thinking assessments relevant to an engineering context that (1) examined demonstrated knowledge or skill, (2) included the creation of at least one visual representation, and (3) were open-ended. The findings include descriptions of the various visualizations used, the dimensions of systems thinking including in the assessments, and the tensions among several affordances tied to assessment administration. These findings have several implications for future systems thinking assessment development, including the opportunity to be more explicit in valuing considerations of a breadth of issues, e.g., social and temporal, and stakeholders. In addition, the findings highlight several considerations for creating assessments that include visualizations, which also translate to assessment selection considerations.

References

[1] M. Frank and S. Kordova, "Four Layers Approach for Developing a Tool for Assessing Systems Thinking," in *Proceedings of the 59th Annual Meeting of the International Society for the Systems Sciences*, 2015.

[2] ABET, "Criteria for Accrediting Engineering Programs. Effective for Reviews during the 2020-2021 Accreditation Cycle," Baltimore, MD, 2019.

[3] National Academy of Engineering, *The Engineer of 2020*. Washington, D.C.: National Academies Press, 2004.

[4] E. Mosyjowski, S. R. Daly, and L. R. Lattuca, “Defining and Assessing Systems Thinking in Diverse Engineering Populations,” in *ASEE Annual Conference & Exposition*, 2019.

[5] E. Mosyjowski, J. Espinoza, L. R. Lattuca, and S. R. Daly, “Student and Practitioner Approaches to Systems Thinking : Integrating Technical and Contextual Considerations,” in *ASEE Annual Conference & Exposition*, 2020.

[6] E. F. Crawley, J. Malmqvist, S. Östlund, D. R. Brodeur, and K. Edström, “Student Learning Assessment,” in *Rethinking Engineering Education*, Cham: Springer, Cham, 2014, pp. 165–180.

[7] D. Wiliam, “What is assessment for learning?,” *Stud. Educ. Eval.*, vol. 37, no. 1, pp. 3–14, 2011.

[8] G. Wiggins and J. McTighe, “What is Backward Design?,” in *Understanding by Design*, 2nd ed., Alexandria, VA: Association for Supervision and Curriculum Development, 2005.

[9] K. M. Castelle and R. M. Jaradat, “Development of an Instrument to Assess Capacity for Systems Thinking,” *Procedia Comput. Sci.*, vol. 95, pp. 80–86, 2016.

[10] M. Frank, “Assessing the interest for systems engineering positions and other engineering positions’ required capacity for engineering systems thinking (CEST),” *Syst. Eng.*, vol. 13, no. 2, pp. 161–174, 2010.

[11] B. M. Olds, B. M. Moskal, and R. L. Miller, “Assessment in engineering education: Evolution, approaches and future collaborations,” *J. Eng. Educ.*, vol. 94, no. 1, pp. 13–25, 2005.

[12] K. E. Dugan, E. A. Mosyjowski, S. R. Daly, and L. R. Lattuca, “Systems Thinking Assessments in Engineering: A Systematic Literature Review.” in revision, 2021.

[13] A. T. Bahill and B. Gissing, “Re-evaluating systems engineering concepts using systems thinking,” in *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, 1998, vol. 28, no. 4, pp. 516–527.

[14] M. Frank and D. Elata, “Developing the Capacity for Engineering Systems Thinking (CEST) of freshman engineering students,” *Syst. Eng.*, vol. 8, no. 2, pp. 187–195, 2005.

[15] P. M. Senge, *The fifth discipline: The art and practice of the learning organization*. New York: Doubleday, 1990.

[16] 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, 2018.

[17] K. Hogan and K. C. Weather, “Psychological and ecological perspectives on the development of systems thinking.,” in *Understanding urban ecosystems*, New York, NY: Springer, 2003, pp. 233–260.

[18] W. M. Trochim, D. A. Cabrera, B. Milstein, R. S. Gallagher, and S. J. Leischow, “Practical challenges of systems thinking and modeling in public health,” *Am. J. Public Health*, vol. 96, no. 3, pp. 538–546, 2006.

[19] D. Kilgore, C. J. Atman, K. Yasuhara, T. J. Barker, and A. Morozov, “Considering Context: A Study of First-Year Engineering Students,” *J. Eng. Educ.*, vol. 96, no. 4, pp. 321–334, 2007.

[20] B. Palmer, A. F. Mckenna, B. J. Harper, P. Terenzini, and D. Merson, “Design in Context: Where do the Engineers of 2020 Learn this Skill?,” in *ASEE Annual Conference &*

Exposition, 2011.
[21] H. K. Ro, D. Merson, L. R. Lattuca, and P. T. Terenzini, “Validity of the contextual competence scale for engineering students,” *J. Eng. Educ.*, vol. 104, no. 1, pp. 35–54, 2015.
[22] M. K. Watson, J. Pelkey, C. R. Noyes, and M. O. Rodgers, “Assessing Conceptual Knowledge Using Three Concept Map Scoring Methods,” *J. Eng. Educ.*, vol. 105, no. 1, pp. 118–146, 2016.
[23] R. Lavi, Y. J. Dori, N. Wengrowicz, and D. Dori, “Model-Based Systems Thinking: Assessing Engineering Student Teams,” *IEEE Trans. Educ.*, vol. 63, no. 1, pp. 39–47, 2020.
[24] S. Gray *et al.*, “Assessing (social-ecological) systems thinking by evaluating cognitive maps,” *Sustain.*, vol. 11, no. 20, pp. 1–11, 2019.
[25] M. Hu and T. Shealy, “Methods for measuring systems thinking: Differences between student self-assessment, concept map scores, and cortical activation during tasks about sustainability,” in *ASEE Annual Conference and Exposition*, 2018.
[26] K. Brandstädter, U. Harms, and J. Großschedl, “Assessing System Thinking Through Different Concept-Mapping Practices,” *Int. J. Sci. Educ.*, vol. 34, no. 14, pp. 2147–2170, 2012.
[27] C. R. Rehmann, D. T. Rover, M. Laingen, S. K. Mickelson, and T. J. Brumm, “Introducing systems thinking to the engineer of 2020,” in *ASEE Annual Conference and Exposition*, 2011.
[28] L. Vanasupa, E. Rogers, and K. Chen, “Work in progress: How do we teach and measure systems thinking?,” in *ASEE/IEEE Frontiers in Education Conference*, 2008, pp. 1–2.
[29] P. Checkland, “Soft systems methodology: a thirty year retrospective,” *Syst. Res. Behav. Sci.*, vol. 17, no. S1, pp. S11–S58, 2000.
[30] S. Burge, “The Systems Thinking Tool Box.” Burge, Hughes, Walsh, pp. 1–5, 2015.
[31] J. Costa, J. C. Diehl, and D. Snelders, *A framework for a systems design approach to complex societal problems*, vol. 5. 2019.
[32] Y. Yin, J. Vanides, M. A. Ruiz-Primo, C. C. Ayala, and R. J. Shavelson, “Comparison of two concept-mapping techniques: Implications for scoring, interpretation, and use,” *J. Res. Sci. Teach.*, vol. 42, no. 2, pp. 166–184, 2005.
[33] B. K. Bala, F. M. Arshad, and K. M. Noh, *System Dynamics Modelling and Simulation*. Singapore: Springer, 2017.
[34] D. Dori *et al.*, “System Definition, System Worldviews, and Systemness Characteristics,” *IEEE Syst. J.*, vol. 14, no. 2, pp. 1538–1548, 2020.
[35] D. Dori, *Model-based systems engineering with OPM and SysML*. New York, NY, USA: Springer, 2016.
[36] International Organization for Standardization, “ISO/PAS 19450:2015 Automation systems and integration - Object-Process Methodology,” 2015. [Online]. Available: https://www.iso.org/standard/62274.html.
[37] R. D. Knight, *Physics for scientists and engineers: a strategic approach*, Third. Pearson, 2013.
[38] J. Funke, “Solving complex problems: Exploration and control of complex systems.,” in *Complex problem solving: Principles and mechanisms*, R. J. Sternberg and P. A. Frensch, Eds. Hillsdale, NJ: Lawrence Erlbaum Associates, 1991, pp. 185–122.
[39] D. H. Jonassen, “Toward a design theory of problem solving,” *Educ. Technol. Res. Dev.*,

vol. 48, no. 4, pp. 63–85, 2000.
[40] J. Segalàs, D. Ferrer-Balas, and K. F. Mulder, "Conceptual maps: measuring learning processes of engineering students concerning sustainable development," *Eur. J. Eng. Educ.*, vol. 33, no. 3, pp. 297–306, Jun. 2008.
[41] J. Segalàs, D. Ferrer-Balas, and K. F. Mulder, "What do engineering students learn in sustainability courses? The effect of the pedagogical approach," *J. Clean. Prod.*, vol. 18, no. 3, pp. 275–284, Feb. 2010.
[42] J. R. McClure, B. Sonak, and H. K. Suen, "Concept map assessment of classroom learning: Reliability, validity, and logistical practicality.," *J. Res. Sci. Teach.*, vol. 36, no. 4, pp. 475–492, 1999.
[43] M. Besterfield-Sacre, J. Gerchak, M. R. Lyons, L. J. Shuman, and H. Wolfe, "Scoring Concept Maps: An Integrated Rubric for Assessing Engineering Education," *J. Eng. Educ.*, vol. 93, no. 2, pp. 105–115, Apr. 2004.
[44] J. D. Novak and A. J. Cañas, "The Theory Underlying Concept Maps and How to Construct Them," 2006. [Online]. Available: http://cmap.ihmc.us/publications/researchpapers/theorycmaps/TheoryUnderlyingConceptMaps.bck-11-01-06.htm.
[45] M. C. Jackson, *Critical Systems Thinking and the Management of Complexity*. Chichester: Wiley, 2019.
[46] R. M. Jaradat *et al.*, "Measuring individuals' systems thinking skills through the development of an immersive virtual reality complex system scenarios," in *ASEE 126th Annual Conference and Exposition, Conference Proceedings*, 2019.