Dugan Kelley E. (Orcid ID: 0000-0002-1859-2230)

Systems Thinking Assessments in Engineering: A Systematic Literature Review

Abstract: Engineers, facing increasingly complex problems, need to understand the technical and contextual aspects of their work to develop effective solutions. Assessments of comprehensive systems thinking skills are needed to support the development of these skills and to inform professional placement. Thus, our study investigated current systems thinking assessments in engineering by systematically reviewing existing assessments. We analyzed which systems thinking skills were emphasized, how they were evaluated, how data were collected, and in what content areas assessments were based. The results revealed a range of assessments, in terms of type, format, and content area, but a lack of assessments that equally prioritized accounting for technical and contextual considerations. This overview of assessments can be used by employers and educators to select assessments appropriate for their contexts and goals. Overall, this study demonstrates a need for comprehensive systems thinking assessments that evaluate performance.

Keywords: Assessment, engineering, systematic review, systems engineering (SE), systems thinking.

1. INTRODUCTION

In an increasingly complex world, where sociotechnical systems have been recognized as the "environment of our lives" (Strijbos, 2003), we face a number of grand challenges that will have impacts on our society globally (Mote, Dowling, & Zhou, 2016). The National Academy of Engineering's 14 Grand Challenges for Engineering (National Academy of Engineering, 2020) necessitate the need for expertise and input across disciplines and professions because these challenges are "engineering *system* problems" (emphasis original) (Mote et al., 2016). Problem complexity, and similarly system complexity, is influenced by the number of variables involved, how connected the variables are, the types of functional relationships between variables, and how stable these different aspects are with respect to time (Funke, 1991; Rousseau, 2019). Many of the complex problems encountered in professional practice are also ill-structured, where judgments must be made about what is or is not part of the problem and the hierarchy of criteria used to evaluate solutions; consequently, there may be many, one, or no solution(s) to the problem as it is constituted (Jonassen, 2000). Problems that sit at this intersection of being both complex and ill-structured include wicked problems (Rittel & Webber, 1973) and some design problems (Jonassen, 2000).

Interventions that attempt to address or resolve these complex problems risk being ineffective or even harmful if relevant technical and contextual aspects of a problem are not considered or are ignored (Grotzer, 2012). Employers, policy makers, and scholars alike recognize the need for engineers who can integrate connections between technical aspects of their work, as well as, the larger context in which their work is situated and call for education and training to prepare engineers to better account for this complexity (Hayden, Rizzo, Dewoolkar, Oka, & Neumann, 2010; National Academy of Engineering, 2004; Rebovich, 2006). The call for educational interventions must go hand-in-hand with the development of ways to assess the success of such interventions. Assessments that can evaluate the extent to which skills are taught, valued, and developed are thus a necessary component to address this call. Assessments provide evidence and understanding of skill development (E. F. Crawley, Malmqvist, Östlund, Brodeur, & Edström, 2014; Wiliam, 2011) and should inform the development of training and curriculum materials (Wiggins & McTighe, 2005). Specific to systems thinking, some researchers have also argued for the importance of assessments in effectively evaluating professional competence and fit (Castelle & Jaradat, 2016; Frank, 2010).

Some fields refer to the skill of integrating connections between technical and contextual aspects of their work into decision making as systems thinking (ST) (Hogan & Weather, 2003; Trochim, Cabrera, Milstein, Gallagher, & Leischow, 2006), and we conceptualize systems thinking as an essential skillset in addressing complex problems. In engineering, ST research has often emphasized recognition of the constituent elements of an immediate problem (e.g., Bahill & Gissing, 1998; Frank & Elata, 2005; Senge, 1990), but frequently underplays the range of contextual factors that interact with the problem. Several recent studies have recognized the importance of integrating context in engineering solutions, but have not explicitly tied contextual competence to systems thinking (Kilgore, Atman, Yasuhara, Barker, & Morozov, 2007; Palmer, Mckenna, Harper, Terenzini, & Merson, 2011; Ro, Merson, Lattuca, & Terenzini, 2015). Systems thinking is related to other competencies, abilities, and frameworks, including interdisciplinary competence (Lattuca, Knight, Ro, & Novoselich, 2017), socio-technical thinking (Mazzurco & Daniel, 2020), and the holistic contextual framework for design (Aranda-Jan, Jagtap, & Moultrie, 2016). Interdisciplinary competence is a multidimensional concept that includes students' ability to synthesis within-discipline information, beliefs regarding the nature of engineering problems, and valuation of interdisciplinary work (Lattuca et al., 2017). Socio-technical thinking is "the ability to integrate social and technical dimensions in solving a design problem" (Mazzurco & Daniel, 2020). The holistic contextual framework aims to aid designers in understanding contextual factors when working in low-resource settings (Aranda-Jan et al., 2016). While systems thinking often necessitates drawing on various aspects of these competencies and abilities (e.g., including contextual factors and working across disciplines while problem solving), and can benefit from existing assessments and frameworks of related competencies, systems thinking differentiates itself with its attention to and concern with complexity, particularly the interconnectedness of various aspects of a problem.

How people think about systems varies by their ontological and epistemological perspectives. One distinction in the way systems thinkers understand systems is the "hard" system stance, where the world is made up of determinate systems, versus the "soft" system stance, where systems thinkers perceive the world as complex and although they cannot know what this complexity is, i.e., it is indeterminate, they can think about it as a system (Checkland, 1983, 2000). Rather than engaging in the debate on the nature of systems or reality more broadly, for the purposes of this paper we hold a pragmatic position, where systems thinking is helpful in solving complex problems because it foregrounds an awareness of relationships and tradeoffs. We do not take into consideration if those systems are framed as existing or require interpretation as a system.

Recognizing the importance of both technical and contextual factors in engineering work, we advance a definition of *comprehensive* systems thinking as a holistic approach to problem solving in which connections and interactions between constituent parts and the immediate work, stakeholder needs, broader contextual aspects (e.g., social and environmental), and potential impacts over time are identified and integrated into decision making (Authors, 2019; Authors 2020). This definition is informed by literature that describes elements of systems thinking, such as relationships between components, stakeholder needs, social and environmental contexts, and temporal dimensions (Bahill & Gissing, 1998; Frank, 2000; Frank & Elata, 2005;

Grohs, 2015; Hogan & Weather, 2003; Senge, 1990). From the perspective of comprehensive systems thinking we acknowledge the challenges related to decomposing complexity and the tradeoffs that arise in this process. However, we believe it advantageous to make explicit the aspects of a problem that engineers attend too, particularly as contextual aspects are often overlooked.

In this study, we analyzed existing systems thinking assessments in engineering to provide an overview of available assessments. We focused on what dimensions of systems thinking were evaluated, how they were evaluated, how data were collected, and the content area within which the assessments were based. The outcomes of this work can guide assessment selection and inform future assessments.

2. Background Literature

Across disciplines there are many different definitions of systems thinking, as well as numerous lists of systems thinking skills (Booth Sweeney & Sterman, 2000; Kordova & Frank, 2018; Rehmann, Rover, Laingen, Mickelson, & Brumm, 2011; Tomko, Nelson, Linsey, Bohm, & Nagel, 2017). One commonly cited definition describes 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.’ It is a set of general principles… It is also a set of specific tools and techniques” (Senge, 1990, p. 68). Such general foundational definitions, combined with the development of systems thinking in several disciplines, had led to a proliferation of varying definitions of systems thinking. It is not within the scope of this paper to review the history and development of systems thinking frameworks and definitions. However, in the context of reviewing systems thinking assessments in engineering, it is important to recognize that there are several aspects of systems thinking that frequently appear in engineering. These aspects include the ideas of holism, “focused on the whole, interested more the big picture” (Castelle & Jaradat, 2016), identifying and analyzing relationships between components, and recognizing changes over time. These aspects are frequently focused around the constituent parts of an engineering problem (e.g., Bahill & Gissing, 1998; Frank & Elata, 2005; Senge, 1990), rather than recognizing and incorporating the broader context in which the constituent parts are embedded. We use c*omprehensive* systems thinking to push for consideration of various stakeholders and broader contextual aspects in addition to the constituent elements of the immediate problem. This work recognizes that how a scholar defines systems thinking guides their operationalization of these skills and can vary from one scholar to another.

The field of engineering makes a distinction between systems thinking as a skillset and systems engineering as a systems development approach (Monat & Gannon, 2018). Though our focus is on systems thinking as a skillset, conversations within systems engineering echo common challenges of addressing problems concerning sociotechnical systems. Similar to patterns in engineering systems thinking research, systems engineering has struggled with trying to separate a system from its context, for example its social context (Kroes, Franssen, van de Poel, & Ottens, 2006). One proposed solution in systems engineering is the expansion of the system's boundaries to include elements such as human agents when working with socio-technical systems (Kroes et al., 2006). Other calls for a broader conceptualization of systems maintain the distinction between the "internal workings of a system" and "external factors" and characterize complexity as ranging from "internal complexity" to "external complexity," noting that as the world's complexity increases engineers will need to deal with problems that have both internal and external complexity more and more frequently (Rousseau, 2019).

Recognition of this complexity is reflected in our definition of *comprehensive* systems thinking that we used to guide our review of ST assessments. This review's focus on assessments stems from best practices in curriculum development that call for educators to "operationalize our goals or standards in terms of assessment evidence as we *begin* to plan a unit or course," (emphasis original) (Wiggins & McTighe, 2005, p. 8). It is this practice of "backward design" that encourages educators to "think like an assessor" before starting to develop lessons (Wiggins & McTighe, 2005, p. 12). Therefore, the availability of systems thinking assessments, the ease of their implementation, and the aspects of systems thinking they operationalize can all impact what knowledge and skills are covered in a course. In this way, assessments signal what content is most valued in a particular context and are essential in supporting and measuring the development of key skills (E. F. Crawley et al., 2014; Wiliam, 2011), including systems thinking. Thus, the focus of our systematic literature review (SLR) was on systems thinking assessments rather than systems thinking definitions, because of the practical and more immediate implications of understanding which aspects of systems thinking have been and are being assessed in engineering.

3. Method

This systematic literature review (SLR) characterized the current state of systems thinking assessments in engineering. Our systematic approach—sometimes called systematic mapping (Gough, Sandy, & James, 2012)—was informed by best practices for SLRs, such as clearly defining the scope, strictly adhering to exclusion and inclusion criteria, detailing the screening process, and summarizing important details of included literature (Borrego, Foster, & Froyd, 2015; Gough et al., 2012).

A. Purpose

The purpose of this SLR was to map the landscape of existing systems thinking assessments across engineering disciplines to identify similarities and differences, including their approaches, structure, substance, focus, and how systems thinking is represented, either implicitly or explicitly, within them. We did not characterize assessment development processes nor assess outcomes associated with each assessment.

The SLR was guided by the following research questions:

1) What assessments of ST exist in engineering?
2) What are the different approaches to assessing ST?
3) What do these existing systems thinking assessments represent about how systems thinking is defined?

B. Searching the Literature

The SLR began with six papers on systems thinking assessments (Castelle & Jaradat, 2016; Grohs, Kirk, Soledad, & Knight, 2018; Jaradat, 2014; Kordova & Frank, 2018; Rehmann et al., 2011; Vanasupa, Rogers, & Chen, 2008), all of which studied postsecondary engineering students, that our research team had previously used in our systems thinking research. These papers informed our SLR search, which consisted of two main decisions: database selection and search string development of keyword, timing, and field selection. Guided by a librarian, we considered four databases (Web of Science, SCOPUS, ERIC, and Engineering Village) with the intent of determining how to best focus our search while identifying relevant papers within engineering across a range of educational and professional engineering contexts. We chose the databases Web of Science and SCOPUS under the assumption that the search results from Web of Science were likely to cover both the results in ERIC and Engineering Village, while having a smaller, and possibly different, breadth of results than SCOPUS. The six assessments we previously identified informed the keywords used in the search string, namely, terms related to the word "assessment." No restriction was imposed on publication date and ultimately the oldest assessment included was published in 2000.

The initial search on 01/07/2020 used the "Article title, Abstract, Keywords" fields to find publications related to systems thinking assessment and returned 1,826 documents. To better identify papers in which systems thinking assessment was foregrounded, we modified the search on the same day to restrict results to article title only, reducing the number of results in SCOPUS from 1,826 to 88 articles. Eighty-five of these SCOPUS results were new additions to the search and Web of Science results contributed an additional three novel papers. Both searches were also defined to include papers that mentioned "engineering" or "engineers" in any of the fields, in order to find the broadest collection of assessments with connections to engineering. Thus, this search string was:

> Article title: "systems thinking" AND (assess* OR measur* OR eval* OR instru* OR metr* OR analy*) AND All fields: engineering OR engineers

A refresh of the above search in the same databases was conducted on 03/15/2020 to check for publications added to the databases since January. Additionally, the keywords were expanded in this search to include analogies to "assessment," including typolog*, inventor*, scale*, test*, and rubric. By refreshing the original search and broadening the search criteria, an additional 14 articles were identified, bringing the total number of unique papers through these searches to 102. A refresh of the broadened search was conducted on 02/27/2021 that identified an additional 21 articles, bringing the total number of unique papers from these searches to 123. While there may be additional publications or other forms of information on the research covered in the database search results, we evaluated studies solely on the information provided in the papers appearing in the database search results. One additional paper (Booth Sweeney & Sterman, 2000) did not appear in any of the database searches, but was added to the review because it was frequently referenced by other papers that met the other inclusion criteria for the SLR. This additional paper combined with the 123 identified in the database searches and the original six known assessments brought the total number of papers for review to 130.

C. Two-Stage Screening Process

The 130 papers were screened (see Figure 1 for a summary of the search and screening processes) to include only those that were available online, available in English, and provided sufficient information on the systems thinking assessment to enable us to address our RQs, specifically by describing an approach to assessing systems thinking in detail. We relied on authors' identification of their work as a means to measure or assess systems thinking regardless of how they characterized systems thinking. Papers that did not include presentation of a systems thinking assessment were excluded (e.g., Plack et al., 2018; York & Orgill, 2020). We also eliminated book excerpts and our own publications, but kept peer-reviewed conference papers/journal articles and dissertations. In order to maintain strict inclusion and exclusion criteria, all articles that showed up in the search and were not screened out were still included even if upon reading the article, it was evident that the primary audience was not engineering students or practitioners.

Following these inclusion and exclusion criteria, the articles were first screened by reviewing their Titles and Abstracts (if available). Those screened using title only were removed only if it was clear from the title that the article was outside the scope of the review. Ninety-five failed to meet the screening criteria and were removed during this first screening; the full texts of the remaining 35 papers were then retrieved and reviewed for the second screening. The second screening resulted in removing an additional five papers. A total of 30 papers, and 27 unique assessments, met the inclusion criteria for this SLR. Several articles discussed the same assessment and a few other articles discussed more than one assessment, leading to the discrepancy between the number of papers and assessments.

D. Thematic Analysis

The final 30 papers were analyzed according to driving questions developed a priori that aligned with the research questions (see Table 1). For some of the analysis categories, the types of results were readily apparent (e.g., "Education Level Targeted" included Professionals, Postsecondary, High School, and Pre-High School). For other categories (e.g. Assessment Type), an inductive analysis approach was used, where papers were grouped based on commonalities and differences in the particular analysis category and then named and described. The justifications used to support these groupings were iteratively refined until common themes were developed across the assessments. During this iterative process, additional analysis categories were added to further detail existing assessments, including geographic location and evaluation criteria.

4. Findings

A. Existing Systems Thinking Assessments

The 27 distinct systems thinking assessments (listed in Table 2 with their sources) were named as follows: a) if the assessment was named in the source article(s), that name was used; b) if a name was not provided, the authors' names were used to identify the assessment; c) if the same authors discussed more than one assessment, we added the labels "A", and "B" alongside the authors' names.

1) *Education Levels and Disciplines Targeted by ST Assessments:* The assessments described a variety of education levels and disciplines for which they were implemented in the papers we reviewed, as shown in Table 3. Education levels and disciplines are those of the majority of participants described in these assessments' source paper(s). The disciplines that were described in the papers we reviewed do not necessarily reflect the full range in which the assessments are potentially relevant. Pre-high and high school participant disciplines were identified by the class or project context in which students were assessed, postsecondary participant disciplines were identified by major, department, or degree program, and professional participant disciplines were identified by job position. Sixteen of the 27 assessments targeted professionals and/or postsecondary students in engineering. Seven targeted postsecondary students not in engineering, five targeted high school students, three targeted pre-high school students, and one targeted an unspecified education level of organic chemistry students.

B. Approaches to Assessing Systems Thinking

We characterized approaches to assessing systems thinking according to: 1) Type; 2) Format; and 3) Content Area. Assessment Type identified what was evaluated and/or how evaluations were made. Assessment Format described how assessments were structured and are discussed in relation to Assessment Type. Assessment Content Area referred to the topic around which an assessment was based. For example, Brandstädter, Harms,

& Großschedl's Assessments (Brandstädter, Harms, & Großschedl, 2012) A and B were both based around the topic of blue mussels. Table 4 shows Assessment Format and Content Area by Type.

1) *Systems thinking assessment types:* Across the 27 assessments in this SLR, our inductive analysis resulted in four types of assessments: 1) behavior-based; 2) preference-based; 3) self-reported; and 4) cognitive activation. The majority of assessments were behavior-based or preference-based.

Behavior-based: Nineteen assessments focused on knowledge or skill(s) based on performance on a specific task, such as drawing or answering open-ended, fill-in-the-blank, or multiple-choice questions. For example, the Systems Thinking Assessment Rubric (STAR) had teams create conceptual models of a selected system and these models were then scored based on how fully they communicated an understanding of each of nine attributes, with one such attribute being the complexity levels of the model (Lavi, Dori, & Dori, 2020; Lavi, Dori, Wengrowicz, & Dori, 2020).

Preference-based: Five assessments characterized values, interest, attitude, and/or aptitude. For example, Jaradat & Castelle's Assessment included the creation of a systems thinking profile, that "established individuals' predisposition to adapt a systemic perspective" based on an their responses to a 39-questions linked to a given scenario (Castelle & Jaradat, 2016, p. 83). Jaradat, Hamilton, Dayarathma et al.'s Assessment used a virtual reality (VR) gaming format to map participant choices in the game to a subset of questionnaire selections in Jaradat & Castelle's Assessment (Jaradat et al., 2019).

Self-reported: One of two self-reported assessments gauged engineering systems thinking self-efficacy (Degen, Muci-Küchler, Bedillion, Huan, & Ellingsen, 2018), while the other asked how well an individual perceived they had learned something (Hadgraft, Carew, Therese, & Blundell, 2008). The unifying characteristic of this category was that these assessments relied on the individual, rather than an outsider, to make the evaluation. For example, Hadgraft, Carew, Therese, & Blundell's Assessment provided students with a list of 14 systems thinking skills and asked them to rate how well they had learned 14 systems thinking skills (Hadgraft et al., 2008).

Cognitive Activation: One assessment, Hu & Shealy's Assessment B, used neuroimaging to monitor brain activity. In this assessment, participants wore a functional near-infrared spectroscopy (fNIRS) cap while concept mapping (Hu & Shealy, 2018).

Assessment Type Not Discernable: Zoller & Scholz Example 2 Assessment, could not be categorized. While Zoller and Scholz (2004) provided the questionnaire, we could not determine if the goal of the assessment was to examine knowledge or skill(s), as in behavior-based assessments, or to characterize values, interests, etc., as in preference-based assessments.

Another assessment, Engineering Systems Thinking Survey, was categorized both as a self-reported and a behavior-based assessment since it had two sections. The first measured self-efficacy with Likert-scale questions, and the second measured knowledge and skills with multiple-choice questions (Degen et al., 2018).

2) *Systems thinking assessment formats:* Assessment Format focused on how assessment data were collected, i.e., what the assessment looked like to participants and how participants were evaluated (e.g., multiple-choice questions or an oral exam). The format groups, which were not mutually exclusive, included mapping, scenario, open-ended, oral, fill-in-the-blank, multiple-choice, virtual reality, and fNIRS cap.

Mapping (M): A "mapping format" included participant creation of some visual representation, that may have contained words, but did not consist exclusively of words. There was variety in how much structure was provided and how much the evaluators helped participants in creating the map. For example, STAR was relatively highly structured, as conceptual models were created following Object-Process Methodology (OPM) (Lavi, Dori, & Dori, 2020; Lavi, Dori, Wengrowicz, et al., 2020). In Vanasupa, Rogers, & Chen's Assessment, the use of rich pictures was unstructured, as the creation of the rich picture was left entirely to the students (Vanasupa et al., 2008). In many of the assessments, there was a mix of evaluator and participant involvement in the process of mapping. In Brandstädter, Harms, & Großschedl's Assessment A, various levels of directedness in concept map creation were tested (Brandstädter et al., 2012). We did not consider Keynan, Benzvi Assaraf, & Goldman's Assessment a mapping assessment because the Repertory Grid maps were created by evaluators after collecting data from students (Keynan, Ben-Zvi Assaraf, & Goldman, 2014). See Appendix A for information on the Repertory Grid Technique. Terms from mapping assessments (in Table 4) are defined in Appendix B.

Scenario (S): While many assessments included some context, a "scenario" assessment elaborated on many details within a problem setting, including background information, needs, and/or constraints. For example, Brandstädter, Harms, & Großschedl's Assessment A directed students to create concept maps of the "development, enemies, living, and feeding of eggs, larvae, young and adult blue mussels" (Brandstädter et al., 2012, p. 2151). Another example of a scenario was Taylor, Calvo-Amodio, and Well's Assessment, where they asked students to draw fish-tank systems for the problem: "You recently purchased a fish tank. After two weeks, you notice the water is turning green in color" (Taylor, Calvo-Amodio, & Well, 2020, p. 10). Conversely, Hu & Shealy's Assessment A was not characterized as a scenario assessment, because while students were asked to draw concept maps on topics related to sustainability the paper did not indicate that students were given a particular prompt to respond to (Hu & Shealy, 2018).

Open-ended (E): Open-ended assessments did not ask to students to draw from pre-populated language or responses. Some of the assessments in this category indicated a particular medium in which the response should be delivered (e.g., a concept map or an oral presentation). While all open-ended assessments inherently allowed an unanticipated amount of variation in responses, there was a range of how much scaffolding was provided for responses. For example, in Systems Assessment Test (SysTest), students were directed to "Read the following customer needs statement and then describe the system as best as possible using technique(s) you have learned" (Tomko et al., 2017, p. 182), whereas in Grohs, Kirk, Soledad, & Knight's Assessment, there were multiple prompts designed to operationalize a number of systems thinking constructs, including problem identification and information needs (Grohs et al., 2018). In Brandstädter, Harms, & Großschedl's Assessment A, which included three different variations of concept mapping, only the "nondirected" variation was considered open-ended (Brandstädter et al., 2012). Keynan, Ben-Zvi Assaraf, & Goldman's Assessment was not considered open-ended because participants were provided with 15 elements, which were terms related to the Shezaf ecosystem (see Appendix A for more general element meaning), and asked to examine three elements by specifying how two elements are similar to each other and yet are different from the third element (Keynan et al., 2014).

Oral (O): The "oral format" asked participants to verbally describe their thoughts. For example, in Rehmann, Rover, & Laingen et al.'s Assessment one aspect of evaluation of a scholars program included oral presentations which were scored on the basis of technical content and presentation details (Rehmann et al., 2011).

Fill-in-the-blank (F): In three assessments participants were provided a partial diagram (Hrin, Milenković, Segedinac, & Horvat, 2017; Timofte & Popuș, 2019) or graph (Booth Sweeney & Sterman, 2000) and asked to fill in elements that were blank or not drawn. For example, the assessments by Timofte & Popuș (2019) and Hrin, Milenković, Segedinac, & Horvat (2017), which used systemic assessment questions (see Appendix B), provided a diagram to be filled in. These assessments were categorized as both mapping and fill-in-the-blank.

Multiple-choice (C): Assessments that had a multiple-choice format utilized questions with several, predefined answers. While the multiple-choice format spanned across three Assessment Types, there were key differences between the multiple-choice questions used in preference-based versus behavior-based assessments. All of the preference-based assessments had a multiple-choice format except for Jaradat, Hamilton, Dayarathna, et al.'s Assessment, which was a VR game (Jaradat et al., 2019). The multiple-choice questions in preference-based assessments were used to determine a value judgment either by having participants indicate their agreement with one of two statements (Castelle & Jaradat, 2016; Frank, 2007, 2009, 2010; Frank & Kordova, 2015; Jaradat, 2014) or the extent to which a statement aligned with their personal values on a set scale (Camelia & Ferris, 2018; Camelia, Ferris, & Cropley, 2018; Kordova & Frank, 2018). In the behavior-based assessments, individuals were asked to select from multiple statements to determine actual knowledge or skill. The Climate Change System Thinking Instrument (CCSTI), a behavior-based assessment, was considered part of the scenario and multiple-choice grouping because the multiple-choice questions provided background information and needs (Meilinda, Rustaman, Firman, & Tjasyono, 2018).

Virtual Reality (V): One assessment, by Jaradat, Hamilton, Dayarathma et al., used a virtual reality (VR) gaming scenario to measure "how students react to situations and manage uncertainty" (Jaradat et al., 2019, sec. Abstract). Participants wore a VR headset and used touch controllers to complete retail store tasks in an immersive setting.

fNIRS Cap (N): One assessment, Hu & Shealy's Assessment B, had participants wear a functional near-infrared spectroscopy (fNIRS) cap to measure global efficiency of connectivity while they were concept mapping (Hu & Shealy, 2018).

The format groups were not intended to be mutually exclusive, and 20 of the 27 assessments were assigned two or more groups. Most of the behavior-based assessments were assigned to the mapping group (11 of 19). Two format groups, scenario and multiple-choice, spanned two Assessment Types, while the multiple-choice format group crossed three Assessment Types. The self-reported, cognitive activation, and "not discernable" Assessment Types used multiple-choice, open-ended, and a fNIRS cap format.

3) *Content Area of systems thinking assessments:* Ten of the 27 assessments utilized environmental topics ranging from sustainability to resource and energy use to ecology (Benninghaus, Mühling, Kremer, & Sprenger, 2019b; Brandstädter et al., 2012; Hu & Shealy, 2018; Keynan et al., 2014; Meilinda et al., 2018; Rehmann et al., 2011; Zoller & Scholz, 2004). No other contexts frequently occurred; rather there was a wide variety of content areas, such as information systems (Lavi, Dori, Wengrowicz, et al., 2020), an export management company (Jaradat, 2014), and heating expenses in a village (Grohs et al., 2018).

C. Assessment Insights Regarding ST Definitions

We also analyzed how assessment authors defined ST, either explicitly or implicitly, and which aspects of ST they foregrounded in their assessments. Definitions of ST signal aspects of systems thinking valued in the field. Within education, assessments can be used to promote learning outcomes (E. F. Crawley et al., 2014), determine if learning outcomes are achieved (E. F. Crawley et al., 2014; Wiliam, 2011), and inform course design (Wiggins & McTighe, 2005). Within industry, assessments guide identifying, placing, and developing professionals (Castelle & Jaradat, 2016; Frank, 2010) and creating training programs (Frank & Kordova, 2015).

We included definitions of both "*systems* thinking" and "*system* thinking" because of the prevalent use of the two terms. For example, in the STAR assessment (Lavi, Dori, & Dori, 2020; Lavi, Dori, Wengrowicz, et al., 2020), one of the assessment's source paper's definition of "*systems* thinking" included a reference to a system architecture text (Lavi, Dori, Wengrowicz, et al., 2020, p. 40) that used "*system* thinking" and defined it as "thinking about a… problem explicitly as a system," (E. Crawley, Cameron, & Selva, 2016, p. 8). The source papers for Jaradat & Castelle's Assessment (Castelle & Jaradat, 2016; Jaradat, 2014), Jaradat, Hamilton, Dayarathna et al.'s Assessment (Jaradat et al., 2019), and Keynan, Ben-Zvi Assaraf, & Goldman's Assessment (Keynan et al., 2014) used both "*systems* thinking" and "*system* thinking." Three assessments (Brandstädter et al., 2012; Meilinda et al., 2018; Zoller & Scholz, 2004) only used "*system* thinking" with two (Brandstädter et al., 2012; Meilinda et al., 2018) including references on "*systems* thinking".

We examined how systems thinking was defined, if at all, in each assessment's source paper(s) from two perspectives: how the author's defined systems thinking in the paper(s) and what aspects of ST were emphasized in the assessment's rubric. We divided the assessments into three mutually exclusive groups based on the availability of their evaluation criteria and the depth with which dimensions of ST were assessed. Tables 5–7 summarize the definitions and dimensions of systems thinking across these three groups. Group 1 contained nine assessments that did not provide specific evaluation criteria. Group 2 contained five assessments that, at a high level, covered a wide range of ST skills, while Group 3 consisted of 13 assessments that, at a more detailed level, typically covered fewer ST skills in comparison to Group 2.

1) *Definitions of Systems Thinking:* Several definitions emphasized holism, including taking a big picture view (Frank, 2010), having a holistic understanding (Jaradat, 2014; Lavi, Dori, Wengrowicz, et al., 2020), and seeing the whole (Kordova & Frank, 2018; Vanasupa et al., 2008). Common themes also included focusing on system elements, which generally referred to aspects of a system's identity or what composed the system (e.g. system parts or physical or informational objects) and the relationships between system elements (Benninghaus et al., 2019b; Brandstädter et al., 2012; Gray et al., 2019; Lavi, Dori, Wengrowicz, et al., 2020). Time-related or temporal considerations were also seen across a few ST definitions (Booth Sweeney & Sterman, 2000; Brandstädter et al., 2012; Jaradat et al., 2019). Some definitions presented systems thinking at a more general level, including as a higher-order cognitive skill (Zoller & Scholz, 2004), a meta-cognitive strategy (Grohs et al., 2018), or as the characteristics of an individual demonstrated while solving complex problems (Jaradat, 2014). The definitions in Tables 5–7 were directly pulled from the source paper(s) unless otherwise stated.

2) *Criteria Foregrounded in ST Assessments:* Group 1 assessments are not discussed here because their source papers lacked sufficient detail regarding evaluation criteria. This made an analysis of the dimensions of ST foregrounded unproductive. The five ST assessments in Group 2, which were preference-based or self-reported, covered a wide range of systems thinking skills. Two included understanding a part in relation to the whole, levels of complexity, the interdisciplinary nature of ST, continuous improvement, and managerial considerations (Camelia & Ferris, 2018; Camelia et al., 2018; Frank, 2007, 2009, 2010; Frank & Kordova, 2015). Four of the five assessments included questions regarding interconnections or interactions between different parts of a system (Camelia & Ferris, 2018; Camelia et al., 2018; Castelle & Jaradat, 2016; Frank, 2007,

2009, 2010; Frank & Kordova, 2015; Jaradat, 2014; Jaradat et al., 2019), e.g. “interconnections and mutual influences between the main tasks and the peripheral task” (Camelia & Ferris, 2018, p. 577; Camelia et al., 2018, p. 119). Within Group 2, another common theme (4 of 5) was recognizing the importance of factors that push beyond a traditionally narrow technical focus within engineering, such as managerial considerations and customer needs (Frank, 2010), so-called “engineering and non-engineering consequences” (Camelia & Ferris, 2018, p. 577; Camelia et al., 2018, p. 119), “non-technical issues” (Jaradat, 2014, p. 264), and political, social, and environmental responsibilities (Camelia & Ferris, 2018; Camelia et al., 2018; Hadgraft et al., 2008).

All 13 of the Group 3 assessments were of the behavior-based Assessment Type. In contrast to Group 2 assessments, Group 3 assessments rarely emphasized aspects of systems thinking beyond element and relationship identification and analysis, the hierarchy of a system, or changes over time. Across these assessments, we identified eight dimensions of systems thinking. Table 7 summarizes the evaluation criteria, dimensions of systems thinking assessed, and presents the dimensions of ST categories for each of these assessments.

Elements (E): Ten of 13 Group 3 assessments included identifying individual aspects of the problem, which we defined as attending to Elements. Elements included objects and processes (Lavi, Dori, & Dori, 2020; Lavi, Dori, Wengrowicz, et al., 2020), components (Meilinda et al., 2018), the system’s structure (Brandstädter et al., 2012), key variables (Rehmann et al., 2011), terms (Keynan et al., 2014), or information cards (Benninghaus et al., 2019b). One example of an assessment that emphasized elements was Hrin, Milenković, Segedinac, & Horvat's Assessment, which rated participants in part based on identifying concepts to fill in the Systemic Synthesis Questions [SSynQs] (Hrin et al., 2017). Their inclusion of criteria about the identification of individual concepts in their scoring rubric showed that the element identification was foundational to their definition of systems thinking. In contrast, Mystery Maps provided participants with information cards, a proxy for the problem’s main elements, thus it did not emphasize element identification (Benninghaus et al., 2019b).

Relationships (R): Twelve of the 13 assessments included identifying and/or analyzing relationships between elements, such as objects (Lavi, Dori, & Dori, 2020; Lavi, Dori, Wengrowicz, et al., 2020) or information cards (Benninghaus et al., 2019b). One example was STAR, which rated participants partially on their understanding of structural and procedural relations (Lavi, Dori, & Dori, 2020; Lavi, Dori, Wengrowicz, et al., 2020). The inclusion of different types of links in the STAR assessment rubric showed that a consideration of relationships was a key aspect of the authors’ definition of systems thinking. Mystery Maps also emphasized relationships, where all the evaluation variants rate participants based on whether they create the appropriate direct or indirect connections (Benninghaus et al., 2019b).

Feedback (F): Three assessments explicitly named feedback processes. As feedback is a type of relationship, this dimension was a subset of the assessments in the Relationships category. Booth Sweeney & Sterman's manufacturing case was an example of incorporating feedback into assessment as it included one negative feedback loop and participant responses were evaluated for meeting certain feedback constraints (Booth Sweeney & Sterman, 2000). In CCSTI, one system thinking indicator was about the ability to identify feedback processes (Meilinda et al., 2018). In Hu and Shealy's Assessment, when Watson, Pelkey, Noyes, and Rodgers' holistic scoring method (2016) was used, part of the criteria for scoring the organization the concept map was the presence of feedback loops (Hu & Shealy, 2018)

Levels (L): Two assessments valued increasing levels of refinement in responses (Lavi, Dori, & Dori, 2020; Lavi, Dori, Wengrowicz, et al., 2020; Meilinda et al., 2018) where levels refers to the "levels of description that can be used to characterize a system with lots of interacting parts," (Wilensky & Resnick, 1999, p. 3). The STAR assessment included "complexity levels" as a systems thinking attribute; the authors defined these as the "number of levels of detail," "refinement of main functions into sub-levels," (Lavi, Dori, Wengrowicz, et al., 2020, p. 42) and the "number of levels in the system's functional hierarchy," (Lavi, Dori, & Dori, 2020, p. 4). The other example was CCSTI, where one system thinking indicator was the ability to identify relationships within "one level of organization," while another indicator was the ability to analyze relations across two different levels (Meilinda et al., 2018, p. 3). These assessment's rubric and system thinking indicators, respectively, revealed that recognition of the numerous levels at which a system can be described were key to their understanding of systems thinking.

Time (T): Assessments grouped in the time category valued temporal considerations, which includes accounting for time as a contextual factor as well as accounting for dynamic behavior. For example, Grohs, Kirk, Soledad, & Knight's Assessment's rubric included identifying both short-term and long-term goals, considering short-term and long-term consequences and challenges, and considering and valuing time as a contextual aspect of problem-solving (Grohs et al., 2018). Their decision to include time as a contextual aspect, along with short-term and long-term considerations, showed that reflection, prediction, and attention to how a problem can evolve or change over time are essential to their definition of systems thinking. Another example of an assessment that emphasized time was Keynan, Benzvi Assaraf, & Goldman's Assessment, which was based on Ben-Zvi Assaraf and Orion's (2005) Systems Thinking Hierarchy model (Keynan et al., 2014). This model included "thinking temporally" as one of the three characteristics in Level C of the hierarchy (Keynan et al., 2014, p. 92). STAR included "temporary objects and decision nodes," as a systems thinking attribute in its rubric, highlighting that engineering systems change over time (Lavi, Dori, & Dori, 2020, p. 5; Lavi, Dori, Wengrowicz, et al., 2020, p. 42)

Breadth (B): The Breadth category included those assessments that pushed beyond the consideration of only technical factors to consider social, economic, environmental, political, legal, etc. aspects of the problem. In contrast to the Group 2 assessments, where pushing beyond a narrow technical focus was a common theme, only three Group 3 assessments met this criterion. The strongest example was Grohs, Kirk, Soledad, & Knight's Assessment, which included consideration of contextual aspects numerous times throughout the rubric. In this assessment, contextual aspects encompassed the following considerations: economic, political, legal, social, cultural, and time (Grohs et al., 2018). In Hu and Shealy's Assessment A, when Watson, Pelkey, Noyes, and Rodgers' categorical scoring method (2016) was used, a complexity index based on concepts and relationships between concepts in social (including stakeholders), economic, and environmental categories was determined (Hu & Shealy, 2018). It is important to note that the categorical method was designed to be used in assessing concept maps related to sustainability and sustainability topics are the content area for Hu and Shealy's Assessments A and B (Hu & Shealy, 2018; Watson et al., 2016). In Hu and Shealy's Assessment A, the holistic scoring method (Watson et al., 2016, p. 129) was used to evaluate how many different "major", e.g. economic, environmental, social, and "advanced" (values, education, actors and stakeholders), dimensions were included in a concept map (Hu & Shealy, 2018). In Rehmann, Rover, & Laingen et al.'s Assessment, the evaluation of rich pictures included whether the rich picture included elements form five of the following seven areas: "engineering, social, ethical, cultural, environmental, business, and political issues" (Rehmann et al., 2011, sec. Introduction).

Stakeholders (S): Only two of the 13 assessments in Group 3, Grohs, Kirk, Soledad, & Knight's Assessment and Hu and Shealy's Assessment A, explicitly valued identifying and engaging with stakeholders. Grohs et al. rated participants, in part, based on their awareness of stakeholders (Grohs et al., 2018). Their inclusion of criteria regarding how many different stakeholders the participant planned to gather input from and engage with demonstrated that considering stakeholders was an important part of their systems thinking definition. As described in the previous sections on breadth, Hu and Shealy's Assessment A included consideration of stakeholders when the categorical and holistic scoring methods (Watson et al., 2016) were used (Hu & Shealy, 2018). In contrast to Grohs, Kirk, Soledad, & Knight's Assessment where awareness of stakeholders is evaluated as its own construct (Grohs et al., 2018), in Hu and Shealy's Assessment, stakeholder considerations are one of several dimensions that impact a dimension (categorical scoring method) or comprehensiveness (holistic scoring method) evaluation as described further by Watson et al. (2016).

Other (O): The Other category was used to draw attention to assessments that explicitly valued some aspect of systems thinking not seen in any of the other 27 assessments. STAR included "intended purpose" as a systems thinking attribute in its rubric, highlighting that engineering systems are created with specific beneficiaries in mind (Lavi, Dori, & Dori, 2020, p. 5; Lavi, Dori, Wengrowicz, et al., 2020, p. 42). Taylor, Calvo-Amodio, &

Well's Assessment included identifying "roles/purposes" for each element (Taylor et al., 2020, p. 9). Grohs, Kirk, Soledad, & Knight's Assessment rubric included a section that checked if participant responses were aligned across different aspects of their response (Grohs et al., 2018). Hrin, Milenković, Segedinac, & Horvat's Assessment explicitly valued the formation of a meaningful whole as it was a requisite of reaching the highest systems thinking level in their scoring rubric (Hrin et al., 2017). SysTest stands apart in that the analysis of responses was guided primarily by participant approaches rather than outcomes. Whereas the majority of assessments focused on what participants attended to or the content of their processes, SysTest focused on what means participants used to attend to different aspects of the problem (e.g. technique use) or their process (Tomko et al., 2017).

5. Discussion

A. Operationalization of Comprehensive Systems Thinking

We inductively identified four different Assessment Types and eight different Formats across 25 assessments. This variety complicated comparisons between assessments and indicates that triangulation through the use of multiple Assessment Types and/or Formats to examine relationships, as done in (Brandstädter et al., 2012; Degen et al., 2018; Hu & Shealy, 2018), could be beneficial. However, to what extent a multiplicity of approaches is more effective than a singular one is unknown.

Examining the ST skills foregrounded in the assessments provided insight into which skills are most valued and how that looks across different Assessment Types. The majority (19 of 27) of the systems thinking assessments were behavior-based, meaning they examined knowledge or skill based on task performance. In addition, all 13 assessments in Group 3, which provided in-depth descriptions of their criteria but typically covered relatively few dimensions of systems thinking, were behavior-based. The most common dimensions among Group 3 assessments were identifying elements of a problem (10 of 13), identifying and/or analyzing connections between elements (12 of 13), and temporal considerations (9 of 13). While the least prominent dimensions were explicit considerations of broader contextual factors (3 of 13) and stakeholders (2 of 13).

In contrast, of the assessments in Group 2 (four which were preference-based and one which was self-reported), all but one (Jaradat et al., 2019) pushed beyond a narrow technical focus. Jaradat, Hamilton, Dayarathna et al.'s Assessment (Jaradat et al., 2019) looked at subset, "Simplicity vs. Complexity", of the systems thinking skills covered in Jaradat & Castelle's Assessment (Castelle & Jaradat, 2016; Jaradat, 2014).

Group 2 covered a wider range of ST skills than Group 3, although it was difficult to directly compare behavior-based assessments to preference-based and self-reported assessments. The majority of Group 2 assessments provided high-level descriptions of relatively many dimensions of systems thinking, and valued broader aspects of systems thinking that those assessments in Group 3. Perhaps this is because assessing performance (Group 3) with regards to stakeholder and broader contextual factors in traditionally technically-focused fields such as engineering and the hard sciences was more difficult than operationalizing interest or self-perception (Group 2). These trends of having few performance-based assessments account for contextual factors reflect and exacerbate the narrow technical focus in engineering fields.

B. Factors for Consideration in Assessment Selection

Our findings suggest several considerations that may inform readers' selection of an assessment, rather than provide general recommendations about which assessments are most useful. These considerations are based a pragmatic analysis of the content and format the assessments described in the source papers, rather than an assessment of the authors' particular ontological framings of systems thinking or reported outcomes of the individual assessments (which risks overinterpretation of journal-length texts that may or may not have allowed substantive discussion of underlying assumptions). Overall, we observed that each of the presented considerations are not always addressed in an assessment's source paper(s) and recognize that ultimately assessment needs will vary by their use context. Thus, as one anonymous reviewer of this paper noted, the value of an assessment may be in its pedagogical affordances, thus challenging face-value judgments. Our discussion thus focuses on use considerations.

One consideration is the effort required for administration. If the assessment will be administered during a single class period (as might be the case with students) or a staff meeting (as might be the case with practitioners), users may prioritize assessments that can be completed in an hour or less, which is the case for many of the assessments in this SLR, e.g. Brandstädter, Harms, & Großschedl's Assessments (Brandstädter et al., 2012), Grohs, Kirk, Soledad, & Knight's Assessment (Grohs et al., 2018), Jaradat, Hamilton, Dayarathna et al.'s Assessment (Jaradat et al., 2019), and Booth Sweeney & Sterman's Assessment (Booth Sweeney & Sterman, 2000), rather than assessments designed to be completed over the course of months, e.g. STAR (Lavi, Dori, & Dori, 2020; Lavi, Dori, Wengrowicz, et al., 2020) and Rehmann, Rover, & Laingen et al.'s Assessment (Rehmann et al., 2011). If the assessment will be administered remotely, e.g., via an electronic survey, it may be easier to select an assessment that uses multiple-choice questions either entirely or mostly, e.g., ESTS (Degen et al., 2018) and Frank and Kordova's Assessment B (Kordova & Frank, 2018), to eliminate some of the clarifying questions that may arise with open-ended questions (e.g., "How long should the response be? How much detail should be included? What file format should be used to share images of created visualizations?"). However, if the assessment will be administered remotely and the creation of a visualization is important to the learning outcomes or goals of the assessment, it may be beneficial to select assessments that create visualizations using freely available software such as MentalModeler[1] or OPCAT[2], as described in Gray et al.'s (2019) and Lavi et al.'s work (2020; 2020), respectively.

Another consideration that may inform assessment selection is the training required for the assessor, with regard to both administering the assessment and interpreting responses (Geisinger, 2015). Factors to consider here may include access to training materials, reliability of the scoring method, and availability of the assessment in a ready-to-distribute format. For example, there is a wide range across source papers in terms of the amount of detail provided when describing the procedure to administer the assessment and the source paper(s) may not include all the instructions that were shared with participants. In addition, some assessments provide scoring protocols or rubrics in the source papers, e.g., Brandstädter, Harms, & Großschedl's Assessment A (Brandstädter et al., 2012), Grohs, Kirk, Soledad, & Knight's Assessment (Grohs et al., 2018), STAR (Lavi, Dori, & Dori, 2020; Lavi, Dori, Wengrowicz, et al., 2020) and Rehmann, Rover, & Laingen et al.'s Assessment (Rehmann et al., 2011), but these vary in terms of how much detail is provided regarding how scores are assigned, and if applicable, how total scores are calculated.

Other considerations are associated with the format of the assessment. For example, selecting assessments with open-ended questions may emphasize the evaluation of higher order thinking such as "creating" from a revision of Bloom's Taxonomy (Anderson & Krathwohl, 2001). In contrast, multiple-choice items can only indirectly assess participants' abilities to create new ideas and solve ill-structured problems (Brookhart & Nitko, 2019). Alternatively, selecting assessments that leverage scenarios may emphasize the importance of context or foreground the context-dependent nature of many complex problems. Thus, people most interested in using an assessment to evaluate ideas students create with respect to real world problems might be more interested in open-ended, scenario-based assessments. On the other hand, assessments that leverage multiple-choice questions may be most useful to those interested in assessing comprehension of "concepts, principles, and generalizations" (Brookhart & Nitko, 2019, p. 184).

C. Limitations of Existing ST Assessments

Many of the assessments included in this SLR were focused on examining an individual's systems thinking and are thus limited because problems that necessitate the use of systems thinking are so complex that, by definition, they need a team to address them. However, our analyses did not foreground the individual versus team focus of the included assessments, as we assume there is a degree of artificiality in any easily scored assessment of systems thinking. Any such assessment is unlikely to fully capture the complex team decision

[1] http://www.mentalmodeler.org/
[2] http://esml.iem.technion.ac.il/opcat-installation/

making dynamics of real-world systems thinking.

The SLR demonstrated that even within engineering and science contexts, there was inconsistent use of terms and in conceptions of what systems thinking means and includes. One inconsistency prevalent in the papers reviewed in this SLR was the use of both "*systems* thinking" and "*system* thinking." This inconsistent use of language inhibits clarity not only because these terms may be understood as distinct types of reasoning, but because the use of one term or the other may signal a specific mode a reasoning that the author did not intend.

Another common issue in discussing and defining systems thinking (Tomko et al., 2017) was evident in many articles in this SLR: the lack of distinction between systems thinking as a discipline-independent skillset or as a prerequisite to/component of systems engineering as a discipline. Tomko et al.'s (Tomko et al., 2017) research on a systems thinking assessment recognized a relationship between how "openly defined" systems thinking is and the variety of systems thinking—or sometimes systems engineering skills—that have been constructed by scholars. The connections to systems engineering ranged from assessments that framed their work as supporting systems engineering education and workforce development (Camelia & Ferris, 2018; Camelia et al., 2018; Frank, 2007, 2009, 2010; Frank & Kordova, 2015), to ESTS that was explicitly described as incorporating both systems thinking and systems engineering knowledge, skills, and abilities (Degen et al., 2018), to STAR that made implicit connections to systems engineering through an evaluation format derived from model-based systems engineering and a rubric based on literature that included systems engineering books (Lavi, Dori, Wengrowicz, et al., 2020). While successful systems engineers need systems thinking skills (Frank & Kordova, 2015), given the complex challenges facing our world today modern engineers need systems thinking skills regardless of their discipline (National Academy of Engineering, 2004). A lack of consistent language to signal if an assessment is targeting systems engineers reduces the usability of systems thinking assessments across engineering disciplines. For instance, Bedir, Desai, and Kulkarni et al.'s Assessment used the terms "systems thinking" and "systems engineering" interchangeably; referring to "objectives and metrics" as "systems thinking concepts" and "two fundamental [systems engineering] topics" while claiming their assessment demonstrated an online module increased "understanding of systems thinking" and "understanding of systems engineering concepts," making it difficult to determine the assessment's intended use (Bedir et al., 2020, sec. Abstract, Methdology-Module Development, Results and Discussion-Hypothesis II, Conclusion).

Perhaps even more importantly, the lack of consistent positioning of systems thinking with respect to systems engineering may leave engineers with the misapprehension that systems engineering, as both a discipline and a methodology, is the only way that systems thinking can be applied in engineering spaces. Systems engineering is one of many systems approaches in applied systems thinking (Jackson, 2019). Jackson (2019) provides a history of the development of systems engineering, among other hard systems thinking approaches, an overview the methodology, methods, and developments, and a summary of different critiques. One critique argues that hard systems thinking may not adequately address extreme complexity because it assumes all factors can be quantified, leading to the tendency to ignore or distort factors that are difficult to quantify. This challenge may be reflected in our finding that broader contextual factors were more frequently attended to in preference-based assessments than in behavior-based assessments. An additional critique of hard approaches relates to their ability to adequately capture different stakeholder perspectives and values. Jackson emphasizes the relevance of considering stakeholder subjectivity—in terms of defining objectives—within management situations, which stands in contrast to "engineering type problems" (2019, p. 193), suggesting engineering problems may be less complex in terms of achieving stakeholder agreement on objectives for a problem. However, our framing of *comprehensive* systems thinking extends this critique to engineering-type problems as well, arguing that engineers must also consider different stakeholders and the complexity of differing objectives. We see this limitation playing out in the SLR findings in that while the majority of the assessments are behavior-based—a perhaps unsurprising finding in the context of engineering—these behavior-based assessments often do not attend to relevant broader contextual factors or stakeholders.

D. Positioning of Comprehensive Systems Thinking in Current Work

We see *comprehensive* systems thinking as a challenge to traditional engineering practice. *Comprehensive* systems thinking is not a methodology; it is a call for a reframing of engineers' approaches to complex problem solving, regardless of discipline, and a framework for what such problem-solving approaches should look like. It advocates a holistic, rather than reductionist, approach, incorporating broader contextual factors in addition to the constituent elements of an immediate problem, and recognizing that, increasingly, the problems engineers work on are sociotechnical problems. Our conceptualization of sociotechnical problems is consistent with Dori's definition of "a socio-technical system, also known as [an] engineering system, is a system that integrates technology, people, and services, combining perspectives from engineering, management, and social sciences," (Dori, 2016, p. 88). Our position thus aligns with critiques of systems engineering as a methodology and recent recognition of the limitations of systems engineering as a discipline, while also acting in parallel to on-going debates regarding systems methodologies. For example, systems engineering as a discipline often relies on reductionist methods (Rousseau, 2019) and the International Council on Systems Engineering (INCOSE) has even recognized the need to shift its "emphasis from reductionism to holism" (INCOSE, 2014, p. 41). Related work includes Rousseau's (2019, 2020) calls for a strengthening of the theory behind holistic approaches in systems research and Yearworth's (2020) suggestion that systems engineering should learn from the social sciences and that soft systems methodology (SSM) should be reconnected to engineering practice. Similarly, Mingers' (2011) argument that soft operations research (Soft OR), which includes SSM, and Hard OR are complements to be used strategically depending on the problem situation, is well aligned with the idea of complementarism, which emphasizes applying all systems approaches relevant to a particular problem context, in critical systems theory (Jackson, 2019). Our conceptualization of comprehensive systems thinking as a general approach is not limited to systems engineering practice as a discipline. While comprehensive systems thinking is not a methodology or theory, there are a number of instances of overlap or alignment with existing methodologies or theories. For example, stakeholder participation or the use of participatory methods is central to SSM and other Soft OR approaches (Mingers, 2011; Yearworth, 2020). While stakeholder participation is not a defining characteristic of comprehensive systems thinking, it is consistent with the approach's emphasis on developing a broad understanding of how various stakeholders are at play in the problem. Critical systems theory has a commitment to "[bring] about those circumstances in which all individuals can achieve the maximum development of their potential," (Jackson, 2019, p. 523). Though this goal is not explicit to our definition of comprehensive systems thinking, such a goal aligns with comprehensive systems thinking's calls for awareness of the potential harm (as well as the benefits) solutions could have, better and ethical service to a range of stakeholders (e.g., not only the funder of a project), and consideration of people-related, cultural, and environmental factors and impacts that are increasingly central to the problems engineers seek to address.

E. Limitations of the SLR

Several limitations may affect how the results of this work are interpreted. One limitation of our SLR was that it may not include all systems thinking assessments in engineering due to our strategic choices with regards to database selection and search string creation, as well as our decision to exclude book excerpts. Another limitation is that because we only reported on details provided in the source papers that met all of our inclusion criteria, additional details of assessments may have been discoverable had we searched beyond our criteria. In addition, our analyses were not strictly limited to assessments designed for engineering students or practitioners, making some assessments potentially less relevant to engineering. All the assessments, however, had some connection to engineering or engineers through the search criteria and we considered it important to show the breadth of relevant assessments in order to illustrate their limitations. Finally, we did not report on or analyze the outcomes of the different assessments. Future work could investigate these outcomes and compare the alignment of outcomes across systems thinking assessments.

F. Implications

Our findings point to several implications for systems thinking assessment development and use. Employers and educators can use this overview of available assessments to select assessment(s) that meet their use cases (e.g., assessment content area) and align with their goals (e.g., evaluating demonstrated skill). This review also highlights, for systems thinking researchers, the importance of explicitly contextualizing the purpose and use cases for their assessments to ensure they can be employed by appropriate populations. The lack of clarity in assessments suggests that systems thinking scholarship as a whole cold benefit from such explicit statements. This study demonstrated a need for behavior-based, comprehensive systems thinking assessments. The Accreditation Board of Engineering and Technology (ABET) continues to include student outcomes that are related to contextual competence (ABET, 2019) and although many of these assessments are relatively new (19 of the 25 assessments were published in the last ten years), this SLR showed that overall, many behavior-based assessments covered system element identification, making connections between elements, and temporal considerations, but few addressed the broader context or stakeholders. Most of the assessments that pushed beyond a narrow technical focus were preference-based; these cannot provide an in-depth understanding of which systems thinking skills engineering students or professionals have or need developed. A lack of behavior-based assessments, means engineering as a field will continue to devalue the importance of understanding contextual aspects of the problem when evaluating the development of systems thinking skills.

6. Conclusions

We identified 27 unique systems thinking assessments across 30 papers that were screened from a total of 130 papers. We characterized these assessments according to Type—what and/or how dimensions of systems thinking were evaluated, Format—their structure, and Content Area—the topic around which it was based. We inductively derived and defined four Assessment Types: behavior-based, preference-based, self-reported, and cognitive activation, and eight Assessment Formats: mapping, scenario, open-ended, oral, fill-in-the-blank, multiple-choice, virtual reality, and fNIRS cap. Our study's findings can support employers and educators in selecting assessment(s) that will meet their contexts and needs. In addition, we analyzed how definitions of systems thinking were conveyed from two perspectives: the author's definition in the source paper(s) and the aspects of systems thinking emphasized in the assessment's rubric. In conclusion, of the systems thinking assessments that pushed beyond a narrow technical focus the majority were preference-based assessments. Overall, most of the assessments were behavior-based, indicating a lack of behavior-based assessments that operationalize aspects of *comprehensive* systems thinking. Comprehensive systems thinking advocates for more holistic problem-solving and provides a framework for what comprehensive approaches should look like, including the explicit attention to stakeholders and contextual factors that are increasingly essential parts of the sociotechnical problems engineers work on. Systems thinking researchers can use this SLR to inform the development of new, *comprehensive* systems thinking assessments that evaluate performance. Without such development, engineering as a field will continue to undervalue the key role of integrating contextual aspects of the problem in the development of successful solutions when assessing systems thinking skill development.

References

ABET. (2019). Criteria for Accrediting Engineering Programs. Effective for Reviews during the 2020-2021 Accreditation Cycle. In *Criteria for Accrediting Engineering Programs*. Retrieved from www.abet.org

Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing*. New York: Longman.

Aranda-Jan, C. B., Jagtap, S., & Moultrie, J. (2016). Towards a framework for holistic contextual design for low-resource settings. *International Journal of Design*, *10*(3), 43–63. https://doi.org/10.17863/CAM.7254

Automation systems and integration - Object-Process Methodology. (2015). *ISO/PAS 19450*, pp. 1–162. Retrieved from https://www.iso.org/standard/62274.html

Bahill, A. T., & Gissing, B. (1998). Re-evaluating systems engineering concepts using systems thinking. *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, *28*(4), 516–527. https://doi.org/10.1109/5326.725338

Bedir, A., Desai, R., Kulkarni, N., Wallet, K., Wells, R., & Smith, M. (2020). Assessing Student Learning of Systems Thinking Concepts in an Online Education Module. *2020 Systems and Information Engineering Design Symposium (SIEDS)*, 1–6. https://doi.org/10.1109/SIEDS49339.2020.9106680

Ben-Zvi Assaraf, O., & Orion, N. (2005). Development of system thinking skills in the context of Earth System Education. *Journal of Research in Science Teaching*, *42*, 518–560.

Benninghaus, J. C., Mühling, A., Kremer, K., & Sprenger, S. (2019a). Complexity in education for sustainable consumption-an educational data mining approach using mysteries. *Sustainability (Switzerland)*, *11*(3). https://doi.org/10.3390/su11030722

Benninghaus, J. C., Mühling, A., Kremer, K., & Sprenger, S. (2019b). The mystery method reconsidered—A tool for assessing systems thinking in education for sustainable development. *Education Sciences*, *9*(4). https://doi.org/10.3390/educsci9040260

Boersma, K., Waarlo, A. J., & Klaassen, K. (2011). The feasibility of systems thinking in biology education. *Journal of Biological Education*, *45*(4), 190–197. https://doi.org/10.1080/00219266.2011.627139

Booth Sweeney, L., & Sterman, J. D. (2000). Bathtub dynamics: Initial results of a systems thinking inventory. *System Dynamics Review*, *16*(4), 249–286. https://doi.org/10.1002/sdr.198

Borrego, M., Foster, M. J., & Froyd, J. E. (2015). What is the state of the art of systematic review in engineering education? *Journal of Engineering Education*, *104*(2), 212–242. https://doi.org/10.1002/jee.20069

Brandstädter, K., Harms, U., & Großschedl, J. (2012). Assessing System Thinking Through Different Concept-Mapping Practices. *International Journal of Science Education*, *34*(14), 2147–2170. https://doi.org/10.1080/09500693.2012.716549

Brookhart, S. M., & Nitko, A. J. (2019). Multiple-Choice and Matching Exercises; Higher-Order Thinking, Problem Solving, and Critical Thinking. In *Educational Assessment of Students* (8th ed., pp. 181–238). New York, NY: Pearson.

Camelia, F., & Ferris, T. L. J. (2018). Validation studies of a questionnaire developed to measure students' engagement with systems thinking. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, *48*(4), 574–585. https://doi.org/10.1109/TSMC.2016.2607224

Camelia, F., Ferris, T. L. J., & Cropley, D. H. (2018). Development and Initial Validation of an Instrument to Measure Students' Learning about Systems Thinking: The Affective Domain. *IEEE Systems Journal*, *12*(1), 115–124. https://doi.org/10.1109/JSYST.2015.2488022

Castelle, K. M., & Jaradat, R. M. (2016). Development of an Instrument to Assess Capacity for Systems Thinking. *Procedia Computer Science*, *95*, 80–86. https://doi.org/10.1016/j.procs.2016.09.296

Checkland, P. (1983). O.R. and the Systems Movement: Mappings and Conflicts. *Journal of the Operational Research Society*, *8*(4), 661–675.

Checkland, P. (2000). Soft systems methodology: a thirty year retrospective. *Systems Research and Behavioral Science*, *17*(S1), S11–S58. https://doi.org/10.1002/1099-1743(200011)17:1+<::aid-sres374>3.3.co;2-f

Crawley, E., Cameron, B., & Selva, D. (2016). *System Architecture: Strategy and Product Development for Complex Systems*. Hoboken, NJ, USA: Pearson Higher Education, Inc.

Crawley, E. F., Malmqvist, J., Östlund, S., Brodeur, D. R., & Edström, K. (2014). Student Learning Assessment. In *Rethinking Engineering Education* (pp. 165–180). https://doi.org/10.1007/978-3-319-05561-9_7

Degen, C. M., Muci-Küchler, K. H., Bedillion, M. D., Huan, S., & Ellingsen, M. (2018). Measuring the Impact of a New Mechanical Engineering Sophomore Design Course on Students' Systems Thinking Skills. *Proceedings of the ASME 2016 International Engineering Congress and Exposition*, 1–9. Pittsburgh, PA.

Dori, D. (2016). Model-based systems engineering with OPM and SysML. In *Model-Based Systems Engineering with OPM and SysML*. https://doi.org/10.1007/978-1-4939-3295-5

Fahmy, A. F. M., & Lagowski, J. . (2014). Systemic Assessment as a New Tool for Assessing Students Learning in Chemistry using SATL Methods: Systemic Matching [SMQS], Systemic Synthesis [SSYNQS], Systemic Analysis [SAN Q, S], Systemic Synthetic – Analytic [SSYN-ANQ, S], as Systemic Question Types. *African Journal of Chemical Education*, *4*(4), 35–55.

Frank, M. (2000). Engineering systems thinking and systems thinking. *Systems Engineering*, *3*(3), 163–168.

https://doi.org/10.1002/1520-6858(200033)3:3<163::aid-sys5>3.3.co;2-k
Frank, M. (2006). Knowledge, abilities, cognitive characteristics and behavioral competences of engineers with high Capacity for Engineering Systems Thinking (CEST). *Systems Engineering*, *9*(2), 91–103. https://doi.org/10.1002/sys.20048
Frank, M. (2007). Towards a quantitative tool for assessing the capacity for engineering systems thinking. *International Journal Human Resources Development and Management*, *7*(3/4), 240–253.
Frank, M. (2009). Capacity for Engineering Systems Thinking (CEST): Literature Review, Principles fro Assessing and the Reliability and Validity of an Assessing Tool. *Internatinal Journal of Information Technologies and Systems Approach*, 1–14. https://doi.org/10.1002/asi.4630330202
Frank, M. (2010). Assessing the interest for systems engineering positions and other engineering positions' required capacity for engineering systems thinking (CEST). *Systems Engineering*, *13*(2), 161–174. https://doi.org/10.1002/sys.20140
Frank, M., & Elata, D. (2005). Developing the Capacity for Engineering Systems Thinking (CEST) of freshman engineering students. *Systems Engineering*, *8*(2), 187–195. https://doi.org/10.1002/sys.20025
Frank, M., & Kordova, S. (2015). Four Layers Approach for Developing System Thinking Assessment Tool for Industrial and Systems Engineers. *Industrial Engineering & Management*. https://doi.org/10.4172/2169-0316.1000178
Funke, J. (1991). Solving complex problems: Exploration and control of complex systems. In R. J. Sternberg & P. A. Frensch (Eds.), *Complex problem solving: Principles and mechanisms* (pp. 185–122). Hillsdale, NJ: Lawrence Erlbaum Associates.
Geisinger, K. F. (2015). Test Evaluation. In S. Lane, M. R. Raymond, & T. M. Haladyna (Eds.), *Handbook of Test Development* (2nd ed.). https://doi.org/10.4324/9780203102961
Gough, D., Sandy, O., & James, T. (2012). *An Introduction To Systematic Reviews* (2nd Ed.; J. Seaman, Ed.). SAGE Publications Ltd.
Gray, S., Sterling, E. J., Aminpour, P., Goralnik, L., Singer, A., Wei, C., … Norris, P. (2019). Assessing (social-ecological) systems thinking by evaluating cognitive maps. *Sustainability (Switzerland)*, *11*(20), 1–11. https://doi.org/10.3390/su11205753
Grohs, J. R. (2015). *Developing a Measure of Systems Thinking Competency* (Virginia Polytehnic Institute and State University). https://doi.org/10.1017/CBO9781107415324.004
Grohs, J. R., Kirk, G. R., Soledad, M. M., & Knight, D. B. (2018). Assessing systems thinking: A tool to measure complex reasoning through ill-structured problems. *Thinking Skills and Creativity*, *28*, 110–130. https://doi.org/10.1016/j.tsc.2018.03.003
Grotzer, T. A. (2012). *Learning Causality in a Complex World: Understandings of Consequence*. Rowman & Littlefield Education.
Hadgraft, R. G., Carew, A. L., Therese, S. A., & Blundell, D. L. (2008). Teaching and assessing systems thinking in engineering. *Research in Engineering Education Symposium 2008*, 230–235.
Hayden, N. J., Rizzo, D. M., Dewoolkar, M. M., Oka, L., & Neumann, M. (2010). Incorporating systems thinking and sustainability within civil and environmental engineering curricula at UVM. *ASEE Northeast Regional Conference Proceedings*, *26*. Rochester, NY.
Hogan, K., & Weather, K. C. (2003). Psychological and ecological perspectives on the development of systems thinking. In *Understanding urban ecosystems* (pp. 233–260). New York, NY: Springer.
Hrin, T. N., Milenković, D. D., Segedinac, M. D., & Horvat, S. (2017). Systems thinking in chemistry classroom: The influence of systemic synthesis questions on its development and assessment. *Thinking Skills and Creativity*, *23*, 175–187. https://doi.org/10.1016/j.tsc.2017.01.003
Hu, M., & Shealy, T. (2018). Methods for measuring systems thinking: Differences between student self-assessment, concept map scores, and cortical activation during tasks about sustainability. *ASEE Annual Conference and Exposition*.
INCOSE. (2014). *A world in motion—Systems engineering vision 2025*. San Diego, CA: International Council on Systems Engineering.
Jackson, M. C. (2019). *Critical Systems Thinking and the Management of Complexity*. Chichester: Wiley.
Jaradat, R. M. (2014). *An Instrument to Assess Individual Capacity for System Thinking* (Old Dominion University). https://doi.org/10.25777/wzh1-2563
Jaradat, R. M., Hamilton, M. A., Dayarathna, V. L., Karam, S., Jones, P., Wall, E. S., … En Hsu, G. S. (2019). Measuring individuals' systems thinking skills through the development of an immersive virtual reality complex system scenarios. *ASEE 126th Annual Conference and Exposition, Conference Proceedings*.
Jonassen, D. H. (2000). Toward a design theory of problem solving. *Educational Technology Research and Development*, *48*(4), 63–85. https://doi.org/10.1007/BF02300500
Keynan, A., Ben-Zvi Assaraf, O., & Goldman, D. (2014). The repertory grid as a tool for evaluating the development of students' ecological system thinking abilities. *Studies in Educational Evaluation*, *41*, 90–105. https://doi.org/10.1016/j.stueduc.2013.09.012
Kilgore, D., Atman, C. J., Yasuhara, K., Barker, T. J., & Morozov, A. (2007). Considering Context: A Study of First-Year Engineering Students. *Journal of Engineering Education*, *96*(4), 321–334.
Kordova, S., & Frank, M. (2018). Systems Thinking as an Engineering Language. *American Journal of Systems Science*, *6*(1), 16–28. https://doi.org/10.5923/j.ajss.20180601.02
Kroes, P., Franssen, M., van de Poel, I., & Ottens, M. (2006). Treating socio-technical systems as engineering systems: some conceptual problems. *Systems Research and Behavioral Science*, *23*(6), 803–814.
Latta, G. F., & Swigger, K. (1992). Validation of the repertory grid for use in modeling knowledge. *Journal of the American Society for Information Science*, *43*(2), 115–129. https://doi.org/10.1002/(SICI)1097-4571(199203)43:2<115::AID-ASI2>3.0.CO;2-I
Lattuca, L. R., Knight, D. B., Ro, H. K., & Novoselich, B. J. (2017). Supporting the Development of Engineers' Interdisciplinary Competence. *Journal of Engineering Education*, *106*(1), 71–97. https://doi.org/10.1002/jee.20155
Lavi, R., Dori, Y. J., & Dori, D. (2020). Assessing Novelty and Systems Thinking in Conceptual Models of Technological Systems. *IEEE*

Transactions on Education, 1–8. https://doi.org/10.1109/TE.2020.3022238
Lavi, R., Dori, Y. J., Wengrowicz, N., & Dori, D. (2020). Model-Based Systems Thinking: Assessing Engineering Student Teams. *IEEE Transactions on Education*, *63*(1), 39–47. https://doi.org/10.1109/TE.2019.2948807
Leat, D. (1998). *Thinking Through Geography*. London, UK: Chris Kington Publishing.
Mazzurco, A., & Daniel, S. (2020). Socio-technical thinking of students and practitioners in the context of humanitarian engineering. *Journal of Engineering Education*, *109*, 243–261. https://doi.org/10.1002/jee.20307
Meilinda, Rustaman, N. Y., Firman, H., & Tjasyono, B. (2018). Development and validation of climate change system thinking instrument (CCSTI) for measuring system thinking on climate change content. *Journal of Physics: Conference Series*, *1013*(1). https://doi.org/10.1088/1742-6596/1013/1/012046
Mingers, J. (2011). Soft OR comes of age-but not everywhere! *Omega*, *39*(6), 729–741. https://doi.org/10.1016/j.omega.2011.01.005
Monat, J., & Gannon, T. (2018). Applying Systems Thinking to Engineering and Design. *Systems*, *6*(3), 34. https://doi.org/10.3390/systems6030034
Mote, C. D., Dowling, D. A., & Zhou, J. (2016). The Power of an Idea: The International Impacts of the Grand Challenges for Engineering. *Engineering*, *2*(1), 4–7. https://doi.org/10.1016/J.ENG.2016.01.025
National Academy of Engineering. (2004). The Engineer of 2020. In *The engineer of 2020: Visions of engineering in the new century*. https://doi.org/10.17226/10999
National Academy of Engineering. (2020). NAE Grand Challenges for Engineering. Retrieved September 23, 2020, from http://www.engineeringchallenges.org/challenges.aspx
Palmer, B., Mckenna, A. F., Harper, B. J., Terenzini, P., & Merson, D. (2011). Design in Context: Where do the Engineers of 2020 Learn this Skill? *ASEE Annual Conference & Exposition*. Retrieved from https://peer.asee.org/17711
Plack, M. M., Goldman, E. F., Scott, A. R., Pintz, C., Herrmann, D., Kline, K., … Brundage, S. B. (2018). Systems Thinking and Systems-Based Practice Across the Health Professions: An Inquiry Into Definitions, Teaching Practices, and Assessment. *Teaching and Learning in Medicine*, *30*(3), 242–254. https://doi.org/10.1080/10401334.2017.1398654
Rebovich, G. (2006). Systems thinking for the enterprise: New and emerging perspectives. *IEEE/SMC International Conference on System of Systems Engineering*, 197–202. https://doi.org/10.1109/sysose.2006.1652297
Rehmann, C. R., Rover, D. T., Laingen, M., Mickelson, S. K., & Brumm, T. J. (2011). Introducing systems thinking to the engineer of 2020. *ASEE Annual Conference and Exposition*.
Reynolds, M., & Holwell, S. (Eds.). (2010). *Systems approach to managing change: A pratical guide*. Springer Science+Business Media.
Rittel, H. W. J., & Webber, M. M. (1973). Dilemmas in a general theory of planning. *Policy Sciences*, *4*(2), 155–169. https://doi.org/10.1007/BF01405730
Ro, H. K., Merson, D., Lattuca, L. R., & Terenzini, P. T. (2015). Validity of the contextual competence scale for engineering students. *Journal of Engineering Education*, *104*(1), 35–54. https://doi.org/10.1002/jee.20062
Rousseau, D. (2019). A vision for advancing systems science as a foundation for the systems engineering and systems practice of the future. *Systems Research and Behavioral Science*, *36*(5), 621–634.
Rousseau, D. (2020). The Theoretical Foundation(s) for Systems Engineering? Response to Yearworth. *Systems Research and Behavioral Science*, *37*(1), 188–191. https://doi.org/10.1002/sres.2671
Schuler, S., Fanta, D., Rosenkraenzer, F., & Riess, W. (2018). Systems thinking within the scope of education for sustainable development (ESD)–a heuristic competence model as a basis for (science) teacher education. *Journal of Geography in Higher Education*, *42*(2), 192–204. https://doi.org/10.1080/03098265.2017.1339264
Senge, P. M. (1990). *The fifth discipline: The art and practice of the learning organization*. New York: Doubleday.
Strijbos, S. (2003). Systems thinking and the disclosure of a technological society: some philosophical reflections. *Systems Research and Behavioral Science*, *20*(2), 119–131.
Taylor, S., Calvo-Amodio, J., & Well, J. (2020). A method for measuring systems thinking learning. *Systems*, *8*(2), 1–36. https://doi.org/10.3390/systems8020011
Timofte, R. S., & Popuș, B. T. (2019). Assessment Tasks to Measure Systems Thinking and Critical Thinking in Organic Chemistry. *Acta Chemica Iasi*, *27*(2), 251–262. https://doi.org/10.2478/achi-2019-0016
Tomko, M., Nelson, J., Linsey, J., Bohm, M., & Nagel, R. (2017). Towards assessing student gains in systems thinking during engineering design. *Proceedings of the International Conference on Engineering Design, ICED*, *9*(DS87-9), 179–188.
Trochim, W. M., Cabrera, D. A., Milstein, B., Gallagher, R. S., & Leischow, S. J. (2006). Practical challenges of systems thinking and modeling in public health. *American Journal of Public Health*, *96*(3), 538–546. https://doi.org/10.2105/AJPH.2005.066001
Vanasupa, L., Rogers, E., & Chen, K. (2008). Work in progress: How do we teach and measure systems thinking? *ASEE/IEEE Frontiers in Education Conference*, 1–2. https://doi.org/10.1109/FIE.2008.4720378
Watson, M. K., Pelkey, J., Noyes, C. R., & Rodgers, M. O. (2016). Assessing Conceptual Knowledge Using Three Concept Map Scoring Methods. *Journal of Engineering Education*, *105*(1), 118–146. https://doi.org/10.1002/jee.20111
Wiggins, G., & McTighe, J. (2005). What is Backward Design? In *Understanding by Design* (2nd ed.). Alexandria, VA: Association for Supervision and Curriculum Development.
Wilensky, U., & Resnick, M. (1999). Thinking in levels: A dynamic systems approach to making sense of the world. *Journal of Science Education and Technology*, *8*(1), 3–19. https://doi.org/10.1023/A:1009421303064
Wiliam, D. (2011). What is assessment for learning? *Studies in Educational Evaluation*, *37*(1), 3–14. https://doi.org/10.1016/j.stueduc.2011.03.001

Yearworth, M. (2020). The theoretical foundation(s) for Systems Engineering? *Systems Research and Behavioral Science*, *37*(1), 184–187. https://doi.org/10.1002/sres.2667
York, S., & Orgill, M. K. (2020). ChEMIST Table: A Tool for Designing or Modifying Instruction for a Systems Thinking Approach in Chemistry Education. *Journal of Chemical Education*, *97*(8), 2114–2129. https://doi.org/10.1021/acs.jchemed.0c00382
Zoller, U., & Scholz, R. W. (2004). The HOCS paradigm shift from disciplinary knowledge (LOCS) - To interdisciplinary evaluate, system thinking (HOCS): What should it take science-technology-environment-society oriented courses, curricula and assessment? *Water Science and Technology*, *49*(8), 27–36. https://doi.org/10.2166/wst.2004.0480

Appendix A

REPERTORY GRID (RG) TECHNIQUE DEFINITION

"a form of highly structured interview, formalizing the interactions of the interview and interviewee and putting into relations personal constructs and given objects of discourse. " (Keynan et al., 2014). "The building blocks of the RG are *elements* (the topics of study), *constructs* (the participants' ideas about these elements) and *ratings* (relations among elements and constructs as viewed by the participants). Elements are the objects that are the focus of the investigation" (Keynan et al., 2014). See (Latta & Swigger, 1992) for more info.

Appendix B

TERMS FROM MAPPING ASSESSMENTS

Term	Definition
Concept Map	"begins 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" (Hu & Shealy, 2018).
Cognitive Map	"include elements that can increase or decrease in quality and quantity and relationships between elements are represented by positive influences (blue lines) and negative influences (red lines)" (Gray et al., 2019).
Fuzzy Cognitive Map	"represent systems as directed and weighted graphs, where the nodes of the graph qualitatively represent elements of the system (i.e., concepts), and the edges between the nodes quantitatively represent the direction and strength of causal relationships between concepts" (Gray et al., 2019).
Mystery Method	"Students are faced with an initial question or problem… They then need to arrange the cards in a way that explains this mysterious question or problem"(Benninghaus, Mühling, Kremer, & Sprenger, 2019a). See (Leat, 1998) for more information.
Influence Diagram	"an influence diagram (or causal-loop diagram), which shows system elements . . . and interrelations (arrows) between them" (Schuler, Fanta, Rosenkraenzer, & Riess, 2018). "example, a qualitative system model, consisting of system elements (nodes) and system relationships (influences, arrows)" (Schuler et al., 2018).
Mystery Map	"We would like to refer to the influence diagrams, emerging form the mystery methods, as *mystery maps*, as the basic principle of connecting the cards is similar to linking concepts in concept mapping." (Benninghaus et al., 2019b). Referred to as a type of concept map (Benninghaus et al., 2019a). Also referred to as an influence diagram (Benninghaus et al., 2019a, 2019b; Schuler et al., 2018).
Causal Loop Diagram	"similar to concept maps, showing how one concept… is linked to another…The difference, however, is that casual loop diagrams depict how *changes* in one concept are linked to *changes* in another" (Vanasupa et al., 2008).
Behavior Over Time Graph	"These are usually schematic depictions of how an important variable behaves over time, although they can also include real data" (Vanasupa et al., 2008).
Rich Pictures	"the aim is to capture, informally, the main entities, structures and viewpoints in the situations, the processes going on, the current recognized issues and any potential ones" (Reynolds & Holwell, 2010, p. 210). " *formalization* via use of ready-made fragments . . . is not usually a good idea " (Checkland, 2000).
SAQs - Systemic Assessment Questions	"The Systemic Assessment questions [SAQ] is a novel assessment tool which combines the ideas from systemic and constructivism and adjusts the in a concept map like structure"(Fahmy & Lagowski, 2014). See (Fahmy & Lagowski, 2014) for [SAQ] design guidelines.
SSynQs - Systemic Synthesis Questions	"we took a specific type of [SAQs] - [SSynQs]" (Hrin et al., 2017). "[SSynQs] required students to recognize relations highlighted on the arrows, as well as initial concept, in unfilled and/or partially filled diagrammatic tasks" (Hrin et al., 2017).
Conceptual Model	"are products of the system representation process in model-based systems engineering (MBSE). Unlike concept maps, conceptual models are constructed using a formal graphical language and are more expressive than concept maps, clearly distinguishing between different types of concepts and interrelationships (Dori, 2016)" (Lavi, Dori, Wengrowicz, et al., 2020).
Object-Process Methodology (OPM)	"a systems modeling paradigm that represents the two things inherent in a system: its objects and process. OPM is fundamentally simple; it builds on a minimal set of concepts: stateful objects—things that exist, and process—things that happen and transform objects by creating or consuming them or by changing their states" (Dori, 2016, p. v). See. ("Automation systems and integration - Object-Process Methodology," 2015) for more information.

TABLE 1
SUMMARY OF ANALYSIS CATEGORIES

	Driving Question	Analysis Category
RQ1	What are the existing assessments in engineering for (ST)? What populations are these assessments targeting?	• Assessment Name (AP) • Education Level Targeted (AP)
RQ2	How are these assessments measuring systems thinking? What is the format of these assessments?	• Assessment Type (AP) • Assessment Format (AP) • Assessment Content Area (E)
RQ3	How is ST defined in the assessment's paper? What dimensions of ST are assessed / foregrounded?	• ST Definition (AP) • Dimensions of ST (AP) • Evaluation Criteria (E)

Key: "AP" means the category was developed a priori, while "E" indicates that a category emerged during the iterative analysis.

TABLE 2
NAME AND SOURCE OF THE 27 UNIQUE ASSESSMENTS IDENTIFIED IN THIS SLR

Assessment	**Source**
Systems Thinking Assessment Rubric (STAR)	(Lavi, Dori, & Dori, 2020; Lavi, Dori, Wengrowicz, et al., 2020)
Taylor, Calvo-Amodio, and Well's Assessment	(Taylor et al., 2020)
Bedir, Desai, Kulkarni et al.'s Assessment	(Bedir et al., 2020)
Mystery Maps	(Benninghaus et al., 2019b)
Gray, Sterling, Aminpour et al.'s Assessment	(Gray et al., 2019)
Jaradat, Hamilton, Dayarathna et al.'s Assessment	(Jaradat et al., 2019)
Timofte & Popuș' Assessment	(Timofte & Popuș, 2019)
Grohs, Kirk, Soledad, & Knight's Assessment	(Grohs et al., 2018)
Frank & Kordova's Assessment B	(Kordova & Frank, 2018)
Engineering Systems Thinking Survey (ESTS)	(Degen et al., 2018)
Camelia, Ferris, & Cropley's Assessment	(Camelia & Ferris, 2018; Camelia et al., 2018)
Hu & Shealy's Assessment A Hu & Shealy's Assessment B	(Hu & Shealy, 2018)
Climate Change System Thinking Instrument (CCSTI)	(Meilinda et al., 2018)
Systems Assessment Test (SysTest)	(Tomko et al., 2017)
Hrin, Milenković, Segedinac, & Horvat's Assessment	(Hrin et al., 2017)
Jaradat & Castelle's Assessment	(Castelle & Jaradat, 2016; Jaradat, 2014)
Frank & Kordova's Assessment A	(Frank, 2007, 2009, 2010; Frank & Kordova, 2015)
Keynan, Ben-Zvi Assaraf, & Goldman's Assessment	(Keynan et al., 2014)
Brandstädter, Harms, & Großschedl's Assessment A Brandstädter, Harms, & Großschedl's Assessment B	(Brandstädter et al., 2012)
Rehmann, Rover, & Laingen et al.'s Assessment	(Rehmann et al., 2011)
Hadgraft, Carew, Therese, & Blundell's Assessment	(Hadgraft et al., 2008)
Vanasupa, Rogers, & Chen's Assessment	(Vanasupa et al., 2008)
Zoller & Scholz Example 2 Assessment Zoller & Scholz Example 4 Assessment	(Zoller & Scholz, 2004)
Booth Sweeney & Sterman's Assessment	(Booth Sweeney & Sterman, 2000)

Assessments are listed from most recent publication to oldest based on the most recent publication date of each assessment's source paper(s).

TABLE 3
NAME, PARTICIPANT EDUCATION LEVEL AND DISCIPLINE OF THE 27 UNIQUE ASSESSMENTS IDENTIFIED IN THIS SLR

Assessment	**Education Level**	**Study Participant Discipline in Source Paper(s)**
Systems Thinking Assessment Rubric (STAR)	Professional Postsecondary	Engineering Engineering, Engineering & Management dual degree
Taylor, Calvo-Amodio, and Well's Assessment	High School Pre-High School	Extracurricular science & math program
Bedir, Desai, Kulkarni et al.'s Assessment	Postsecondary	Arts and Sciences, Engineering and Applied Science
Mystery Maps	High School	Unknown
Gray, Sterling, Aminpour et al.'s Assessment	Postsecondary	STEM and unspecified other majors
Jaradat, Hamilton, Dayarathna et al.'s Assessment	Postsecondary	Engineering
Timofte & Popuș' Assessment	Unknown	Organic Chemistry
Grohs, Kirk, Soledad, & Knight's Assessment	Postsecondary	Engineering
Frank & Kordova's Assessment B	Professional	Engineering
Engineering Systems Thinking Survey (ESTS)	Postsecondary	Engineering
Camelia, Ferris, & Cropley's Assessment	Postsecondary	Engineering
Hu & Shealy's Assessment A Hu & Shealy's Assessment B	Postsecondary	Engineering
Climate Change System Thinking Instrument (CCSTI)	Postsecondary	Biology Education, Physics Education, and Chemistry Education
Systems Assessment Test (SysTest)	Postsecondary	Engineering
Hrin, Milenković, Segedinac, & Horvat's Assessment	High School	Organic Chemistry
Jaradat & Castelle's Assessment	Professionals	Cyber security, Engineering, Management and unspecified others
Frank & Kordova's Assessment A	Professional Postsecondary High School	Engineering Engineering, Technology/Engineering Management N/A
Keynan, Ben-Zvi Assaraf, & Goldman's Assessment	High School	Extracurricular science program
Brandstädter, Harms, & Großschedl's Assessment A Brandstädter, Harms, & Großschedl's Assessment B	Pre-High School	Science, Biology
Rehmann, Rover, & Laingen et al.'s Assessment	Postsecondary	Engineering
Hadgraf, Carew, Therese, & Blundell's Assessment	Postsecondary	Engineering
Vanasupa, Rogers, & Chen's Assessment	Postsecondary	Engineering
Zoller & Scholz Example 2 Assessment	Postsecondary	Science and Science Education
Zoller & Scholz Example 4 Assessment	Postsecondary	Environmental Science
Booth Sweeney & Sterman's Assessment	Postsecondary	Management, Engineering & Management dual degree

Assessments are listed from most recent publication to oldest based the most recent publication date of each assessment's source paper(s).

TABLE 4
FORMAT AND CONTENT AREA OF ASSESSMENTS BY ASSESSMENT TYPE

Assessment(s) Name(s)	Assessment Format(s)	Format Group	Assessment Content Area
Behavior-based			
Systems Thinking Assessment Rubric (STAR)	Conceptual Models following Object-Process Methodology (OPM)	M, E	"Freely available, consumer-focused Web-based information systems" (Lavi, Dori, Wengrowicz, et al. 2020, p. 40); "an authentic design problem" (Lavi, Dori, & Dori, 2020, p. 3)
Hu & Shealy's Assessment A	Concept Mapping while wearing fNIRS cap	M, E	"Sustainability topics about energy, food, climate, and water" (sec. Abstract)
Mystery Maps	Mystery Method combined with Influence Diagrams	M, S	"Water-intensive, export-oriented tomato cultivation in Almería, Spain" (p. 2)
Gray, Sterling, Aminpour et al.'s Assessment	Cognitive Mapping, Fuzzy Cognitive Mapping, & Student Essay	M, E	Scientific & popular articles
Vanasupa, Rogers, & Chen's Assessment	Rich Pictures	M, E	"Successful" & "unsuccessful" engineering student
Taylor, Calvo-Amodio, & Well's Assessment	Scenario-based drawing (Draw fish-tank system)	M, S, E	"You recently purchased a fish tank. After two weeks, you notice the water is turning green in color." (p. 10)
Brandstädter, Harms, & Großschedl's Assessment A	Concept Mapping	M, S, E*	"Development, enemies, living, and feeding of eggs, larvae, young and adult blue mussels" (p. 2151)
Keynan, Ben-Zvi Assaraf, & Goldman's Assessment	Structured Interviews (Repertory Grid)	O	"Local ecological system (Shezaf Nature Reserve in the Arava Valley)" (p. 93)
Timofte & Popuș' Assessment	Systemic Assessment Questions (SAQs)	M, F	Organic chemistry functional group manipulations
Hrin, Milenković, Segedinac, & Horvat's Assessment	Systemic Synthesis Questions [SSynQs]	M, F	Organic chemistry
Booth Sweeney & Sterman's Assessment	Graphs of Expected Behavior Over Time	M, S, F	Bath Tub/Cash Flow Tasks & Manufacturing Case. See full text for complete tasks.
Rehmann, Rover, & Laingen et al.'s Assessment	Oral Project Presentations (Sophomore Seminar)	M, O, E	"Two of the projects dealt with renewable energy, while the others focused on safe roads, sustainable agriculture, protection from disasters, and clean water." (sec. Observations – Sophomore Seminar)
Zoller & Scholz Example 4 Assessment	4 hour written exercise "that very much resembles Example 2" and a 20 min oral exam	O, E	"Aquatic systems, terrestrial systems, atmosphere, or human-environment systems." (p. 34)
Grohs, Kirk, Soledad, & Knight's Assessment	Community Level Problem Scenario	S, E	Heating expenses and heating use in the Village of Abeesee. See full text for a paragraph of the scenario.
Systems Assessment Test (SysTest)	Abstract Problem Statement	S, E	"Automated system for lawn debris (such as tree leaves) collection" See full text p. 182 for the full scenario.
Bedir, Desai, Kulkarni et al.'s Assessment	5 Written Response Questions	S**, E	Systems engineering objectives and metrics
Engineering Systems Thinking Survey (ESTS)**	2nd section: 12 sets of contextual multiple-choice questions	C	Contextual questions - "provided for context a product or system that was thought to be familiar to most engineering students, such as a computer" (p. 3)
Climate Change System Thinking Instrument (CCSTI)	37 Multiple Choice Questions	S, C	Climate change
Brandstädter, Harms, & Großschedl's Assessment B	Procedural and structural system thinking questionnaire	S, C	Blue mussels
Preference-based			
Frank & Kordova's Assessment A	Interest Inventory with 40 pairs of statements	C	Engineering workplace or "context that high school and college students could relate to" depending on participant's education level (Frank, 2010, p. 171)
Frank & Kordova's Assessment B	Questionnaire with 40 items *(13 of the 19 that overlap with Frank* (Frank, 2010)*, are examining "systems thinking capability")*	C	Not Available
Camelia, Ferris, & Cropley's Assessment	Questionnaire with 16 items *(original 30 had about 60% overlap with Frank & Kordova's Assessment A)*	C	Camelia & Ferris (Camelia & Ferris, 2018, p. 578) note that Frank's (Frank, 2010) instrument "was modified to suit undergraduate engineering students"
Jaradat & Castelle's Assessment	Instrument with 39 binary questions	S, C	"Large scale export management company that ships a variety of goods and services worldwide" (Jaradat, 2014, p. 261)
Jaradat, Hamilton, Dayarathna et al.'s Assessment	Virtual reality (VR) gaming scenario	S, V	Retail store
Self-reported			
Engineering Systems Thinking Survey (ESTS)†	1st section: 37 Likert-scale questions	C	Likert-scale questions - Not Available
Hadgraft, Carew, Therese, & Blundell's Assessment	Survey instrument with 6 questions	E, C	"Lived experience of the individual" (p. 2)
Cognitive Activation			
Hu & Shealy's Assessment B	Concept Mapping while wearing fNIRS cap	N	"Sustainability topics about energy, food, climate, and water" (sec. Abstract)
Assessment Type Not Discernable			
Zoller & Scholz Example 2 Assessment	Scenario with 7 open-ended questions	S, E	"Resources and energy" (p. 32)

Format Group Key: M = Mapping (terms defined in Appendix B); S = Scenario; E = Open-ended; O = Oral; F = Fill-in-the-blank; C = Multiple-choice; V = Virtual Reality; N = fNIRS Cap. *Brandstädter, Harms, & Großschedl's Assessment A looked at "highly-directed" and "nondirected" concept maps; only the maps created with the non-directed method belong to the open-ended grouping (Brandstädter et al., 2012). **Bedir, Desai, Kulkarni et al.'s Assessment had one question that asks for a response "based on the scenario provided." †ESTS is listed under two assessment types because it has two sections.

TABLE 5

SUMMARY OF GROUP 1 EVALUATION CRITERIA, DIMENSION OF ST ASSESSED, AND DEFINITIONS OF ST

Assessment	Evaluation Criteria	Dimensions of ST Assessed	Systems Thinking (ST) Definition
Bedir, Desai, Kulkarni et al.'s Assessment	Score of 0 (naïve level), 1 (apprentice level), or 2 (competent level)	Not Available	"Systems thinking begins with formulating well-conceived objectives and metrics that track how well those objectives are achieved" (sec. Introduction)
Gray, Sterling, Aminpour et al.'s Assessment	Independent ranking by faculty as high, medium, or low ST; No formal criteria provided to guide ranking	Not Available	"We then suggest four fundamental dimensions of ST that provide a framework for understanding degrees of ST, which include evaluating student understanding of: (1) system structure, (2) system function, (3) identification and negotiation of leverage points for change, and (4) trade-off analysis" (p. 1)
Timofte & Popuș' Assessment	Not Available	Not Available	Arnold and Wade (2017) defined ST as "the ability to think holistically, to observe the non-obvious connections between the parts of the system, and to understand why parts of a system act in the way they act." (p. 253)
Frank & Kordova's Assessment B	Participant rates extent of agreement or disagreement with items on a scale of 1-5; responses are averaged for relevant items	Not Available	"Systems thinking - a field that deals with seeing the system as a whole and examining the processes that occur within it and its surrounding environment." (p. 16)
Engineering Systems Thinking Survey (ESTS)	Not Available	Not Available	**Not Available** ST and Systems Engineering (SE) skills mentioned but no explicit definition of ST presented.
Hu & Shealy's Assessment B	"Global efficiency (E) of connectivity was measured, which describes the cognitive effort to transfer information between brain regions.	"complexities and comprehensiveness of [ST]"	"Systems thinking is a necessary skill towards solving complex civil engineering problems with interconnected environmental, social, and economic inputs and outputs." (sec. Abstract)
Vanasupa, Rogers, & Chen's Assessment	Not Applicable	Not Applicable	"Systems thinking requires seeing the whole." (sec. Introduction)
Zoller & Scholz Example 2 Assessment	Not Available	Not Available	**Not Available** "Higher-order cognitive skills (HOCS) Capability; i.e. question-asking, critical system thinking, decision making and problem solving" (p. 27)
Zoller & Scholz Example 4 Assessment	Not Available	Not Available	

Assessments are listed from most recent publication to oldest based the most recent publication date of each assessment's source paper(s).

TABLE 6

SUMMARY OF GROUP 2 EVALUATION CRITERIA, DIMENSION OF ST ASSESSED, AND DEFINITIONS OF ST

Assessment	Evaluation Criteria	Dimensions of ST Assessed	Systems Thinking (ST) Definition
Frank & Kordova's Assessment A	Participant selects "A", "B", or "C" depending if they prefer the 1st, 2nd, or neither statement; 2 points for selecting the systems thinking answer and 1 point for selecting the other statement	For example, part "A" of one question is "When I take care of a product, it is important for me to see how it functions as part of the system." See full text for additional sample items, but note that the full inventory is copyrighted.	Frank (Frank, 2006) defined engineering ST as "the ability to: 1. See the big picture . . ., 2. Implement managerial consideration . . ., 3. Acquire and use interdisciplinary knowledge . . ., 4. Analyze the needs/requirement . . . 5. Be a systems thinker." (Frank, 2010, p. 170-171)
Camelia, Ferris, & Cropley's Assessment	7 point Likert scale for each item; scores range from 0 (very low) to 7 (very high) for each item	For example, one question is how much you agree with the statement "I like to be bold and take risks." See full text for all 30 questions.	**Not Explicit** ST is having a holistic understanding of a system that is located in a specific environment.
Jaradat & Castelle's Assessment	Responses lead to a profile with 7 letters (each letter corresponds to a preference dimension) and can be converted to a label of Reduction, Middle, High, or High-Holistic Systems Thinker See full text for scoring sheet (Jaradat, 2014, p. 160)	7 preference pairs: - Complexity vs. Simplicity - Integration vs. Autonomy - Interconnectivity vs. Insolation - Holism vs. Reductionism - Emergence vs. Stability - Flexibility vs. Rigidity - Embracement of Requirements vs. Resistance of Requirements	ST "can provide a holistic thinking paradigm that opens new channels and opportunities to think differently about complex systems as a whole unit."(Castelle & Jaradat, 2016, p. 80) "The perspective taken for systems thinking characteristics for this research is taken as the set of abilities, preferences and skills characteristics that individuals exhibit in dealing with a complex problem domain" (Jaradat, 2014, p. 14)
Jaradat, Hamilton, Dayarathna et al.'s Assessment	Scores range from 0 to 6; VR scenarios based on 6 questions from Jaradat & Castelle's Assessment. See full text for the breakdown of scene vs. ST measurements.	1 preference pair: - Complexity vs. Simplicity	"Systems thinking (ST) is considered an active framework to better manage complex system problem domains. It focuses on how the constituent parts of a system pertain to the whole system and the way the systems work within larger systems over time. This approach contrasts with traditional analysis whose aim is to study the individual pieces of a system separately." (sec. Introduction)
Hadgraft, Carew, Therese, & Blundell's Assessment	Not Applicable	See full text for list of 14 systems thinking skills.	"Systems thinking is touted as a core engineering competence. It is a meta-attribute with value in all engineering disciplines and many non-engineering disciplines as well." (p. 2)

Assessments are listed from most recent publication to oldest based the most recent publication date of each assessment's source paper(s).

TABLE 7
SUMMARY OF GROUP 3 EVALUATION CRITERIA, DIMENSION OF ST ASSESSED, AND DEFINITIONS OF ST

Assessment	Evaluation Criteria	Dimensions (Dim) of ST Assessed	Dim Category	Systems Thinking (ST) Definition
Systems Thinking Assessment Rubric (STAR)	Score of 0 (no expression of attribute understanding), 1, 2, or 3 (full expression of attribute understanding)	For example, one attribute (of nine) is "Complexity levels - Number of levels of detail; refinement of main functions into sub-levels." See source papers for more detail.	E, R, L, T, O	"In the context of technological, engineered systems, it [ST] can be considered as holistic understanding of the system's function, structure and behavior, and how the latter two interact to deliver the former (E. Crawley et al., 2016; Dori, 2016)" (Lavi, Dori, Wengrowicz, et al., 2020, p. 40).
Taylor, Calvo-Amodio, & Well's Assessment	Rubric provided to classify elements, relationships, and roles/purposes with respect to 3 learning levels: 1 – sensibility, 2 – literacy, & 3 - capability	For example, elements classified as "concrete, internal, essential elements" are at the learning level of sensibility or "awareness of systems"	E, R, O	"Systems thinking is comprised of four underlying concepts or skills: distinction-making, organizing systems, inter-relating, and perspective taking." (p. 1)
Mystery Maps	7 evaluation schemes described across 3 reference types	For example, a complete reference is one "containing all the connections for which any number of experts indicated that a direct causal link exists." See full text for more detail.	R	**Not Explicit** ST is the ability to identify key system elements and the interrelationships between these elements.
Grohs, Kirk, Soledad, & Knight's Assessment	Rubric provided with scores for each of 7 constructs ranging from 0 to 3	Problem Identification, Information Needs, Stakeholder Awareness, Goals, Unintended Consequences, Implementation Challenges, Alignment	E, R, T, B, S, O	"Such situations call for a metacognitive strategy - a flexible way of framing, reasoning, and acting within multiple dimensions, which we conceptualize as 'systems thinking.'" (p. 111)
Hu & Shealy's Assessment A	Watson, Pelkey, Noyes, & Rogers' (2016) Concept Map Scoring Methods: Traditional, Holistic, & Categorical Scoring	Traditional scoring method accounts for the number of concepts, highest level of hierarchy, and the number of cross links. See full text for more detail.	E, R F, H, T, B, S	"Systems thinking is a necessary skill towards solving complex civil engineering problems with interconnected environmental, social, and economic inputs and outputs." (sec. Abstract)
Climate Change System Thinking Instrument (CCSTI)	4 Levels of ST indicators from I "pre-requirement" to IV "coherent expert" are defined and an example multiple choice question is shown for each level	For example: Level II c. "Able to identify process of feedback which happens to the system." See full text for more detail.	E, R, F, L, T	"The framework of system thinking from Boersma (Boersma, Waarlo, & Klaassen, 2011) is the one that is developed into an indicator of system thinking in this research." (p. 3)
Systems Assessment Test (SysTest)	Responses categorized by use of techniques and yes or no (1 or 0) question responses	Techniques included: Functional Model, Black box, and Pugh Chart. An example of a yes or no question is "Follow design process?" See full text for more detail.	O	"Systems thinking is seeing the interactions and relationships that reinforce the system as a whole." (p. 180)
Hrin, Milenković, Segedinac, & Horvat's Assessment	Rubric provided with scores ranging 0 to 4	A score of 4 corresponds to "All concepts and subsystems are interconnected, constituting a meaningful whole."	E, R, O	**Not Explicit** ST involves identifying system elements and the interrelationships between these elements, understanding emergent outcomes, and analyzing outcomes in a broader context.
Keynan, Ben-Zvi Assaraf, & Goldman's Assessment	Students are compared based on their "expression of constructs" according the Ben-Zvi Assaraf & Orion's (2005) Systems Thinking Hierarchy (STH) Model	For example, "Identifying the components and process of a system (level A)." See full text for more detail.	E, R, T	"An exploration of learners' system thinking capacities should be based on a theoretical framework… One such framework is the Systems Thinking Hierarchy (STH) model developed by Ben-Zvi Assaraf and Orion (2005)" (p. 92)
Brandstädter, Harms, & Großschedl's Assessment A	McClure, Sonak, and Suen's (1999) Relational Scoring Method. See full text for adapted scoring protocol.	Correctness of concept map propositions, which consist of two concepts, a linking word, and an arrow.	E, R, T	"Structural system thinking is the ability to ability to identify a system's relevant elements and their interrelationships […] Procedural system thinking is the ability to understand the dynamic and time-related processes that emerge from the systems' structure, particularly occurring in within systems' elements and subsystems." (p. 2148)
Brandstädter, Harms, & Großschedl's Assessment B	Procedural - sum of 6 items; Structural - sum of 13 items	For example, a structural system thinking question was "How are Blue Mussels able to stick together?" See full text appendices for questionnaire questions.	E, R, T	
Rehmann, Rover, & Laingen et al.'s Assessment	Rubric provided with scores for each of the 7 aspects of technical content ranging from "0 = not addressed" to "4 = well addressed"	Problem description, Key variables, Rich pictures to show connections, Causal-loop diagrams to show relationships, Graphs to show behavior over time, Lessons learned, Sources	E, R, T, B	**Not Explicit** ST is taking a holistic view to identify factors, explain connections, and understand dynamic behavior, where such factors may be "from inside and outside of engineering."
Booth Sweeney & Sterman's Assessment	Multiple performance criterion provided for each task / case	For example: "The stock should not show any discontinuous jumps (it is continuous)." See full text for more detail.	R, F, T	"Most advocates of systems thinking agree that much of the art of systems thinking involves the ability to represent and assess dynamic complexity . . .both textually and graphically." (p. 250) See full text for description of ST skills.

Dimension Code Key: E = Elements; R = Relationships; F = Feedback; L = Levels; T = Time; B = Breadth; S = Stakeholders; O = Other;
Assessments are listed from most recent publication to oldest based the most recent publication date of each assessment's source paper(s).

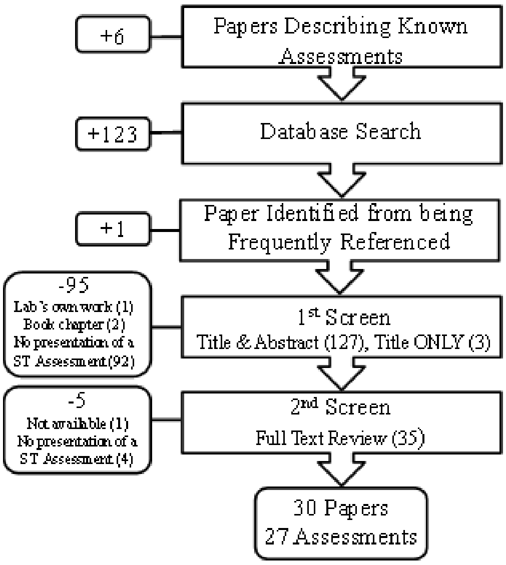

FIGURE 1 SLR Search Process, Screening Process, and Results.

SRES_2808_Figure 1.tif

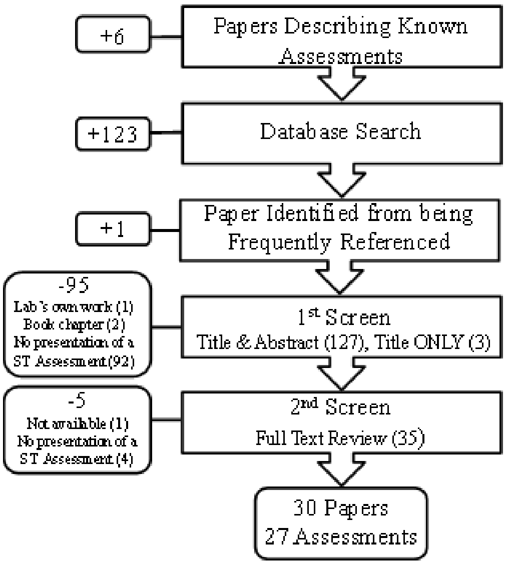

FIGURE 1 SLR Search Process, Screening Process, and Results.

SRES_2808_Figure 1.tif