International Journal *of* Designs for Learning
2024 | Volume 15, Issue 1 | Pages 27-37

DESIGN CONSIDERATIONS FOR A MULTIPLE-CHOICE ASSESSMENT OF SOCIO-SCIENTIFIC SYSTEMS THINKING

Eric A. Kirk[1], Troy D. Sadler[1], Li Ke[2] & Laura A. Zangori[3]
[1]*University of North Carolina at Chapel Hill;* [2]*University of Nevada–Reno;* [3]*University of Missouri—Columbia*

This design case details the design process of a multiple-choice assessment of socio-scientific systems thinking. This assessment is situated within a larger project that aims to understand the ways students use multiple scientific models to understand complex socio-scientific issues. In addition to the research component, this project entails developing curriculum and assessment resources that support science teaching and learning. We begin this paper by framing the needs that motivated the design of this assessment and introducing the design team. We then present a narrative outlining the design process, focusing on key challenges that arose and the ways these challenges influenced our final design. We conclude this paper with a discussion of the compromises that had to be made in the process of designing this instrument.

Eric A. Kirk is a doctoral student in the Learning Sciences and Psychological Studies program at the University of North Carolina at Chapel Hill. His work focuses on how students use models to make sense of complex systems implicated in societal issues.

Troy D. Sadler is the Thomas James Distinguished Professor of Experiential Learning at the University of North Carolina at Chapel Hill. His work focuses on how to support students in negotiating socioscientific issues.

Li Ke is an Assistant Professor of Science Education at the University of Nevada, Reno. His work focuses on supporting learners in using models and engaging in systems thinking, particularly with regards to socioscientific issues.

Laura A. Zangori is an Associate Professor of Science Education at the University of Missouri, Columbia. Her work focuses on supporting students in using models to support scientific reasoning about complex systems.

INTRODUCTION

The COVID-19 pandemic is a classic example of a socio-scientific issue (SSI): a societal dilemma that is entangled with scientific practices and scientific knowledge (Zeidler, 2014). As partisan debates about masking, business closures, and vaccinations embroil our population, it has become abundantly clear that there is a need to better support public understanding of the interactions between science and society and better develop our populations' ability to discern non-immediate consequences of actions (and inaction). Decision-making in socio-scientific issues is often a high-stakes affair. Conflicting interests between diverse stakeholders preclude straightforward solutions based on simple, linear cause-effect reasoning. Individuals must recognize the inherent complexity of these issues and consider the complex interactions between the entangled components of the systems they operate within should they wish to predict behaviors, and design solutions that minimize unintended consequences (Sadler et al., 2007).

In this paper, we present the design process of an instrument that assesses skills associated with systems thinking in the context of COVID-19. The resulting instrument contains 19 multiple-choice items and is intended to be used as a diagnostic tool to help the authors understand the cognitive processes involved in understanding models of complex systems. After introducing the design team, we provide an overview of the design motivation and a high-level summary of the design process. Next, we elaborate on specific design challenges, the ways we addressed those challenges in our design, and the rationale behind our design choices. We

https://doi.org/10.14434/ijdl.v15i1.34204

CHALLENGE	DESIGN FEATURES ADDRESSING CHALLENGES
Selecting an intervention	• A focus on system models
Framing how we conceptualize systems thinking	• A focus on domain-general skills • Mehren et al. (2018) assessment framework
Establishing a test structure	• Fewer than 20 items • Multiple-choice items • Delivery through Qualtrics
Translating Ecosystems to COVID-19 Systems	• Focusing on identifying causal relationships • Prioritizing readability • Identifying Relationship Directionality
Accounting for variations in prior knowledge. Adaptable for future pandemics.	• Embedded content supports. • Factors likely to impact future pandemics

TABLE 1. Summary of Challenges and Associated Design Features,

conclude this paper with a discussion of the compromises that shaped the final design.

DESIGN TEAM

This project was carried out by a team consisting of four members as part of a larger project aimed at better understanding how students use models to understand respiratory viruses. At the time this assessment was developed, Eric Kirk was a first-year doctoral student pursuing a degree in the Learning Sciences with a focus on science instruction using socio-scientific issues. Prior to beginning his degree, he served as a research assistant on a project studying how students use modeling to understand complex biological systems. His recent experiences teaching high school environmental science and biology during the COVID-19 pandemic informed how he approached this design task. Troy Sadler has been teaching and conducting research on socioscientific issues for 20 years. In this work, he has conceptualized the issues as complex systems and has recognized student negotiation of these issues as a form of systems thinking. Li Ke has been working with K-12 teachers and students to promote systems thinking through modeling for 10 years. His recent work focuses on socio-scientific issues as a larger system that includes both scientific and social dimensions, and how students could use modeling to better understand the complexity of the underlying issue. Laura Zangori has been working with K-16 students in teaching and research on biological systems for 15 years. In her work, she uses the practices of modeling to support students in making their systems understanding explicit so they can use their modeling tools as cognitive aids for reasoning scientifically about system-specific phenomena.

DESIGN PROCESS

Throughout the design process, there were several key challenges that emerged as we worked to balance the needs of the research program and the practical concerns of doing classroom research. This section elaborates on these challenges as well as the design decisions made by our team to overcome them. Furthermore, this section illustrates why the choice of the assessment framework described was best suited to our needs. For a summary of challenges and design features, see Table 1.

Challenge: Selecting an Intervention

As a team, our mission is to develop tools that allow educators and researchers to better understand how students navigate complex and contentious SSI. If we wish to support our students' ability to face these challenges, it is important we develop the skills that support thinking about these issues in productive ways. Before working to better understand these processes, our teams' first task was to identify what resources we felt were most promising as supports for students so that we could tailor our research to understanding the specific processes involved in using those resources.

Scientific models are one way of supporting students that is of particular interest to our team. Ke et al. (2021) advocate for the increased use of modeling in SSI-based curricula as one way of addressing these concerns. Using the term "socio-scientific models" to refer to models that account for both scientific and social factors, the authors argue that these models have the potential to be particularly useful for students negotiating complex societal issues by helping students draw connections between scientific knowledge and relevant social dimensions while considering possible solutions to these issues.

What sets socio-scientific models apart from traditional scientific models is their interdisciplinarity. The complex and interdisciplinary nature of socio-scientific issues necessitates an approach that emphasizes the relationships and interconnectedness of the many facets of these issues. Whereas

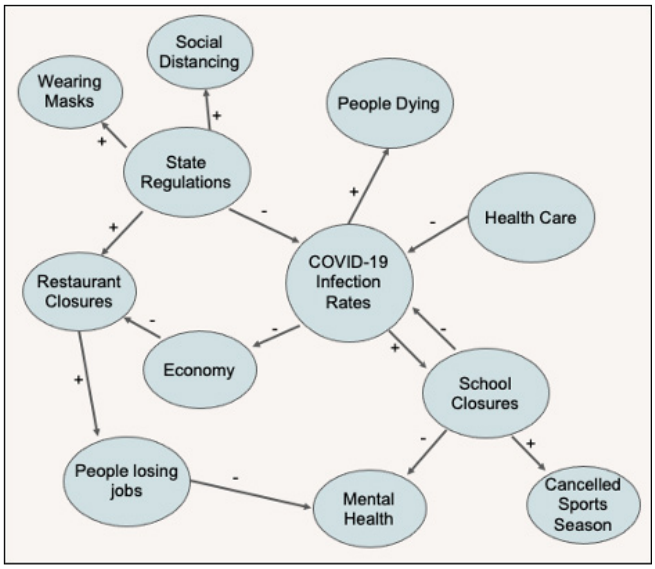


FIGURE 1. Example of a COVID-19 Pandemic Socio-Scientific System Model presented in the accompanying curricular materials. From COVID-19 Curricular Materials by Sadler et al., 2021, https://tarheels.live/seel/teaching-materials/curriculum/covid-19-teacher-materials/. Copyright 2021. Reprinted with permission.

scientific models only seek to explain scientific phenomena and rely on scientific evidence, socio-scientific models incorporate knowledge from social domains such as the economic, historical, or political dimensions tied to a phenomenon. For example, a scientific model that represents a fishery collapse would focus on the unfolding ecosystem dynamics (e.g., predator/prey relations and water quality measures). A socio-scientific model may expand upon the scientific model by also incorporating the economic impact on the local fishing industry as well as relevant laws and regulations that dictate how many fish may be harvested, illustrating how these factors and the ecosystem dynamics shape one another.

Design Consideration: A Focus on System Models

Although socio-scientific models can take many forms, we focused our attention on socio-scientific system models (referred to as system models throughout the remainder of this article). In these models, important system components are represented within a labeled circle, and relationships between factors are conveyed using arrows running between two or more interrelated factors (see Figure 1). Creating these models provides students with the opportunity to deconstruct complex organized systems to their basic components, supporting their ability to discover causal mechanisms for system-level behaviors (Bechtel & Abrahamsen, 2005). Socio-scientific system models differ slightly from other system models traditionally encountered in science classes like biogeochemical cycles and food webs, however. These models explicitly highlight the relationship between the scientific and social factors of the issue in question.

Our team feels these models are particularly promising for several reasons. First, teachers readily rely on traditional, scientific system models to help students understand complex phenomena encountered in science education. Although these models rarely include societal factors, the general approach is likely to be familiar to both teachers and students, leaving fewer obstacles to implementing these models in classrooms. Students and teachers can focus their efforts on understanding the behaviors of a complicated system rather than learning a technique from scratch.

Second, these models naturally align with key aspects of the United States' *Next Generation Science Standards* (NGSS Lead States, 2013). These standards underscore the importance of engaging students in the practice of constructing, revising, using, and critiquing models in science instruction. In representing the components that make up a system and the relationships that exist between these components, these models also address cause and effect, and systems and system models: two crosscutting concepts named by these standards. Focusing our efforts on models that are aligned with national standards documents increases the likelihood of these models being embraced by instructors.

Finally, these models target a key challenge to navigating socio-scientific issues: complexity. Students often struggle to identify complex causal relationships such as domino causality, feedback effects, and non-obvious causes (Grotzer, 2012; Grotzer & Tutwiler, 2014), all of which can be immensely consequential to the behavior of a complex socio-scientific issue. By explicitly identifying and representing the causal relationships of a complex system, students may be in a better position to navigate the complexity of these issues as they work to understand system behaviors, predict system changes, or design interventions to achieve a desired outcome.

Challenge: Framing How We Conceptualize Systems Thinking

After arriving at system models as our intervention of interest, we turned our attention to understanding the skills needed to use these models in ways that are likely to occur in the classroom. For students to use system models to navigate SSI such as the COVID-19 pandemic, students must possess systems thinking skills: skills that support an ability to understand and interpret complex systems (Evagorou et al., 2009). These systems thinking skills are applicable across different contexts, functioning as scaffolds that support student thinking about the specific context being investigated (Yoon, 2018). These systems thinking skills align with our overarching goals as a research team. Systems thinking skills allow students to consider solutions to complex problems in ways that may not be possible when relying on simple, linear

causal reasoning, which can help minimize the likelihood of unpredicted or unwanted outcomes by considering the problem holistically (Mehren et al., 2018).

Even though complex systems are regularly found in scientific and social settings the skills needed to interpret these systems have been shown to be incredibly difficult to develop. Often, doing so requires making significant changes to personal epistemologies and ontologies as well as the schema used to understand causation (Grotzer, 2012; Jacobson & Wilensky, 2006; Wilensky & Jacobson, 2014).

Because of the research-focused nature of our assessment, it was important to establish a sturdy theoretical foundation for our research. To address this, we turned to the robust and rapidly growing body of research on systems thinking in science education (Yoon et al., 2018). During our review, we found many potential frameworks that could be used to structure our assessment. As a result, it became clear that we needed to identify specific aspects of systems thinking that we hoped to understand so that we could narrow our focus enough to select a framework.

Design Consideration: A Focus on Domain General Skills

Systems thinking skills can be classified as domain-general or domain-specific depending on whether they can be transferred across contexts. Although researchers have identified many domain-general skills (cf., Ben-Zvi Assaraf & Orion, 2010; Mehren et al., 2018), these skills find common ground around the ability to identify components and processes that constitute a system, to understand dynamic relationships among the components within the system, and to organize these components into a usable framework to explain and predict behavior (Yoon, 2018).

Focusing on domain-general skills such as identifying structural features of a system or predicting system behaviors increases the versatility of an assessment, allowing these assessments to be adapted to future contexts through the modification of domain-specific details without requiring extensive modification of the deep structure of the assessment. The project this assessment was developed for was framed around the goal of supporting student learning about global pandemics caused by respiratory viruses such as COVID-19 through the design of curricular materials (Sadler et al, 2021) and research. Although COVID-19 is the specific anchoring phenomenon, we sought to develop materials that could be adapted and used during future viral pandemics. Should another pandemic arise, we want practitioners and researchers to have access to ready-made curricula and instruments that support the teaching and learning of these topics. Thus, it was important to design an assessment that was flexible enough to be adapted to future contexts with minimal effort and minimal threat to item integrity.

Design Consideration: Mehren et al.'s (2018) Assessment Framework

Having made the decision to focus on domain-general skills, we began the process of evaluating several potential frameworks as potential frameworks for our assessment. Although there have been several instruments developed in recent years to assess students' ability to understand complex systems (e.g., Grotzer et al., 2016; Mehren et al., 2018), the assessment developed by Mambrey et al. (2020) most closely aligned with our intended design goals. This assessment focuses on three domain-general skills identified by Mehren et al. (2018):

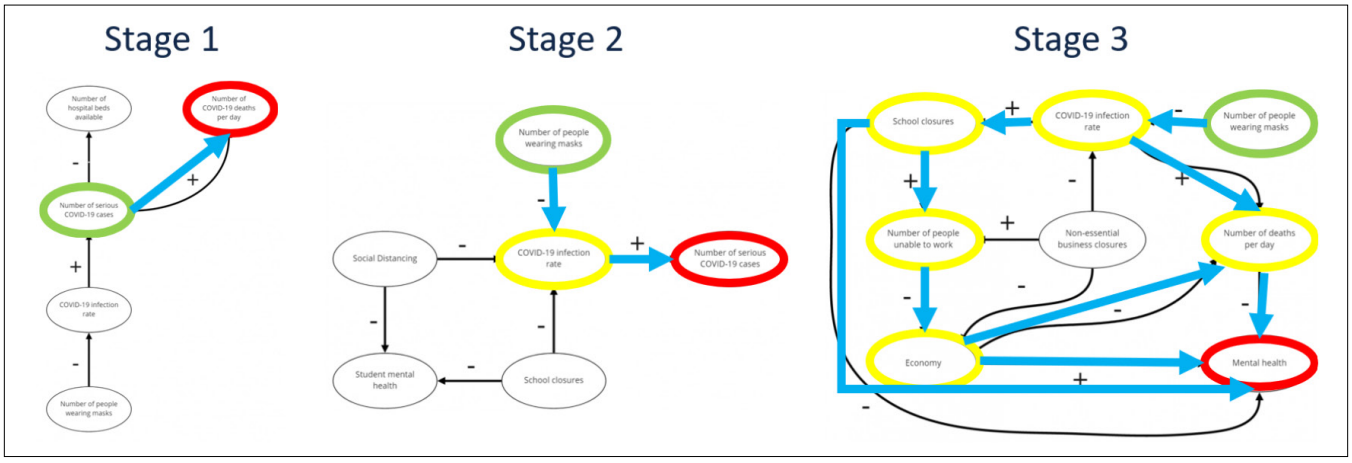


FIGURE 2. Examples of Systems and Relationships Across Competence Stages. Green and red circles denote two factors of interest that students would be asked to consider the relationship between. Yellow circles represent intermediate factors that must be understood to predict the relationship between the two factors of interest. Blue arrows denote pathways that students must consider when considering the relationship between the two factors of interest. Note that there are no intermediate factors in Stage 1. Stage two may contain intermediate factors, but the factors of interest are connected in a non-branching pathway. Stage 3 contains multiple intermediate factors as well as multiple branching pathways between factors of interest

- System organization (SysOrg)—identifying the components of the system in question and understanding how those components are organized in relation to one another.
- System behavior (SysBeh)—understanding how systems behave when a system component is modified.
- System-adequate intention to act (referred to as system modeling in this paper, SysMod)—proposing manipulations to a system to achieve a desired outcome.

Originally, the skills in this framework were assessed using qualitative and quantitative items in the context of geography, featuring systems that include both social and scientific factors. These skills closely resemble those identified by Ben-Zvi Assaraf and Orion (2005; 2010) in their hierarchical conception of systems thinking skills, providing further empirical support for the selection of this framework.

Although like Ben-Zvi Assaraf and Orion's (2005; 2010) model of systems thinking, Mehren et al. (2018) extend their model beyond simply accounting for skills, specifying three stages of competence within each skill. Given our interest in tracking and supporting students' skill development, having a framework that explicitly outlines a progression of competence stages was particularly appealing. In this framework, students progress through competence stages as they demonstrate their ability to interpret increasingly complex models and answer increasingly complex questions.

Mehren et al. (2018) identify 3 features of systems that add to their complexity: the number of system components, the number of connections between system components, and the ways in which system components are connected. A structure index calculation is used to calculate a system's structural complexity (Mehren et al., 2015). Guided by this calculation, structural complexity increases as students progress throughout the stages, with Stage 1 featuring the simplest systems, and Stage 3 featuring the most complex.

As students progress through these stages, they are also able to answer questions featuring increasingly complex causal relationships. In Stage 1, students answer questions about simple, direct relationships (X influences Y). For Stage 2, students evaluate more complex systems that present non-branching but indirect relationships (X influences Y, and Y influences Z, therefore X influences Z). Finally, students encounter the most complex systems in Stage 3; they analyze complex, indirect relationships. These relationships differ from those featured in Stage 2 in that there are multiple pathways between two factors that must be considered rather than one direct path (e.g., W influences X and Y, X influences Y and Z, Y influences Z). Examples of systems and relationships used in this assessment can be seen in Figure 2.

Challenge: Establishing a Test Structure

Whereas many assessments of systems thinking have relied on qualitative methods, we worried that this approach would not suit our goals as researchers. Our instrument was developed to be used for a large-scale, multi-year study on the use and integration of multiple models in collaboration with several high school teachers. We also intended this assessment to be administered to large sample sizes during instructional time. Because of this, it was important to design the assessment in such a way that it could be administered easily by teachers and scored rapidly. Additionally, it was important to protect teachers' instructional time; something teachers often express concerns about given the amount of time it takes to implement an SSI unit (Ekborg et al., 2013; Tidemand & Nielsen, 2017).

Design Consideration: 20 Multiple Choice Items Delivered Through Qualtrics

To address these challenges, we made the decision to use only multiple-choice, single-select questions for our assessment. We also decided to limit the length of the test to 20 items or less. These decisions address the aforementioned concerns in several ways. First, multiple-choice items are easily adapted to online formats such as Qualtrics, allowing the test to be rapidly disseminated and collected by researchers without unnecessary time investment by teachers, thus protecting our partner teachers' instructional time. Imposing a limit of 20 items also supported the goal of protecting instructional time, ensuring that the amount of class time taken to administer the test is kept to a minimum. Similarly, by delivering multiple-choice items through Qualtrics, researchers and teachers alike are saved from the process of manually scoring items. This provides teachers with rapid feedback on their students' performance while also expanding our capacity to work with larger sample sizes.

Whereas Mehren et al. (2018) originally used qualitative items to assess systems thinking with this framework, Mambrey et al. (2020) used that same framework to design an assessment to evaluate ecology systems thinking skills in German 5th and 6th-grade biology students using only multiple-choice items. Despite this assessment having been designed with a different population and anchoring phenomenon in mind, it demonstrates that it is indeed possible to assess systems thinking skills through a multiple-choice assessment using the framework described above. The entirely multiple-choice design of Mambrey et al.'s (2020) assessment balances the theoretical affordances of Mehren et al.'s (2018) framework with the practical concerns related to teachers' time and having the capacity to evaluate large samples of students.

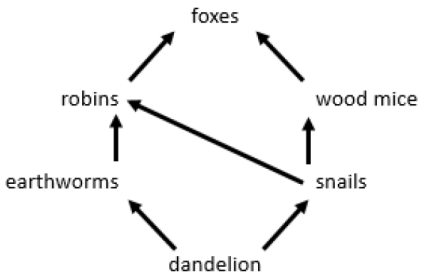
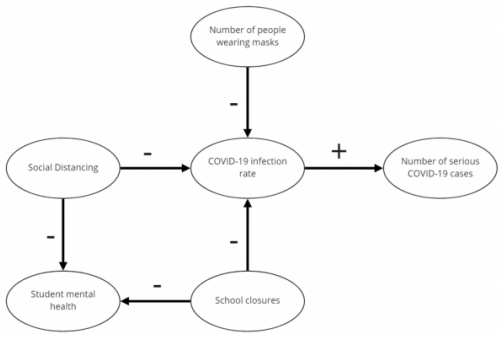
	FOOD WEB ASSESSMENT (MAMBREY ET AL., 2020)[1]	COVID-19 SYSTEMS THINKING ASSESSMENT
Systems		
ITEMS		
SysOrg	Tick the box that shows which connection is depicted in the food web. a. robins—wood mice b. robins—dandelion c. foxes—snails **d. dandelion—wood mice**	Select which of the following options is a relationship that is shown on this system map. **a. Social distancing impacts number of serious COVID-19 cases** b. Number of people wearing masks impacts school closures c. Student mental health impacts the COVID-19 infection rate d. Student mental health impacts school closures
SysBeh	Imagine, the number of earthworms is raising. Tick the box that shows what happens with the number of foxes and what caused this. **a. The number of foxes increases because the foxes can feed on more robins.** b. The number of foxes rises because the numbers of all creatures in the food web increase, if there are more earthworms. c. The number of foxes rises because there are fewer robins who feed on the fox. d. The number of foxes stays the same because there is no arrow that directly connects them with the earthworms.	Imagine school closures are increasing. Select the statement that best describes what will happen to the number of serious COVID-19 cases according to the figure shown. a. The number of serious COVID-19 cases will increase **b. The number of serious COVID-19 cases will decrease** c. The number of serious COVID-19 cases will stay the same d. The figure does not provide enough information to determine what will happen to the number of serious COVID-19 cases.
SysMod	The number of foxes shall be reduced. Tick the box that shows how this could be done and why it is possible. a. One could reduce the number of snails because then the numbers of all creatures in the food web would decrease. b. One cannot decrease the number of foxes by changing the numbers of other creatures, which are not directly connected with an arrow. **c. One could reduce the number of snails because then the foxes could feed on fewer wood mice.** d. One cannot decrease the number of foxes because the foxes are on the top of the food web.	Your goal is to increase the number of people wearing masks. Using this figure, choose the option that would best accomplish this. a. You could increase social distancing. b. You could decrease the number of serious COVID-19. c. You could decrease school closures. **d. It is impossible to increase the number of people wearing masks by changing any of the points in this system.**

TABLE 2. A Comparison of Example Stage 2 Systems and Questions. Correct answers are bolded.

Challenge: Translating Ecosystems to COVID-19 Systems

Having identified a suitable framework and how that framework could be applied such that it balances researcher and teacher interests, we were finally in the position to begin designing items. Mehren et al.'s (2018) framework, helped ensure items were designed systematically, ensuring that variations occur in deliberate ways and follow specific rules, so that we can better explore differences that emerge in domain-general skills across students as they engage with socio-scientific systems. We also relied heavily on Mambrey et al.'s (2020) assessment as a model for our COVID-19 systems assessment.

Thanks to these two resources, we were able to standardize the language used to measure each of these skills across assessment items. By standardizing language, we can more confidently attribute variations in student responses across items to variations in ability, rather than interpretation of the items. Likewise, by naming these three skills, we were able to design questions specifically to test one skill at a time, helping us identify differences in individual skills that support systems thinking. Examples of these items can be found in Table 2.

Still, we encountered several challenges as we worked to translate the assessment developed by Mambrey et al. (2020) into our context. Although our assessment would follow a similar structure and measure the same constructs as the one designed by Mambrey et al. (2020), the nature of the systems we hoped students to engage with are vastly different than the food webs incorporated in Mambrey et al.'s assessment. Adapting Mambrey et al.'s items would not be a simple matter of changing a few words to better match the content. As a result, we found it was necessary to make significant changes to the features and structures depicted within the system models themselves, as well as the questions themselves.

Design Consideration: Focusing on Identifying Causal Relationships

SysOrg items in Mambrey et al.'s (2020) assessment asked students to identify predator-prey relationships. Because these relationships are non-existent in our systems of interest, we needed to reconsider how these items should be designed to meaningfully measure students' ability to analyze the structure of a system in our context.

Rather than focusing on predator-prey relationships, we focused on causal relationships more broadly. Changes to upstream items (causes) drive changes to items downstream (effects). For these items, we did not ask students to identify what changes would occur, simply which factors would impact one another. Causality was represented in systems using arrows, with arrows leaving upstream causes and pointing towards downstream effects. Participants will have received instruction on how these conventions are used through the accompanying curriculum materials (e.g., Figure 1) as well as a reminder at the beginning of the assessment (see Figure 3).

A comparison between the systems and items developed by Mambrey et al. (2020) with those developed for our assessment can be found in Table 1. Ultimately, the assessment framework detailed in Mambrey et al. (2020) paper provided us with a matrix of question possibilities that we drew upon as we designed the COVID systems thinking test. Our goal was to create six different system models spread across three stages of difficulty. Each stage features two distinct system models. Although each system model was unique, we feel this framework ensured that this test was structured such that student systems thinking ability can be assessed in a clear, reliable, systematic way.

Design Consideration: Prioritizing Readability

As we began to construct the models that we would ask students to consider, we struggled to translate the organized structure of food webs into our context. Whereas energy always flows from prey to predator as it ascends the trophic pyramid, relationships in the complex socio-scientific system we aimed to assess do not always follow a common heuristic. Instead, our system does not feature an intuitive directionality. Because of this, we abandoned the "bottom to top" organization of a food web in favor of structures that presented the most user-friendly visual representations of our systems. Ensuring that connecting arrows did not overlap or intersect helped avoid unnecessary confusion. By prioritizing readability during system design, we hope to improve the reliability of our assessment. Decreasing the likelihood of errors caused by misinterpretations of system features and relationships helps us to be more confident that the variations in student scores are due to differences in their systems thinking ability, not their ability to interpret difficult-to-read graphics.

Design Consideration: Identifying Relationship Directionality

Similarly, unlike the relationships within a food web, the relationships we hoped to depict in our systems of interest are not always correlated in easily predictable ways. Whereas interactions commonly portrayed between trophic levels in food webs follow predictable rules (e.g., an increase in energy availability in lower trophic levels can support larger predator populations) the relationships exhibited in socio-scientific systems such as the COVID-19 pandemic do not align with a uniform set of governing rules. Thus, we found it important to include this complexity in our system models.

We addressed this concern by explicitly labeling arrows with "+" or "—" to indicate whether a relationship is positively (if X

increases so does Y) or negatively (if X increases Y decreases) correlated. This feature also helped address another prominent challenge unique to this assessment: the unfolding, controversial nature of the relationships within the system and the large amount of misinformation and disinformation that may impact student content knowledge. See Figure 3 for an example of these labeling conventions.

Challenge: Accounting for Variations in Prior Knowledge

Because of our interest in understanding how students interpret and use system models, we found it necessary to design this instrument to minimize the impact content knowledge could have on our results. As mentioned previously, our goal was to better understand domain-general skills, skills that can be applied across system contexts. Because these skills may be influenced by system-specific knowledge (Mambrey et al., 2020), varying levels of content knowledge and exposure to misinformation could obscure patterns in the application of the domain-general skills we are ultimately interested in measuring.

Design Consideration: Embedded Content Supports.

To minimize these confounding factors, we provided content supports for students directly within the assessment. To encourage students to rely on the information we provide, directions and items explicitly instructed students to determine their answers using only the information presented in the model being displayed. We felt this to be an acceptable approach because models are inherently over-simplifications of a phenomenon, and our aim was to assess how students interpret and use models presented to them—not their conceptual understanding of the COVID-19 pandemic.

Second, we designed these models to make the connections between factors explicit and whether these connections were positively or negatively correlated. By labeling arrows to represent positively and negatively correlated relationships, we hope to provide students with sufficient information to successfully answer items that depict relationships they know little about. These are the same conventions used in the curriculum materials designed as a part of this project (Sadler et al., 2021), ensuring this assessment is aligned with the instruction that they are likely to receive during the unit this assessment accompanies. We also created a splash page with a diagrammatic and text-based description of these conventions that students must click through to begin the test. This page communicates these conventions to students who are not already familiar with them or may have forgotten them. Excerpts from this splash page can be found in Figure 3.

We acknowledge the impact content knowledge and misconceptions can have on student systems thinking performance (Mambrey et al., 2020, 2022). The measures

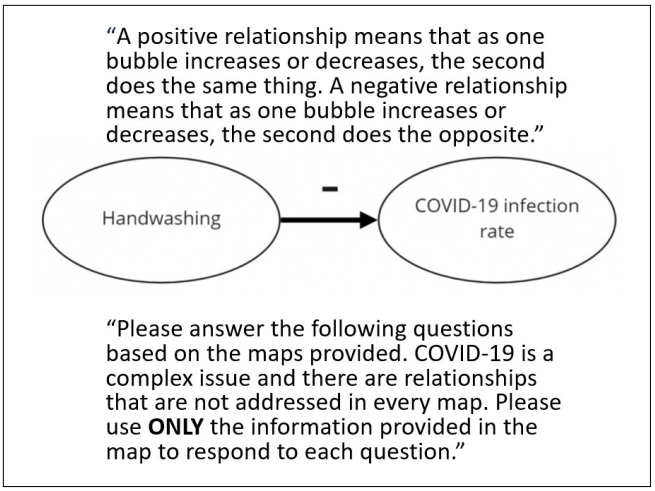


FIGURE 3. Splash Page Excerpts and Example Depiction of Negative Correlation Between Factors.

described above are designed to act as content supports for students with varying levels of exposure to these ideas. For example, whether a student rightly believes that masking is an effective way to manage infection rate, the model they are using explicitly specifies this relationship. Students correctly interpreting the model and following the directions provided should answer based on this information, not their prior knowledge. Despite this support, it is possible that students may not follow this assumption. This represents a possible direction for future research. Future design work will be dedicated to examining and refining these supports.

Challenge: Adaptability to Future Pandemics

Although there are assessments that have been designed to be content-agnostic (e.g., Moore et al., 2010; Sweeney & Sterman, 2000), the interaction between domain-specific knowledge and systems thinking (Mambrey et al., 2020) suggests that these assessments may not provide information that can reliably address our aims of understanding systems thinking about specific content areas. Despite the cross-cutting nature of domain-general skills, we recognize that the assessment we designed for COVID-19 may not translate to other pandemic contexts without modification. When designing this assessment, we hoped to create an instrument that could assess the ways students think about systems specific to the COVID-19 pandemic, while also being easy to adapt to future, similar pandemics should they arise.

Design Consideration: Factors Likely to Impact Future Pandemics

In their assessment, Mambrey et al. (2020) effectively designed two, parallel tests that were administered at the same time. One test was based on an aquatic habitat, whereas the other was based on a terrestrial habitat. Mambrey et al. changed the organisms portrayed in their food webs;

	COMPETENCE STAGE 1		COMPETENCE STAGE 2		COMPETENCE STAGE 3	
	SYSTEM 1A	SYSTEM 1B	SYSTEM 2A	SYSTEM 2B	SYSTEM 3A	SYSTEM 3B
Item 1	SysOrg	SysOrg	SysOrg	SysOrg	SysOrg	SysOrg
Item 2	SysBeh	SysBeh	SysBeh	SysBeh	SysBeh	SysBeh
Item 3	SysMod	SysMod	SysMod	SysMod	SysMod	SysMod (1)
Item 4						SysMod (2)

TABLE 3. Assessment Item Structure. Note: For examples of System 1a, System 2b, and System 3b please see Figure 2. For examples of SysOrg, SysBeh, and SysMod items, please see Table 2.

however, the structures of the systems remained unchanged between habitats. We took this as an indication that it may be possible to design systems in such a way that we could modify factors portrayed in our system models to suit future contexts without making significant changes to the system's structure. The parallel nature of their test items is relevant to our design rationale as it demonstrates the potential for this test to be adapted across multiple systems that share a similar underlying structure without threatening the assessment's validity.

It stands to reason that future viral pandemics will share many features with the currently unfolding pandemic. For example, although there are many factors that predict COVID-19 vaccine uptake in the United States; political affiliation is one of the most significant (Milligan et al., 2021). It is possible that political affiliation may not predict vaccine uptake in a future pandemic. This system factor could be replaced with a predictive factor that is more applicable (e.g., access to preventative care resources) without modifying the underlying structure of the system or the nature of the questions asked of students. By including relationships and factors that are similar to those we might expect in future pandemics, we help decrease the amount of work needed to adapt this assessment to future contexts as they arise.

FINAL DESIGN

Once we had developed initial prototypes of item types and systems, we reviewed and critiqued these items as a team. This resulted in a refined set of items that we then used as guides for developing the remaining items. We then asked two educators who were not familiar with the assessment to review the full list of items. We incorporated these educators' feedback into the next iteration. Overall, feedback throughout the process only resulted in minor changes to the wording of factors included in maps. For example, the factor "strain on hospital resources" was changed to "number of hospital beds available" to improve clarity.

With the help of a partner teacher, we then piloted the assessment with a small sample of high-school students in a school located in the Midwest United States (N=34) via Qualtrics to identify any glaring issues such as problematic items or difficulties in administering the assessment. These students were selected out of convenience; they had already given consent to participate in the larger study this assessment was developed as a part of. From this pilot, we found that the assessment could be administered quickly and easily by educators, with the test-taking approximately 15 minutes of class time. Because there were no concerns raised by the partner teacher or student scores no additional changes were made.

The resulting assessment asks students to analyze six systems across varying levels of complexity (see Table 3). Our assessment is comprised of 19 multiple-choice, single-select items. Each item is designed to assess one of three specific systems thinking skills (i.e., SysOrg, SysBeh, and SysMod). Items are scored dichotomously with a maximum overall score on this assessment being 19. This assessment features systems with varying levels of structural complexity, ranging from simple systems composed of five factors to highly interconnected systems featuring up to eight factors. Three items accompany each system such that all three skills were tested for each system. The final, most complex system features a fourth item resulting in a total of 19 items. For a visualization of how systems and items were organized, see Table 3.

LIMITATIONS

Although the design process unfolded smoothly with no major failures, the final design does reflect several compromises and assumptions that were made by the research team. One of these compromises was the decision to rely entirely on multiple-choice items rather than other more information-rich item formats frequently used to assess systems thinking, such as drawings or short answer items. Although a multiple-choice assessment is limited in its ability to capture a full range of student thinking and modeling practices; we found this limitation to be acceptable for two reasons. First, we designed this test to understand how students interpret and understand system models, not construct them. Second, we address this limitation through observation data collected in the classroom as students work together to construct models. Supplementing these quantitative findings with qualitative data allows us to make inferences not possible with either method alone. Ideally, the data generated by this test and through observational work

will be placed in conversation with one another, providing us with a richer understanding of the interactions between the practices involved in the creation of models and the skills needed to interpret models.

Another notable compromise we made while designing this assessment stems from the ways we depict systems. The static presentation of systems in this test limits our ability to understand how students think about many features of complex systems and causal relationships that have temporal components like steady-states and simultaneous causality, or cyclical features such as feedback loops (Grotzer, 2012). Our representations inherently constrained the extent to which our findings could be applied to systems with these features.

Similarly, we deliberately omitted probability and magnitude from the relationships depicted in the models. Although complex systems rarely operate in a sequential, deterministic, "all or nothing" fashion and often result in nonlinear (e.g., exponential) relationships, we felt these understandings would be best assessed in a more naturalistic setting. Incorporating these features would require students to engage in mathematical calculations, introducing mathematical ability as a confounding variable and dramatically increasing the amount of time necessary to administer the test—a burden we were not willing to impose on our partner teachers. Because this test is to be paired with student observations and interviews, and the unit this test is situated in contains a computational modeling component that could support student exploration of these ideas, we were willing to accept this limitation for our purposes.

Finally, the extent to which our embedded content supports are effective remains undetermined. It is very possible that students are relying on their prior knowledge and intuition to solve these problems entirely, and that these content supports do not factor into how students approach these tasks. Although we are confident in our decision to include content supports, we acknowledge that this was a conjecture-driven choice.

CONCLUSION

This paper presents the rationale that drove the design of an instrument to help us better understand the ways students think about complex, socio-scientific systems. Although SSI-based instruction can advance students' ability to think about the social dimensions of scientific issues, more targeted interventions are needed to support student reasoning as they grapple with the complexity of systemic issues like the COVID-19 pandemic.

ACKNOWLEDGEMENTS

This work was supported by U.S. National Science Foundation [grant number 2023088].

ENDNOTES

[1] Food Web Assessment items can be found in the online supplemental materials associated with Mambrey et al., (2020). Adapted from "Appendix S1: Supporting Information" by S. Mambrey, J. Timm, J. Landskron, P. Schmiemann, 2020, *Journal of Research in Science Teaching, 57*(10) (https://doi.org/10.1002/tea.21649). CC BY 4.0

REFERENCES

Bechtel, W., & Abrahamsen, A. (2005). Explanation: A mechanist alternative. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 36(2), 421–441. https://doi.org/10.1016/j.shpsc.2005.03.010

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(5), 518–560. https://doi.org/10.1002/tea.20061

Ben-Zvi Assaraf, O., & Orion, N. (2010). System thinking skills at the elementary school level. *Journal of Research in Science Teaching*, 47(5), 540–563. https://doi.org/10.1002/tea.20351

Ekborg, M., Ottander, C., Silfver, E., & Simon, S. (2013). Teachers' experience of working with Socio-scientific Issues: A large scale and in depth study. *Research in Science Education*, 43(2), 599–617. https://doi.org/10.1007/s11165-011-9279-5

Evagorou, M., Korfiatis, K., Nicolaou, C., & Constantinou, C. (2009). An investigation of the potential of interactive simulations for developing system thinking skills in elementary school: A case study with fifth-graders and sixth-graders. *International Journal of Science Education*, 31(5), 655–674. https://doi.org/10.1080/09500690701749313

Forum on regenerative medicine, Board on Health Sciences Policy, Health and Medicine Division, & National Academies of Sciences, Engineering, and Medicine. (2021). Applying Systems Thinking to Regenerative Medicine: *Proceedings of a Workshop* (S. Addie, M. Hackmann, L. Teferra, A. Nicholson, & S. H. Beachy, Eds.; p. 26025). National Academies Press. https://doi.org/10.17226/26025

Grotzer, T. A. (2012). *Understandings of consequence: learning causality in a complex world*. Rowman & Littlefield Education.

Grotzer, T. A., & Tutwiler, M. S. (2014). Simplifying causal complexity: How interactions between modes of causal induction and information availability lead to heuristic-driven reasoning. *Mind, Brain, and Education*, 8(3), 97-114.

Grotzer, T. A., Tutwiler, M. S., Kamarainen, A. M., Derbiszewska, K. M., Metcalf, J., & Dede, C. J. (2016). Students' reasoning tendencies about the causal dynamics of ecosystems and the impacts of MUVE vs. Non-MUVE instructional contexts. In The Next Phase of Research in Complex Systems in Science Education, 29. https://clic.gse.harvard.edu/selected-publications

Jacobson, M. J., & Wilensky, U. (2006). Complex systems in education: scientific and educational importance and implications for the learning sciences. *Journal of the Learning Sciences*, 15(1), 11–34. https://doi.org/10.1207/s15327809jls1501_4

Ke, L., Sadler, T. D., Zangori, L., & Friedrichsen, P. J. (2021). Developing and using multiple models to promote scientific literacy in the context of socio-scientific issues. *Science & Education*, *30*(3), 589–607. https://doi.org/10.1007/s11191-021-00206-1

Mambrey, S., Schreiber, N., & Schmiemann, P. (2022). Young students' reasoning about ecosystems: The role of systems thinking, knowledge, conceptions, and representation. *Research in Science Education*, 52(1), 79–98. https://doi.org/10.1007/s11165-020-09917-x

Mambrey, S., Timm, J., Landskron, J. J., & Schmiemann, P. (2020). The impact of system specifics on systems thinking. *Journal of Research in Science Teaching*, 57(10), 1632–1651. https://doi.org/10.1002/tea.21649

Mehren, R., Rempfler, A., Buchholz, J., Hartig, J., & Ulrich-Riedhammer, E. M. (2018). System competence modelling: Theoretical foundation and empirical validation of a model involving natural, social and human-environment systems. *Journal of Research in Science Teaching*, 55(5), 685–711. https://doi.org/10.1002/tea.21436

Mehren, R., Rempfler, A., Ulrich-Riedhammer, E. M., Buchholz, J., & Hartig, J. (2015). Wie lässt sich Systemdenken messen? Darstellung eines empirisch validierten Kompetenzmodells zur Erfassung geographischer Systemkompetenz / How to measure systems thinking? Presentation of an empirically validated competence model for assessing geographic system competence. Geographie aktuell & Schule, 37(215), 4–16.

Milligan, M. A., Hoyt, D. L., Gold, A. K., Hiserodt, M., & Otto, M. W. (2021). COVID-19 vaccine acceptance: Influential roles of political party and religiosity. *Psychology, Health & Medicine*, 0(0), 1–11. https://doi.org/10.1080/13548506.2021.1969026

Moore, S. M., Dolansky, M. A., Singh, M., Palmeri, P., & Alemi, F. (2010). *The Systems Thinking Scale.* Unpublished manuscript.

Next Generation Science Standards: For States, By States. (2013). *The National Academies Press*. https://doi.org/10.17226/18290

Sadler, T. D., Barab, S. A., & Scott, B. (2007). What do students gain by engaging in socioscientific inquiry? *Research in Science Education*, 37(4), 371–391. https://doi.org/10.1007/s11165-006-9030-9

Sadler, T. D., Rawson, R., Kirk, E. A., Elsner, J., Ke, L., Apple, S., Elmey, C., Kinslow, A. T., McKee, R., Miller, D., Plato, J., Rockett, J., Wagner, B., Friedrichsen, P. J., & Zangori, L. (2021). COVID-19: A model-oriented issues-based science unit. University of North Carolina at Chapel Hill School of Education. https://tarheels.live/seel/teaching-materials/curriculum/covid-19-teacher-materials/

Sweeney, L. B., & 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

Tidemand, S., & Nielsen, J. A. (2017). The role of socioscientific issues in biology teaching: From the perspective of teachers. *International Journal of Science Education*, 39(1), 44–61. https://doi.org/10.1080/09500693.2016.1264644

Wilensky, U., & Jacobson, M. J. (2014). Complex systems and the learning sciences. In R. K. Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences* (2nd ed., pp. 319–338). Cambridge University Press. https://doi.org/10.1017/CBO9781139519526.020

Yoon, S. A. (2018). Complex systems and the learning sciences. In F. Fischer, C. E. Hmelo-Silver, S. R. Goldman, & P. Reimann (Eds.), *International Handbook of the Learning Sciences* (1st ed., pp. 157–166). Routledge.

Yoon, S. A., Goh, S.-E., & Park, M. (2018). Teaching and learning about complex systems in K–12 science education: A review of empirical studies 1995–2015. Review of Educational Research, 88(2), 285-325. https://doi.org/10.3102/0034654317746090

Zeidler, D. (2014). Socioscientific issues as a curriculum emphasis: Theory, research and practice. In N.G. Lederman & S.K. Abell (Eds.), *Handbook of research on science education* (Vol.*2*, pp. 697–726). Routledge.