Investigating the Relative Difficulty of Complex Systems Ideas in Biology

Sao-Ee Goh[a], Susan A. Yoon[a], Joyce Wang[a], Zhitong Yang[a], Eric Klopfer[b]
[a]Graduate School of Education, University of Pennsylvania, 3700 Walnut Street, Philadelphia, PA 19104
goh@dolphin.upenn.edu, yoonsa@gse.upenn.edu, joyce.s.wang@gmail.com, zhitongy@hotmail.com
[b]Scheller MIT Teacher Education Program, 77 Mass. Ave., MIT Building 10-337, Cambridge, MA 02139
klopfer@mit.edu

Abstract: A number of students' biology misconceptions can be attributed to their lack of understanding in complex systems. Comprehending complex systems can be counterintuitive and difficult. A literature review reveals that we have yet to systematically determine a learning approach to address these learning challenges. In this paper, we propose that learning progressions research offers a methodical approach to organize the learning pathways students take to improve conceptual competence in complex systems. As a first step, we articulate a sequence of complex systems ideas - from the least to most difficult - by analyzing students' written responses. Using an Item Response Theory model, we found that the easiest ideas to grasp are those that relate to the interconnected nature of these systems whereas the most difficult ideas are those concerning the decentralized organization of the system, and the predictability of the system effects.

Introduction

Studies in science education have revealed robust misconceptions about concepts and processes in school biology that directly impact students' abilities to understand scientific topics. Misconceptions have been found across scales, from atoms (Taber & Garcia-Franco, 2010), to cells and genes (Garvin-Doxas & Klymkowsky, 2008), to organisms and ecology (Gotwals & Songer, 2010). Difficulties in understanding the relationships between these various scales have also been documented (Sewell, 2002). Researchers have speculated that these problems may exist due to a lack of understanding of the complex systems realms in which these scientific phenomena reside and interact (Charles & d'Apollonia, 2004; Chi, 2005). Thus, educational agencies in the US have urged science curriculum and instruction to emphasize systems content (AAAS, 2009; NRC, 2011).

A complex system can be defined as an organization of interconnected components that as a whole, exhibits patterns and properties not obvious from those of the individual components (Mitchell, 2009). The way in which herds of ungulates roam across the savannah provides one of the most vivid examples of how individual actions and interactions lead to large scale patterns. To understand complex systems is to be able to reason that the individual components interact with one another in multiple and nonlinear ways, and recognize that the patterns observed at the system level emerge from these interactions at the component level (Sweeney & Sterman, 2007; Yoon, 2008). Research has documented that grasping complex systems ideas can be counterintuitive and difficult (Chi, 2005; Jacobson & Wilensky, 2006). While there have been several studies that promote the learning of particular complex systems or specific complexity ideas (e.g., Klopfer et al., 2009; Levy & Wilensky, 2009), the literature also suggests that we have yet to systematically determine a learning approach to address the challenges associated with a comprehensive understanding of the various ideas (Hmelo-Silver & Azevedo, 2006).

We believe that the learning progressions methodology offers one such systematic approach in structuring the learning of various complex systems ideas. Learning progressions are defined as sequences of ordered descriptions that illustrate the learning pathways students can take to improve conceptual competence in science (Alonso & Steedle, 2006; CPRE, 2009). For example, Mohan and her colleagues (2009) identified levels of increasing sophistication in students' perception of carbon-transforming events in complex socio-ecological systems. These ordered descriptions represent a research-informed framework for structuring the learning of core scientific ideas (NRC, 2007). Curriculum and instructional activities can in turn be mapped onto the learning progressions so as to better facilitate the learning of the core ideas (Gotwals & Songer, 2010; Songer et al., 2009).

Our study extends the research on learning progressions by focusing on complex systems ideas grounded in high school biology content. Our broad goal is to eventually design valid curricular and instructional activities based on a learning progression of complex systems ideas. As a first step, in this paper we report on an exploratory work that investigates a conceptual sequence of complex systems ideas, from the least to the most difficult, by analyzing a diverse group of students' written responses on the effects of geese on a park ecosystem.

Theoretical Considerations

Our research lies at the intersection of two fields – understanding complex systems and learning progressions. We first discuss known challenges in understanding complex systems, before illustrating how learning progressions as a methodology can help address these challenges.

Understanding complex systems

As previously mentioned students typically possess misconceptions about complex systems (Ben-Zvi Assaraf & Orion, 2005; Hmelo, Holton, & Kolodner, 2000; Jacobson & Wilensky, 2006). For example, students tend to adopt a linear approach when thinking about the relationships among system components (Gotwals & Songer, 2010; Hogan, 2000; Riess & Mischo, 2010; Sweeney & Sterman, 2007). That is, they perceive single cause-and-effect relationships, where small actions lead to small effects. However, a complex system with its multiple connections among the components can often result in an action that gives rise to widespread effects throughout the system. Students also generally conceive of systems being controlled by certain components, and systemic patterns as designed with specific functions in mind (Penner, 2000; Resnick, 1996; Taber & Garcia-Franco, 2010). In other words, students are unable to interpret the decentralized nature and emergent patterns of complex systems. Wilensky and Resnick (1999) argue that this attribution of centralized control and intentionality to system processes is due to students' confusion about the various levels of representation (e.g., microscopic and macroscopic) of complex systems. Despite the valuable information these studies provide, they examine single aspects of complex systems understanding. Given the current overall systems thrust in science education (AAAS, 2009; NRC, 2011) it is prudent to turn our attention to developing a more comprehensive view of complex systems content.

Some studies have already articulated analytical frameworks that encompass multiple complex systems ideas and beliefs (Hmelo-Silver et al., 2004; 2007; Jacobson, 2001; 2011). For example Jacobson and his colleagues (2001) developed a framework that delineates distinct categories of complex systems ideas. Using this framework in an expert-novice study, he discovered that pre-college students tended to believe that systems operate in reductive, centralized and predictable way, whereas, science experts described phenomena as non-reductive, decentralized, and non-linear. We adapt this framework and evolutions of it (Jacobson, 2011; Yoon, 2008; 2011) to analyze levels of understanding for various ideas of complex systems. In doing so, we unpack and reveal the relative levels of difficulties that can eventually inform the development of a learning progression.

Learning Progressions

Learning progressions provide a methodical approach to organize curriculum and instructional activities to learn and teach about complex systems ideas, as they illustrate the cognitive pathways and skills students are likely to follow over a period of time in mastering the scientific concepts (Alonso & Steedle, 2006; NRC, 2007). Learning progressions structure the sequences based on what cognitive researchers know about science learning from empirical research findings (CPRE, 2009).

To date, there have been three published works on learning progressions on biology topics related to complex systems concepts (Gotwals & Songer, 2010; Mohan et al., 2009; Songer et al., 2009). These studies have collectively ascertained that students' learning of biological systems, or the development in the way they think about systems, follows some trajectory. However, these progressions address particular science content such as ecosystems and biodiversity rather than a more generalized complexity perspective. Although they have not characterized their work in terms of learning progressions, Ben-Zvi Assaraf and Orion (2005; 2010a; 2010b) examine students' development of systems thinking abilities, and construct a hierarchical model of the stages by which students' reason about complex systems. They explain that at the most rudimentary level students are able to identify the components of a system and processes within the system. At the intermediate level, students were able to identify relationships between system components, and organize the systems' components, processes, and their interactions within a framework of relationships. At the most sophisticated level students can perceive hidden system components, explain the history of current patterns, and make predictions on the evolution of the system. However, their model comprises only a subset of essential complex systems components as identified in Jacobson (2001). A sequence or markers for learning about other equally salient ideas such as decentralization and emergence are also necessary to ensure that students acquire an adequate overall understanding of complex systems.

As an initial development of a learning progression on complex systems ideas, a comprehensive set of ideas is first differentiated in order of difficulty levels. Such delineation can help organize the sequence of complexity ideas to be learned. There are generally two broad approaches in constructing the initial learning sequences (CPRE, 2009). The first approach begins by conjecturing a possible sequence from existing literature and then validating it. For example, Songer and her team (2009) first hypothesized their learning progression on biodiversity from literature reviews, and then tested the validity of the progression by comparing student learning outcomes in a control-treatment study. The second approach starts by analyzing a cross-sectional sample of students' responses to provide indicators of their understanding and then derives the levels of

sophistication from the analysis. For instance, Mohan and her colleagues (2009) analyzed almost 300 elementary through high school students' written accounts of the biogeochemical carbon cycling processes in socio-ecological systems for distinct levels of sophistication, and constructed their progression based on these levels. The latter approach presents an advantage when existing literature does not inform much about which ideas should be learned before others. In our exploratory study, we adopt this more 'grounds-up' approach in drawing up the initial sequence of our learning progression.

Methods

Context and Participants

This study is part of a larger National Science Foundation-funded research project in which we design and develop curriculum and instructional activities using computational modeling tools in four high school biology units – *Chemistry of Life*, *Population Ecology*, *Community Ecology*, and *Evolution* that promote student learning of complex systems. While the activities are being constructed and tested, we focus concurrently on building the learning progression that informs them. To begin, we developed eight open-ended short-answer questions, two for each biology unit. These were selected or derived from the OECD Programme for International Student Assessment (PISA, 2006) international science test and Ph.D. level biology content experts. We then administered this test to Grades 8 to 12 students in order to determine the range of conceptual difficulties.

To recruit students to participate in our study, we enlisted the support of science teachers with whom we had previously worked in another study. We targeted different kinds of schools (including 3 magnet and 4 public) to account for differences in science learning experience. All teachers and students were recruited from a large urban school district in the northeast part of the US. Two days prior to the implementation, a researcher went to the teachers' classes and gave a short presentation regarding the goals of the study and the roles of the students (should they participate). Those who participated were self-selected and were offered a $10 gift card in exchange for their participation. Their involvement lasted about one hour after school in which they answered questions on the open-ended biology test. The students were given as much time as they required to answer the questions. In total 44 students of various ethnicities from 7 urban schools participated. There were 20 males and 24 females in grade 8 (5), grade 9 (13), grade 10 (14), and grade 12 (12). The recruitment and data collection occurred over 3 months.

Data Sources, Data Coding and Analyses

For this paper, only the responses to one question were analyzed: "Imagine a flock of geese arriving in a park in [your city], where geese haven't lived before. Describe how the addition of these geese to the park affects the ecosystem over time. Consider both the living and non-living parts of the ecosystem." This question, written by an expert in biology, sought to solicit students' understanding of biology and complex systems in an ecological context.

A content analysis of the students' responses was performed using six categories of complex systems understanding - *Agent Effects, Action Effects, Networked Interactions, Multiple Causes, Order,* and *Processes* - derived from Jacobson's (2001; 2011) and Yoon's (2008; 2011) studies. Table 1 provides an abridged description of each category. To account for variation in students' understanding of the complexity ideas within each category, each response was coded six times - once for each category - for four levels of increasing sophistication: Completely Clockwork (Level 1), Somewhat Clockwork (Level 2), Somewhat Complex (Level 3), and Completely Complex (Level 4). Clockwork responses encompassed those that showed linear, deterministic, isolated, centralized, single-cause, and predictable system interactions or states, whereas complex responses included those that demonstrated non-linear, non-deterministic, networked, decentralized, multiple-causes, and random system interactions or states.

After the coding manual was constructed and vetted, its reliability was assessed with two independent doctoral student raters coding 20% of the written responses. An inter-rater agreement of 0.8 was achieved collectively across categories using the Cronbach alpha reliability test. The remainder of the responses was subsequently coded by two of the authors using the coding scheme, with any discrepancies discussed.

Below we provide an example of a response that achieved mostly Levels 3 or 4 for all categories of complexity ideas to show how the coding was done:

> *Well if geese arrived it would probably help the ecosystem. The bird droppings might make the soil fertile. It would start to look a lot greener. The problem with that is erosion. The increase of plants and root size might cause paths or walkways to be damaged or destroyed. Statues might start to fall apart from the constant weight of birds. Plus the increase of plants in amount and size make O2 levels higher. Which could cause a warm and wet ecosystem, much similar to a swamp. Over a long period of time of course.* (Student response, March 2011)

Table 1: Complex Systems Category Code Descriptions

Category	Completely Clockwork (Level 1)	Somewhat Clockwork (Level 2)	Somewhat Complex (Level 3)	Completely Complex (Level 4)
Agent Effects The emphasis is the predictability of the effects caused by the part in question.	Response shows that the way in which a part operates or affects other parts is completely predictable. No alternative is offered in the response.	Response shows that the way in which a part operates or affects other parts is somewhat predictable. There are 1-2 possibilities suggested in the response.	Response shows that the way in which a part operates or affects other parts is somewhat unpredictable. There are 3-4 possibilities suggested in the response.	Response shows that the way in which a part operates or affects other parts is completely unpredictable. There are many possibilities suggested in the response.
Action Effects Three components are considered: (i) the relative scale of outcomes caused by action; (ii) the cascading effects of the action; and (iii) the time scale at which changes happen.	Response indicates (i) small actions only lead to small effects; (ii) there is a sense that the action causes localized changes only; and (iii) the changes are immediate and do not sustain for a long time.	Response contains one complex component (out of three) of action effects. (See Level 4.)	Response contains two complex components (out of three) of action effects. (See Level 4.)	Response indicates (i) small actions can lead to large effects; (ii) the action can produce both localized changes and cascading effects; and (iii) the changes can take place both immediately and over a long period of time.
Multiple Causes The focus is on the number of causes that may/will attribute to the outcome(s) of an event.	Response attributes the outcome(s) of an event to one cause/factor.	Response attributes the outcome(s) of an event to two causes/factors.	Response attributes the outcome(s) of an event to three causes/factors.	Response attributes the outcome(s) of an event to four or more causes/factors.
Networked Interactions Three components are assessed: (i) interdependency among parts in the system; (ii) nonlinearity in reasoning; and (iii) emergent patterns over scale.	Response indicates that (i) the parts of a system are isolated with no interdependency among them; (ii) the interactions between parts are linear with no feedback; and (iii) the patterns at the system level are the same from those at the component level.	Response contains one complex component (out of three) of networked interactions. (See Level 4.)	Response contains two complex components (out of three) of networked interactions. (See Level 4.)	Response indicates that (i) the parts are interdependent; (ii) the interactions between parts are non-linear with feedback; and (iii) the patterns at the system level are emergent.
Order The focus is the organization of the system or phenomenon – centralized or decentralized.	Response indicates that the system is controlled by one central agent, that is, all action is dictated by a leader. Order in the system is established 'top-down' or determined with a specific purpose in mind.	Response indicates that the system is largely controlled by 2-3 central agents, i.e., there are other parts that may dictate how the system behaves. Order in the system is established 'top-down.'	Response indicates that the system is largely decentralized and the control lies with 4-5 components. However, there is little evidence to show that the order in the system is self-organized.	Response indicates that the system is decentralized and control lies with a myriad (more than 5) of parts. Order in the system is self-organized or 'bottom-up', and emerges spontaneously.
Processes Processes refer to the dynamism of the mechanisms that underlie the phenomena; in other words, how the system works or is thought to work.	Response indicates that the system is composed of static events. While perturbations in the system cause change to occur, the change terminates once an outcome is achieved (i.e., a definite end).	Response indicates that the system is somewhat composed of static events with suggestions that these events take time to reach the outcome(s).	Response indicates that the system is somewhat of an on-going process. Perturbations take a long time to reach the final outcomes, which are at larger scale than the initial event(s).	Response indicates that the system is an on-going, dynamic process. System continues to be in a state of flux. The parts adapt or evolve, and continue to do so accordingly.

In this exemplar response, the student repeatedly used non-deterministic words, such as "might" and "could," and suggested three possibilities of the effects due to the geese's arrival, which indicates his uncertainty in the effects of the geese's arrival (Level 3 *Agent Effects*). There are suggestions that the geese's

arrival can cause cascading effects to the soil, the plant population, the oxygen levels, may lead to large-scale and long-term effects such as causing a "warm and wet ecosystem" (Level 4 *Action Effects*). More than four factors (e.g., droppings, soil, erosion, plants, oxygen) have been identified as attributing to possible outcomes (Level 4 *Multiple Causes*). The response also hints of an understanding of the interdependence of the various components in the park ecosystem, the feedback mechanisms present in the system (i.e., paths or walkways damaged by more plants), and the possible emergence of a systemic pattern (i.e., warm and wet environment) (Level 4 *Networked Interactions*). In addition, the idea of decentralization is clearly demonstrated as four actors (i.e., geese, plants, soil, and oxygen) are said to be involved (Level 3 *Order*). Furthermore, the response implies that the perturbations in the park ecosystem are somewhat on-going, and may take a long time to arrive at a final state (Level 3 *Processes*).

As we are interested in determining the differences in conceptual difficulties among the six categories of complex systems ideas, and given that we have employed successively ordered rating scales (i.e., 4 levels of understanding), the *Generalized Partial Credit Model* (GPCM: Muraki, 1992), which is an item response theory model, was deemed appropriate for the analysis of the responses. The items, or categories of ideas in our case, are conceptualized as a series of hierarchical levels of performances where respondent receives partial credit for successfully performing at a particular level. Since our coding included four levels of understanding for six categories, the GPCM was an appropriate means to provide the information that we wanted. Students' raw scores were first standardized on a continuum between -3 to +3 to reveal the difficulty level of each category. On this continuum of logit scale, 0 is set as the mean of the item difficulty parameter. On the positive direction toward +3, each increase indicates that the item is becoming more difficult; conversely, on the negative direction toward -3, each decrease signals that the item is turning less difficult. In addition, the model is able to show us how well each item can distinguish students with different abilities by the discrimination parameter. The discrimination parameter typically ranges between 0.5 and 2.5 in value. The larger this parameter, the more effective the item can distinguish students with varying levels of understanding.

Results

We rated the written responses for their understanding in each category of complex systems ideas. A breakdown of how they scored is given in Table 2. A simple frequency count shows most of the responses were at a level 2 (somewhat clockwork) or level 3 (somewhat complex) understanding of various complex systems ideas, with very few showing a level 4 (completely complex) understanding. We expected this because as described earlier, the literature indicated that most students face difficulty in grasping complex systems ideas.

Table 2: Breakdown of scoring for each category of complex systems ideas

Category	Level 1	Level 2	Level 3	Level 4
Agent Effects	16	14	12	2
Action Effects	8	11	14	11
Multiple Causes	11	19	9	5
Networked Interactions	6	12	22	4
Processes	11	23	7	3
Order	15	16	12	1
Total:	67	95	76	26

We then ran the GPCM on the 44 sets of ratings, using PARSCLE 4.1. The model converged at critical value = 0.005, and had no items fit statistical significance, indicating a good model fit. The mean difficulty of the six items or categories is 0.49 (*SD* = 0.45), and mean discrimination parameter was 1.10 (*SD* =0.39). The categories were found reliable to measure students' understanding of complex system (composite reliability = 0.87). The total test information peaked between 0 and 1, indicating the categories could measure students with slightly higher than average ability with most precision. Table 3 presents the six categories, their difficulty parameters, and their discrimination parameters. As indicated by the difficulty parameters, *Agent Effects, Order,* and *Processes* categories of ideas were found to be the most difficult among these students; *Multiple Causes* was found to be at the intermediate level; and *Networked Interaction* and *Action Effects* categories were the easiest. All six categories can also distinguish students based on their abilities, as their discrimination parameters ranged from 0.69 to 1.64. That is, the question is suitable for measuring the categories theorized. In addition, an examination of the test characteristic curves for this model indicated that the six items collectively provide adequate information for students with ability levels both below and above average.

Table 3: Generalized Partial Credit Model

Category	Difficulty (SE)	Slope (SE)

Action Effects	-0.17 (.018)	1.42 (0.50)
Networked Interaction	0.04 (0.22)	1.64 (0.58)
Multiple Causes	0.54 (0.21)	1.24 (0.51)
Processes	0.77 (0.26)	0.69 (0.39)
Order	0.86 (0.22)	0.72 (0.98)
Agent Effects	0.87 (0.29)	0.91 (0.34)

$N = 44$.

Discussion and Implications

In the introduction, we highlighted that a number of students' misconceptions in biology can be attributed to their lack of understanding in complex systems. Comprehending complex systems ideas can be counterintuitive and difficult, and a review of the literature has revealed that we have yet to systematically determine a learning approach to address these learning challenges. We have proposed that learning progressions research can help to systematize the learning pathways students take to improve conceptual competence in complex systems. These learning pathways can in turn guide the curriculum and instructional activities in promoting complex systems learning, which is our broad goal.

This paper articulates the initial development of a learning progression of complex systems ideas. A sequence of a comprehensive set of these ideas, from the least to the most difficult to understand, has been determined. As there is no prior literature to guide us in hypothesizing such a sequence, we have adopted a 'grounds-up' approach in constructing the progression similar to Mohan et al.'s study (2009). Using the GPCM, we analyzed 44 students' written responses to a biology question on the effects of geese arrival on a park ecosystem for their understanding of various complexity ideas. We acknowledge the small sample size in this exploratory study and the implication this may have in our analyses, but we believe we have taken sufficient steps to ensure that different grade levels and types of schools are represented so as to make our findings as generalizable as possible.

We found that two of the easiest categories of ideas for students to comprehend are those that relate to the effects of actions in complex systems (*Action Effects*), and the interconnected and emergent nature of these systems (*Networked Interactions*). A "completely complex" understanding of these ideas includes the ability to reason that: small actions can lead to large effects; these actions can produce both localized and cascading changes; the changes can occur immediately and over a long period of time; the parts in the system are interdependent; the relationships among the parts in the system are nonlinear with feedback mechanisms; and the parts interact to produce emergent patterns that are not obvious at the component level. Students seem to experience more difficulty in interpreting that there are multiple causes that may attribute to the outcome(s) of a change (*Multiple Causes*), and perceiving that the system continues to undergo adaptation or evolution (*Processes*). The categories of complex systems ideas that apparently are the most complicated to understand are *Order* and *Agent Effects*; these complexity ideas concern the decentralized organization of the system, and the predictability of the effects caused by the parts of the system.

We believe this ordered sequence of complexity ideas is valid – at least in the context of our test question. In a question that frames the complexity within a familiar park ecosystem, the students might have found it easier to observe the manifestation of complexity ideas associated with the interconnectedness and interdependency of the various plant and animal species in the park, the cascading, emergent, and non-linear effects the arrival of the geese can have on the rest of the ecosystem (i.e., the *Action Effects* and *Networked Interactions* categories). In contrast, complexity ideas in the *Order* and *Agent Effects* categories are less 'visible;' the ideas that there is a myriad of other components (e.g., plants, predators and prey to geese, climate) that may contribute to how the system is organized, and that it is not possible to predict with precision the effects of geese's arrival at the park could well fall outside of students' perceptual abilities. Chi (2005) also proposes that some scientific concepts and ideas may be easier to comprehend because of the ontological categories they belong to. Chi explains that "ontological categories refer to the basic categories of realities or the kinds of existence in the world, such as concrete objects, events, and abstractions" (p. 163). The 'more visible' ideas of complex systems may have represented ontologically-easier categories for students to understand.

In sum, our findings align well with current literature on complex systems understanding, and extend it by establishing an initial progression of the various ideas. Resnick (1996) and other researchers (e.g., Wilensky & Resnick, 1999; Wilensky & Reisman, 2006; Jacobson et al., 2011) have repeatedly iterated the deterministic and centralized mindset students typically have. Likewise, Charles and d'Apollonia (2004) and others (Ben-Zvi Assaraf & Orion, 2010a; Hogan, 2000; Lin & Hu, 2003; Perkins & Grozter, 2005) have observed that students tend to think linearly about relationships among the components or parts of complex systems. Collectively, these researchers have shown that when students are exposed to appropriate instruction, the learning challenges can be overcome. However, as we have argued, most of the interventions are targeted at improving student understanding of particular systems or specific complexity concepts rather than promoting a more holistic understanding of complex systems. A learning sequence of complexity ideas that specifies levels of difficulty

which we have found seeks to address this gap. This initial conception of difficulty levels is only the first, but important, step to developing a systematic approach in designing curriculum and instructional activities for the learning of complex systems ideas. In subsequent studies we aim to further validate the sequence with a larger sample size and use questions that involve other types of complex systems in order to determine how robust it is as a general complex systems heuristic.

Reference

Alonzo, A. & Steedle, J. (2008). Developing and assessing a force and motion learning progression. *Science Education*, *93*, 389-421. doi: 10.1002/sce.20303.

American Association for the Advancement of Science, AAAS (2009). *Benchmarks for scientific literacy*. Retrieved from http://www.project2061.org/publications/bsl/online/index.php on Mar 1, 2010.

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. doi: 10.1002/tea.20061.

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

Ben-Zvi Assaraf, O., & Orion, N. (2010b). Four case studies, six years later: Developing system thinking skills in junior high school and sustaining them over time. *Journal of Research in Science Teaching*, *47*(10), 1253-1280. doi: 10.1002/tea.20383.

Charles, E. S., & d'Apollonia, S. (2004). Developing a conceptual framework to explain emergent causality: Overcoming ontological beliefs to achieve conceptual change. In K. Forbus, D. Gentner, & T. Reiger (Eds.), *Proceedings of the 26th annual cognitive science society*. Mahwah, NJ: Lawrence Erlbaum Associates.

Chi, M.T.H. (2005). Commonsense conceptions of emergent processes: Why some misconceptions are robust. *The Journal of the Learning Sciences, 14*(2), 161-199. doi: 10.1207/s15327809jls1402_1.

Consortium for Policy Research in Education, CPRE (2009). *Learning Progressions in Science: An Evidence-Based Approach to Reform*. CPRE Research Report # RR-63. New York: Teachers College Columbia University. Retrieved from http://www.cpre.org/images/stories/cpre_pdfs/lp_science_rr63.pdf on August 20th, 2010.

Embretson, S.E., & Reise, S.P. (2000). *Item response theory for psychologists.* Mahwah, NJ: Erlbaum.

Garvin-Doxas, K.,& Klymkowsky, M.W. (2008). Understanding randomness and its impact on student learning: Lessons learned from building the biology concept inventory (BCI). *Life Science Education*, *7*, 227-233. doi: 10.1187/cbe.07-08-0063.

Gotwals, A.W., & Songer, N.B. (2010). Reasoning and down a food chain: Using an assessment framework to investigate students' middle knowledge. *Science Education, 94,* 259-281. doi: 10.1002/sce.20368.

Hmelo-Silver, C.E., & Azevedo, R. (2006). Understanding complex systems: some core challenges. *The Journal of the Learning Sciences*, *15*(1), 53-61. doi: 10.1207/s15327809jls1501_7.

Hmelo, C. E., Holton, D., & Kolodner, J. L. (2000). Designing to learn about complex systems. *The Journal of the Learning Sciences*, *9*, 247–298. doi: 10.1207/S15327809JLS0903_2.

Hmelo-Silver, C.E., Marathe, S., & Liu, L. (2007). Fish swim, rocks sit, and lungs breathe: Expert-novice understanding of complex systems. *The Journal of the Learning Sciences*, *16*(3), 307-331. doi: 10.1080/10508400901413401.

Hmelo-Silver, C.E., & Pfeffer, M.G. (2004). Comparing expert and novice understanding of a complex system from the perspective of structures, behaviors, and functions. *Cognitive Science*, *28*, 127-138. doi: 10.1207/s15516709cog2801_7.

Hogan, K. (2000). Assessing students' systems reasoning in ecology. *Journal of Biological Education*, *35*(1), 22-28.

Jacobson, M. (2001). Problem solving, cognition, and complex systems: Differences between experts and novices. *Complexity, 6*(3), 41-49. doi: 10.1002/cplx.1027.

Jacobson, M. J., Kapur, M., So, H.J., & Lee, J. (2011). The ontologies of complexity and learning about complex systems. *Instructional Science, 39*, 763-783. doi: 10.1007/s11251-010-9147-0.

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

Johnson, S. (2002). *Emergence*. New York: Touchstone.

Klopfer, E., Scheintaub, H., Huang, W., Wendel, D., & Roque, R. (2009). The simulation cycle: combining games, simulations, engineering and science using StarLogo TNG. *E-Learning*, *6*(1), 71-96.

Liu, L., & Hmelo-Silver, C.E. (2009). Promoting complex systems learning through the use of conceptual representations in hypermedia. *Journal of Research in Science Teaching*, *46*(9), 1023-1040. doi: 10.1002/tea.20297.

Levy, S.T., & Wilensky, U. (2009). Students' learning with the connected chemistry (CCI) curriculum: Navigating the complexities of the particulate world. *Journal of Science Education and Technology*, *18*, 243-254. doi: 10.1007/s10956-009-9145-7.

Levy, S.T., & Wilensky, U. (2008). Inventing a "mid level" to make ends meet: Reasoning between the levels of complexity. *Cognition and Instruction*, *26*(1), 1-47. doi: 10.1080/07370000701798479.

Levy, S.T., & Wilensky, U. (2011). Mining students' inquiry actions for understanding of complex systems. *Computers & Education*, *56*, 556-573. doi: 10.1016/j.compedu.2010.09.015.

Mitchell, M. (2009). *Complexity: A Guided Tour*. New York: Oxford University Press.

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement, 16,* 159-176. doi: 10.1177/014662169201600206.

National Research Council, NRC (2007). Generating and evaluating scientific evidence and explanation (Chapter 5). In, R. Duschl, H. Schweingruber, and A. Shouse (Eds.), *Taking Science to School: Learning and Teaching Science in Grades K-8* (pp. 129-167). Washington D.C.: National Academies Press.

National Science Foundation, NSF. (2011). *Empowering the Nation through Discovery and Innovation – NSF Strategic Plan for Fiscal Years (FY) 2011-2016*. Arlington, VA: National Science Foundation.

Mohan, L., Chen, J., & Anderson, C.W. (2009). Developing a multi-year learning progression for carbon cycling in socio-ecological systems. *Journal of Research in Science Teaching*, *46*(6), 675-698. doi: 10.1002/tea.20314

Program for International Student Assessment, PISA. (2006). *PISA Released Items – Science*. Paris: OECD. Retrieved from http://www.oecd.org/dataoecd/13/33/38709385.pdf on Oct 14, 2011.

Penner, D.E. (2000). Explaining systems: Investigating middle school students' understanding of emergent phenomena. *Journal of Research in Science Teaching*, *37*(8), 784-806. doi: 10.1002/1098-2736(200010)37:8<784::AID-TEA3>3.0.CO;2-E.

Resnick, M. (1996). Beyond the centralized mindset. *Journal of the Learning Sciences, 5,* 1–22. doi: 10.1207/s15327809jls0501_1.

Riess, W., & Mischo, C. (2010). Promoting systems thinking through biology lessons. *International Journal of Science Education*, *32*(6), 705-725. doi: 10.1080/09500690902769946.

Sewell, A. (2002). Constructivism and student misconceptions. Why every teacher needs to know about them. *Australian Science Teachers Journal, 48*(4), 24-28.

Songer, N.B., Kelcey, B., & Gotwals, A.W. (2009). How and when does complex reasoning occur? Empirically driven development of a learning progression focused on complex reasoning about biodiversity. *Journal of Research in Science Teaching*, *46*(6), 610-631. doi: 10.1002/tea.20313.

Wilensky, U., & Reisman, K. (2006). Thinking like a wolf, a sheep, or a firefly: Learning biology through constructing and testing computational theories-an embodied modeling approach. *Cognition and Instruction*, *24*(2), 171-209. doi: 10.1207/s1532690xci2402_1.

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. doi: 10.1023/A:1009421303064.

Yoon, S.A. (2008). An evolutionary approach to harnessing complex systems thinking in the science and technology classroom. *International Journal of Science Education*, *30*(1), 1-32. doi: 10.1080/09500690601101672.

Yoon, S. A. (2011). Using social network graphs as visualization tools to influence peer selection decision-making strategies to access information about complex socioscientific issues. *Journal of the Learning Sciences*, 20(4), 549-588. doi: 10.1080/10508406.2011.563655.

Acknowledgement

This material is based upon work supported by the National Science Foundation under Grant No. 1019228. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.