Journal of Leadership and Management®
Cabell's Listed
An International Journal of Strategic, Behavioral and Public Aspects of Leadership in Management
A journal of the Institute of Leadership and Management Association, PO Box 564, Douglassville PA 19518. 484-332-3331. https://www.jleadershipmanagement.org

Effects of AI-Based Model-Facilitated Learning and Stealth Assessment Framework for Developing Systems Thinking Leadership Skills

Deniz Eseryel, Aurora Consulting, deniz@eseryel.com
U. Yeliz Eseryel, East Carolina University

Online published: Jan 2026 Print published: Jan 2026 Editor: Sue Kong	Authorship Roles and Conflict of Interest Statement is on file at the Journal of Leadership and Management offices, available on request. Contact editors@jleadershipmanagement.org

Abstract

In a world increasingly characterized by dynamic complexity, effective decision-making necessitates leaders to develop systems thinking skills. This involves developing tools to understand the intricate structures of complex systems. Yet theoretical and didactic reflections on developing systems thinking are generally hard to find in the leadership literature. On the other hand, modern research on cognition and human learning, combined with emerging AI technologies, offers new possibilities for facilitating and assessing higher-order thinking skills such as systems thinking. This paper proposes an AI-based model-facilitated learning and stealth assessment framework for promoting systems thinking leadership competencies. It presents a study to demonstrate the implementation of two different intelligent scaffolding methods, (1) learning by using models and (2) learning by building models. The paper investigates these models' effectiveness in promoting systems thinking leadership skills in the context of a complex lake ecosystem. The findings suggest that while both approaches are effective in promoting learners' systems thinking skills, learning by building models becomes more effective when the level of problem complexity increases. The implications of this study are discussed in the context of integrated approaches to AI-based learning and assessment of complex problem solving.

KEYWORDS
Systems Thinking; Understanding Complex Systems; Dynamic Decision Making; Complex Problem Solving; System Dynamic Modeling; Model Based Learning; AI-Based Stealth Assessment

Introduction

Leadership in the era of artificial intelligence (AI) requires deep understanding of complex systems. AI has made systems thinking more crucial by introducing unprecedented complexity, interconnectedness, and autonomous behaviors that traditional, linear, and piece-meal analysis cannot handle (Qudrat-Ullah, 2025). As artificial intelligence becomes increasingly integrated into societal, economic, and organizational systems, it introduces new dependencies, potential biases, and feedback loops that necessitate a comprehensive, systemic approach to mitigate harmful consequences.

Recent *Future of Jobs Report* published by the World Economic Forum (2023) highlights the importance of systems thinking on top of the list of ten skills predicted to grow in importance for workers over the next seven years. A review of empirical studies also shows that systems thinking is highly correlated with higher leadership performance (Dörner, 1989; Ellis et al,, 1995; Funke, 1989; Gomez and Probst, 1987; Ossimitz, 1990, 1996; Palaima & Skaržauskienė, 2010).

Systems thinking calls for understanding complex systems. Complex systems have a large number of components that have dynamic interrelationships among them. These components produce synergies that are not easily foreseen by the observer. Complex systems involve multiple interconnected levels that operate at multiple time scales. For instance, in order for a leader to make an ecological decision about the impact of building housing around a lake, they must *understand the complex lake ecosystem*, in other words, the complex interplay among the biotic components, such as plants, animals, and microorganisms, and the physical and chemical interactions of abiotic components, such as sunlight, water, temperature, soil, rocks, etc.

Numerous studies have highlighted various challenges in comprehending concepts essential for understanding complex systems across diverse fields (see, for instance, Dörner, 1996; Hmelo-Silver & Azevedo, 2006; Jacobson, 2000; Kozma, 2000; Milrad, Spector, & Davidsen, 2002). Reasoning about a complex system, due to its numerous interconnected components, demands a significant amount of working memory resources and often leads to counterintuitive conclusions. (Feltovich, Coulson, & Spiro, 2001; Narayanan & Hegarty, 1998). Thus, it is no surprise that extant research (e.g., Dörner, 1996; Funke, 1991; Hmelo-Silver & Pfeffer, 2004) shows that most people perceive complex systems as collections of parts, lacking a deep understanding of the dynamic interrelationships among system components and the overall functioning of the system. Humans generally lack the ability to provide causal and structural explanations, nor can they anticipate and explain changes in underlying causes and structures. (Dörner, 1996; Feltovich, Spiro, Coulson, & Feltovich, 1996). Hence, one of the main challenges that largely hinders the development of systems thinking leadership skills is the lack of ability to build a mental model of a complex system (Eseryel & Law, 2012b; Frensch & Funke, 1995; Seel, 2006).

Despite its importance, little is known about how to assess and facilitate the development of systems thinking leadership skills. Theoretical and didactic reflections on developing systems thinking are generally hard to find in the leadership literature (Ossmitz, 2000). The purpose of this paper is to describe an AI-based model-facilitated learning and stealth assessment framework to address this gap. Following the description of the framework, a study is presented that demonstrates two approaches of AI-based model-facilitated learning and stealth assessment framework, (i) learning from models and (ii) learning by building models and to investigate their effectiveness in facilitating leadership students' development of systems thinking skills in the context of dynamic decision-making regarding a complex lake eco-system.

AI-Based Model-Facilitated Learning and Stealth Assessment Framework

There is evidence to suggest that model-facilitated learning could be effective in promoting deeper learning and understanding of complex systems (Clement & Rea-Ramirez, 2008; Gibbons, 2001, 2003; Jonassen, Strobel, & Gottdenker, 2005; Milrad, Spector, & Davidsen, 2002; Seel, 2003; Spector, 2003). While there are different interpretations of model-facilitated learning, our previous investigations show that modeling techniques adapted from the field of system dynamics has the most promise to be used as a basis of both facilitating and assessing the development of systems thinking leadership skills (Eseryel, Ifenthaler, & Ge, 2013; Eseryel & Law, 2010).

Founded by Forrester (1961) as a way to model complex business and organizational processes to aid dynamic managerial decision-making, system dynamics, also represents a way to support deep learning of complex systems (Senge, 1990; Sterman, 1994). System dynamics professionals employ two primary tools to model the dynamic feedback relationships among the components of a complex system: (1) *causal influence diagrams* and (2) *stock-and-flow diagrams*.

As illustrated in Figure 1, a causal influence diagram consists of arrows denoting the causal links among system variables (Eseryel, 2015). Each causal link is assigned a polarity, either positive (+) or negative (-), to signify the direction of change in the dependent variable when the independent variable changes. A positive link signifies a direct relationship, meaning that as the cause increases, the effect also increases, and vice versa. For example, in Figure 1, an increase in the number of contagious people means an increase in the number of incubating mosquitos above what it would otherwise have been. On the other hand, a negative link indicates an inverse causal relationship; in other words, if the cause increases the value of the effect decreases or vice versa. For instance, in Figure 1, an increase in the number of incubating people means the number of susceptible people will fall below what it would otherwise have been.

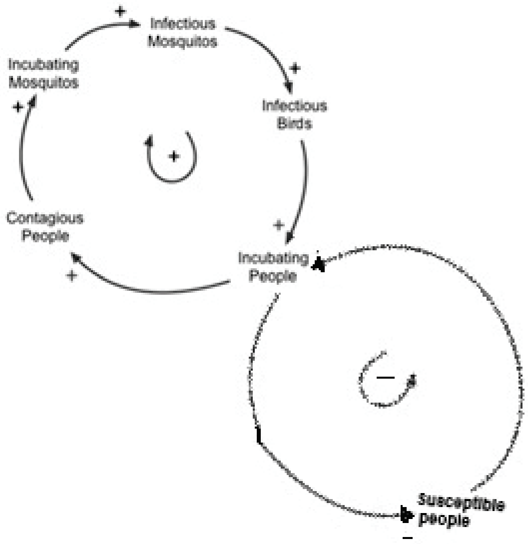


Figure 1. Sample causal influence diagram of an epidemic

In a causal influence diagram, important feedback loops are emphasized by a loop identifier that indicates whether the loop is a positive (reinforcing) or negative (balancing) feedback loop. For instance, in Figure 1, there are two primary feedback loops; the first one on the left is a positive feedback loop while the other one on the right is a negative feedback loop.

In a positive feedback loop, a variable continuously reinforces its own growth or collapse by continuously feeding back upon itself. Several familiar phrases, such as the snowball effect or the vicious cycle, characterize the phenomenon of positive feedback: a worsening of one element in a causal chain brings about further degradation of the element; conversely, positive changes in a system element trigger further improvement. On the other hand, negative feedback loop is characterized by a goal-directed behavior. If the current value of the variable of interest exceeds the desired level, the loop structure decreases its value, while if it falls below the desired level, the loop structure increases its value. Terms like self-governing, self-regulating, self-equilibrating, homeostatic, or adaptive all suggest the presence of a goal. These terms define negative feedback loops or systems. When a positive and a negative loop are combined, as in Figure 1, a variety of patterns are possible. For instance, it's possible that a positive feedback loop causes early exponential growth, but then, after a delay, a negative feedback loop takes over and dominates the system's behavior.

Causal influence diagrams are well-suited to represent the interdependencies and feedback processes of a complex problem situation. A general advantage of a causal influence diagram is that it supports a holistic view of a complex and dynamic system in a single figure represented on a single page or screen. Such representations address a common deficiency in human reasoning, namely, the tendency to ignore significant portions of a complex system (Dörner, 1980; Senge, 1990). As Figure 1 indicates, delayed effects can also be represented, which is an additional challenge; all too often, humans expect to see nearly instantaneous effects of a decision or action but real systems often involve significant delays.

One of the most significant limitations of causal diagrams is their inability to represent the stock and flow structure of systems, which are the two fundamental concepts in system dynamics theory. Hence, following the creation of a causal influence diagram, system dynamicists typically transform it into a stock-and-flow model, which include mathematical equations to represent the stocks (i.e., accumulators), flow rates, variables, and any constraints that may be assumed to govern the system (Figure 2). In this way, the simpler causal influence diagram is elaborated and transformed into the basis for a mathematically driven simulation model that can be manipulated to test overall system behavior when certain variables in the model are changed.

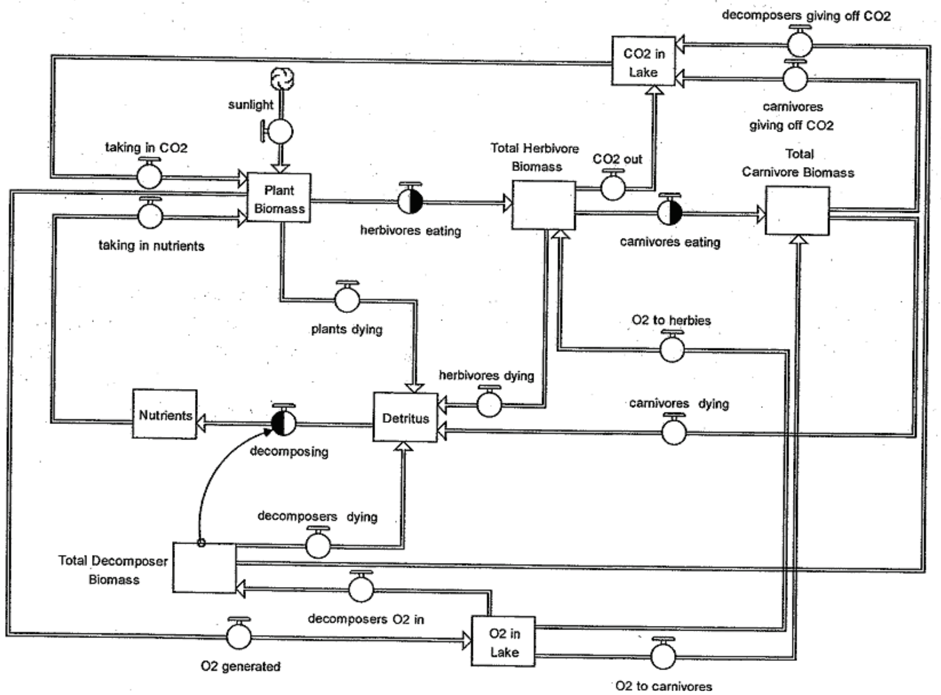


Figure 2. Stock-and-flow model of the ecosystem of Lake Mirabile that guide the Food Chain

A mathematically elaborated stock-and-flow diagram can serve as the basis for an executable simulation, with the help of computer-based systems such as STELLA (Steed, 1992), that shows how all of the components change over time or how the behavior changes when certain critical variables are manipulated. What cannot be represented in a causal influence diagram are the effects of non-linear relationships among system components, but stock-and-flow diagrams that are mathematically based can do just that. Understanding the impact of non-linear relationships in a complex system is a particular shortcoming of human reasoning. Stock-and-flow models can address that weakness in human decision-making and policy formulation and support learning about complex systems. Stock-and-flow models inform a number of so-called management flight simulators that are designed to allow users to manipulate one or more critical components in order to achieve a desired outcome, such as a stable and sustainable population growth (Davidsen, 1996; Sterman, 1994). When the critical components are changed, the behavior of the system changes, and users can see the consequences of making a particular decision or following a particular policy.

In the proposed AI-based model-facilitated learning and assessment framework, the tools developed by system dynamicists could be used in two basic ways to support the development of systems thinking skills: (1) learning from system dynamics models; and (2) learning by creating system dynamics models (Alessi, 2000; de Jong, 2006b; Spector, 2000). Either type of model (causal influence or stock-and-flow) can be used to support either approach.

In *learning from system dynamics models*, an executable stock-and-flow diagram can be used to help leaders identify which variables seem to have a significant impact on system behavior in various circumstances. Leaders' overall learning process revolves around exploring a model by altering the values of input variables and observing the resulting values of output variables. In this process, they encounter the rules of the simulated domain or uncover certain aspects of these rules (de Jong, 2006a). By changing one or more variables and observing system behavior, a learner can develop such an understanding. In addition, stock and flow diagrams can help learners understand the nature and extent of delayed effects. Moreover, by allowing learners to manipulate one or more key variables, a stock-and-flow model can help learners realize the impact of non-linear relationships on system behavior; such effects are often counter-intuitive as humans are more accustomed to reasoning about linear relationships.

In *learning by building system dynamics models*, leaders must construct a system dynamics model that can be simulated to replicate phenomena observed in a real system. Students can be provided with a description of a complex system or a scientific problem situation and asked to create either a causal loop diagram or, for more sophisticated learners, a stock-and-flow model of a certain phenomenon. The primary objective of a leader is to construct a model in a manner that closely resembles the behavior of a theoretical model or a real-world phenomenon. (van Joolingen, de Jong, Lazonder, Savelsbergh, & Manlove, 2005; White, 1993). This approach aligns with the fundamental concepts of constructionism (Harel & Papert, 1991; Kafai, 2006), of which the main focus is on "knowledge construction that takes place when students are engaged in building objects" (Kafai & Resnick, 1996, p.2).

Learning by building system dynamics models can be a highly engaging means of promoting model-facilitated learning (learning what happens, when, and why). It is important to realize that one cannot simply expect those new to a complex domain with little or no knowledge of system dynamics to begin creating very good causal influence or stock-and-flow diagrams. However, one can help learners develop the capacity to create meaningful representations of complex systems by a graduated set of activities that involve active modeling. Asking learners to transform a text-based description of a complex system or problematic situation into a causal influence diagram could be an initial step. Such an initial step need not involve any of the language of causal influence diagrams or system dynamics. Rather, it is possible to simply present the problem situation, ask learners to indicate the key factors influencing the situation, describe each factor, indicate how those factors are related, and describe those relationships. That is in fact the essence of an annotated causal influence diagram and can be depicted as such, either by the learner or by a computer-based system. The learner is then in a position to reflect upon the diagram and compare it with diagrams created by others. This would be a simple learning by modeling activity.

With learners who know something about stock-and-flow diagrams, a relatively simple learning by modeling activity would be to present a partially complete stock-and-flow model along with data about system behavior. Learners can then be asked to provide the missing parts of the model in order to account for the reported behavior. A more challenging activity would be to provide the students with a robust causal influence diagram and data about actual system behavior, and have the learner create a stock-and-flow model that when executed provides data consistent actual data.

AI-Based Stealth Assessment of the Development of Systems Thinking Skills

In addition to serve as a basis for model-facilitated learning, system dynamic tools can also support assessment of the development of systems thinking leadership skills. For instance, there is growing literature on the use of causal influence diagrams as an alternative form of assessment for complex problem solving competencies (Eseryel, 2006; Eseryel et al., 2013; Spector & Koszalka, 2004). To construct a causal influence diagram for a complex problem, one must identify the problem variables that influence the problem state and the causal relationships between these variables. Additionally, it's crucial to articulate the various causes that affect a problem state, the available solution approaches, and the trade-offs associated with each solution approach. As such, causal influence diagrams offer affordances that make them an ideal tool for eliciting learners' structural knowledge and causal reasoning. These diagrams present new, authentic, and unencountered scenario-based questions that require learners to make predictions about future events or draw inferences about past occurrences. (Dörner, Kreuzig, Reither, & Stäudel, 1983; Jonassen, 2000, 2004; Jonassen & Cho, 2008; Jonassen & Wang, 1993). Therefore, causal influence diagrams are uniquely suited to elicit and continuously track a learner's mental model progression during model-facilitated learning. Continuous progression of a learner's causal influence diagram of a complex system can be compared with that of a domain expert to assess whether the instructional intervention is facilitating desired conceptual changes or whether particular misconceptions in students' mental models are preventing their learning. Figure 3 depicts this assessment framework.

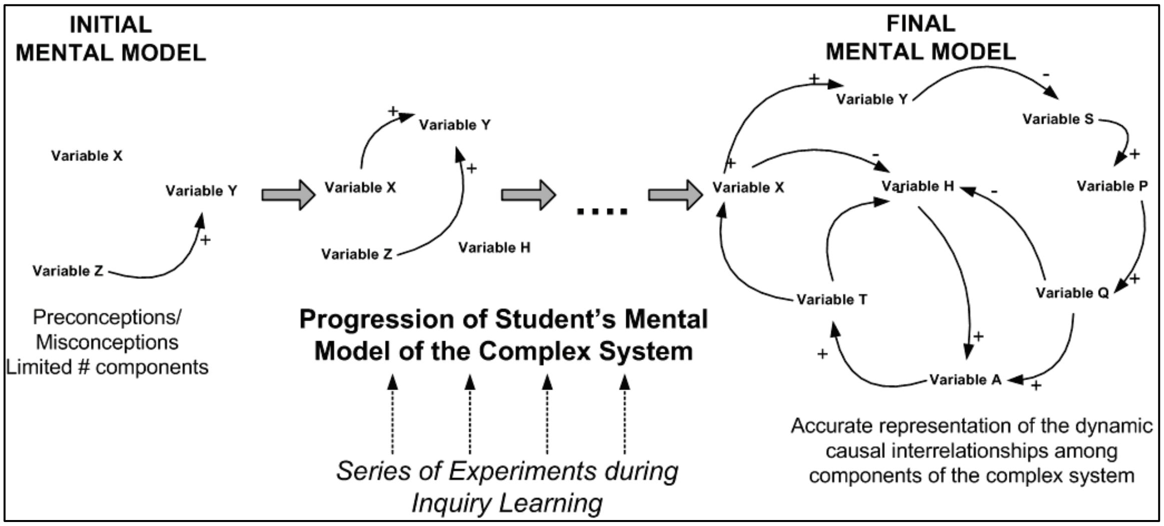


Figure 3. Model-facilitated learning and embedded assessment framework

Based on previous research, this assessment framework assumes that novices initially have preconceptions or misconceptions about the problem space (Eseryel & Law, 2012b; Ifenthaler, Masduki, & Seel, 2011; Snow, 1990). As one's expertise level increases, these initial assumptions are gradually replaced by more comprehensive and causal explanations (Berliner, 2002; Eseryel et al., 2013; Seel, Al-Diban, & Blumschein, 2000; Snow, 1990). Hence, in the context of complex problem-solving, effective learning can be described as a process that facilitates the transition of learners' problem spaces from the state of preconceptions or misconceptions to the state of comprehensive, causal explanations (see Figure 3).

To make this framework a reliable assessment tool, we need to create a similarity metric that compares how novice problem spaces evolve over time through instructional interventions. This metric should help us determine whether the problem conceptualizations of advanced learners or experts are similar to or different from those of novices, allowing us to evaluate the effectiveness of any specific instructional strategy.. In studies with different complex domains, it was found that a similarity metric to compare two causal influence diagrams should have three measures: (1) surface similarity; (2) semantic similarity; and (3) structural similarity (Eseryel, 2006; Spector, Christensen, Sioutine, & McCormack, 2001; Spector & Koszalka, 2004). Recently, Pirnay-Dummer, Ifenthaler, and Spector (2010) devised a computer-based tool, called HIMATT (Highly Integrated Model Assessment Technology and Tools), for comparing the similarity of sets of causal influence diagrams based on an expanded version of these three measures. By using specific automated comparison algorithms, HIMATT calculates similarities between a given pair of frequencies f1 (e.g. expert solution) and f2 (e.g. learner solution) across four structural (surface, graphical, structural, and gamma matching) and three semantic (concept, propositional, and balanced propositional matching) measures, which results in a measure of $0 \leq s \leq 1$, where $s = 0$ is complete exclusion and $s = 1$ is identity (Ifenthaler, 2010a, 2010b; Pirnay-Dummer & Ifenthaler, 2010; Pirnay-Dummer et al., 2010).

Reliability scores exist for the single measures integrated into HIMATT (see Pirnay-Dummer et al., 2010). They range from $r = .79$ to $r = .94$ and are tested for the semantic and structural measures separately and across different knowledge domains. Validity scores are also reported separately for the structural and semantic measures. Convergent validity lies between $r = .71$ and $r = .91$ for semantic comparison measures and between $r = .48$ and $r = .79$ for structural comparison measures (see Pirnay-Dummer et al., 2010). In a recent study by Eseryel and colleagues (2013), the assessment framework and the accompanying HIMATT analysis method was validated against an established research and analysis method for complex problem solving, namely the protocol analysis method (cf. Ericsson & Simon, 1993).

The power of this assessment method is that it can be seamlessly integrated in an AI-based model-facilitated learning environment and act as a *stealth* assessment (cf., Shute, 2015). It continuously tracks students' mental model progressions and compares individual student progress to the reference expert model, making it an integral part of the learning environment. Such immediate and continuous feedback could help instructors in introducing just-in-time information to address apparent misconceptions or to modify the design of the instructional intervention to bring about desired changes in learners' mental models. In addition, it is possible to program automated AI-based scaffolds, for instance, in the form of a question-prompt, that would provide instant guidance to support learners' understanding of a complex system.

Purpose of the Study

The purpose of this study was to demonstrate the utility of the AI-based model-facilitated learning and stealth assessment framework described above and to investigate the effects of its different approaches, namely (1) *learning from system dynamics models* (a.k.a. model-using); and (2) *learning by building system dynamics models* (a.k.a. model-building), on leadership students' dynamic decision-making of a complex lake ecosystem. Thus, the following research question was posed: *Given different levels of problem complexity, is learning-from-system-dynamics models more effective than learning-by-building-system-dynamics-models in a model-facilitated learning environment?*

Method

Participants

Participants included 273 mixed-ability ninth-grade students in the leadership track from an ethnically and economically diverse rural high school in the midwest of the United States. This school's core value is a focus on leadership and runs a program that creates a safe environment for students to learn and practice leadership. The school's mission is geared toward every student developing personal leadership, interpersonal leadership, team leadership, and organizational leadership.

Each of the 273 students were randomly assigned to one of ten classes. Out of the ten classes, five were randomly assigned to the experimental condition, which involved using a model (i.e., model-using group), while the other five were randomly assigned to the control condition, which involved building a model (i.e., model-building group). Out of the 237 students who provided both consent and parental assent forms, 118 were assigned to the experimental group, while 119 were assigned to the control group. There were 50.63% (n=120) males and 49.36% (n=117) females.

Materials

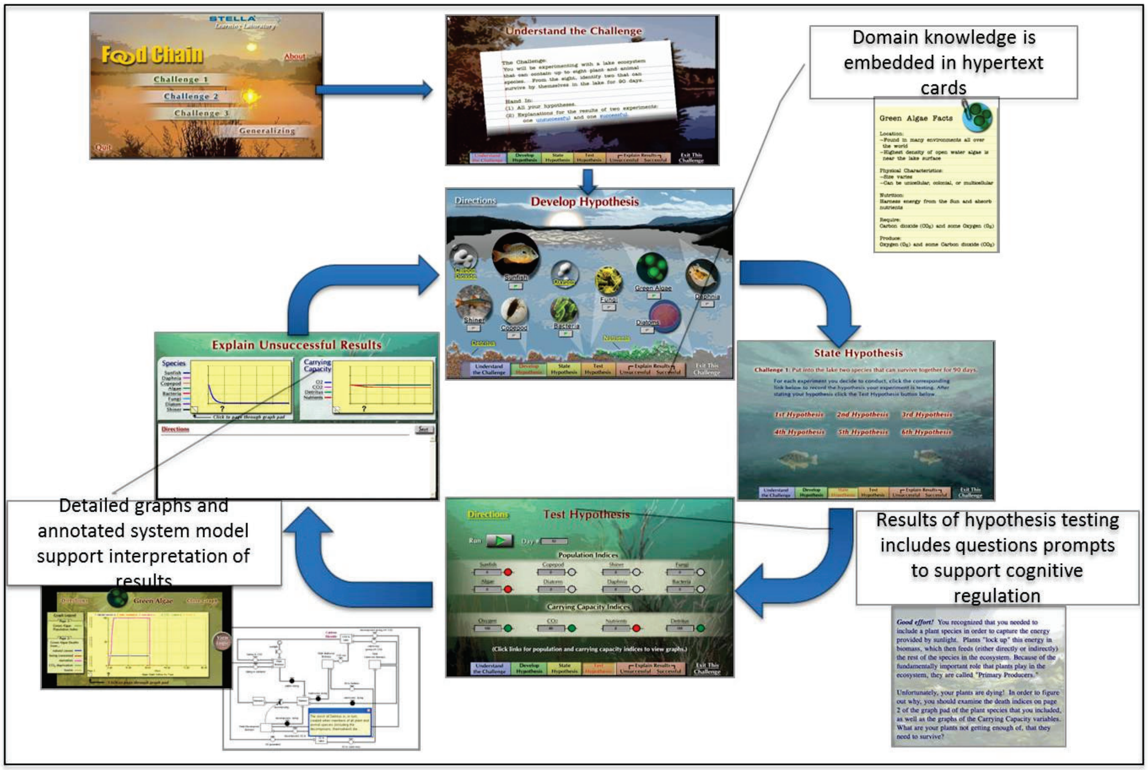
For the purposes of this study, a model-facilitated inquiry learning environment called *Food Chain* was utilized. *Food Chain* was designed to facilitate students' deep learning of the complex ecosystem of Lake Mirabile, an hypothetical lake that contains eight species, two from each of the four trophic levels: sunfish and shiners (the carnivores); copepods and daphnia (herbivores); green algae and diatoms (the primary producers); bacteria and fungi (the decomposers). Food Chain was built on the STELLA system dynamics modeling platform developed by the isee systems. The stock-and-flow model (see Figure 2) guiding the system simulation is based on the expert domain model, which depicts the dynamic interrelationships among the various species of the lake, including additional environmental factors. Students can also develop their own stock-and-flow diagrams related to a given problem scenario in the STELLA platform and simulate what happens in Lake Mirabile. Hence, *Food Chain* can support both learning-from-models and learning-by-modeling approaches.

When students enter the Food Chain environment, they are provided with a series of problems in increasing complexity similar to the idea of model progression also found in earlier work by White and Frederiksen (1990). For instance, the first inquiry challenge in Food Chain asks students to find two species out of eight that can live together in a lake for 90 days. Hence, the main task of the learners is to infer the characteristics of the model underlying the simulation in Figure 2 by first discovering that a higher-order categorization of the species is possible as carnivores, herbivores, primary producers, and the decomposers; and then discovering the inter-dependencies among these categories within the lake ecosystem.

In the *learning-from-models* mode, *Food Chain* supports students while they carry out a sequence of activities that correspond to the steps in the scientific inquiry method (Figure 4). Each challenge in *Food Chain* is organized around the steps in the scientific method. After understanding the challenge, the students are guided to, in sequence, develop hypothesis; state hypothesis; test hypothesis; and explain the results. In each step, Food Chain scaffolds the students. For instance, in the develop hypothesis step, when the students click on any species a hypertext card opens up that provides requisite domain knowledge about that species such as the properties of its preferred location, its physical characteristics, nutritional requirements, and the atmospheric gases (i.e., carbon dioxide and oxygen) that it produces and requires (see Figure 4). In the state hypothesis step, worksheets are provided for the students to write their formal hypothesis in the correct form along with their justification explaining the rationale behind their hypothesis. Hypotheses of the form, "We hypothesize that sunfish and daphnia will be able to survive in the lake for 90 days because...." are expected.

During the test hypothesis step, students test their hypotheses via simulating. Only the species that have been clicked on in the develop hypothesis step are considered to be in the lake in the ensuing simulation. When the students click on the Run button they start to visualize the changes in the lake ecosystem as Food Chain simulates what happens during the 90 days when the selected two species are put in the lake together. Status indicator lamps for each species in the lake will initially glow green to indicate that they are being included in the simulation, and that their initial number is within the normal bounds. As the simulation progresses, these lamps may begin to glow yellow, indicating that the associated population has either grown large or small enough to be considered at risk. Should a lamp begin to go red and flash that species and/or carrying capacity variable is either at peril of disappearing from the ecosystem or has achieved unsustainable proportions.

At the end of the simulation, changes in the population indices of the species and in carrying capacity indices of oxygen, carbon dioxide, nutrients, and detritus are provided by various graphs and charts, which show the causes of death of the species, such as natural causes, starvation, and asphyxiation (see Figure 4). Interactions with these charts and graphs help students regulate their cognition and discover the interrelationships among the different species. During the explain the results step, the students are provided with a worksheet to explain their findings, articulate their understanding of what had happened during the experiment, and state which hypothesis should be tested next.



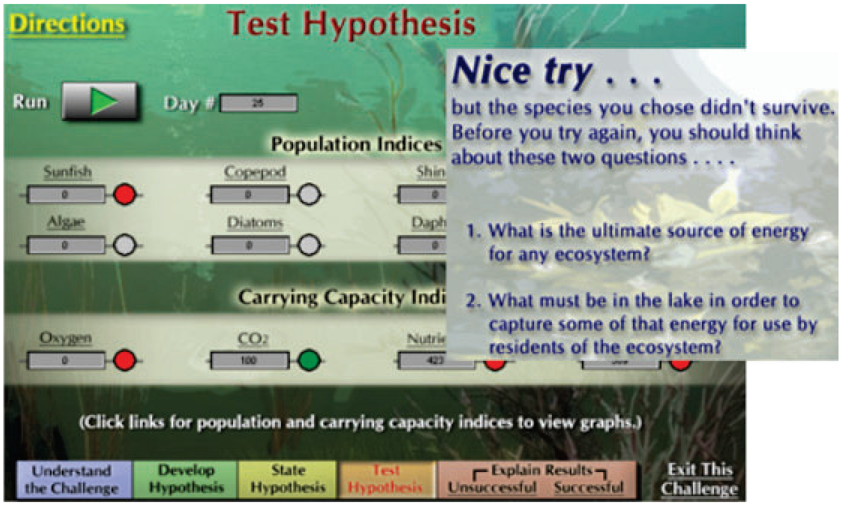
Domain knowledge is embedded in hypertext cards

Detailed graphs and annotated system model support interpretation of results

Results of hypothesis testing includes questions prompts to support cognitive regulation

Figure 4. Food Chain model-facilitated inquiry learning environment

Finally, *Food Chain* scaffold students' inquiry process by intelligent question prompts. As students construct their hypothesis about possible interrelationships among constituents of the complex lake ecosystem, the embedded assessment system automatically compares student hypothesis with the expert stock-and-flow model (see Figure 2) running in the background of the simulation. After the students run the hypothesis they see the consequences of their assumptions and what happens in the lake ecosystem through various graphs and other visuals. As the students are guided to explain these findings to construct a new hypothesis they receive an intelligent question-prompt that stems from the analysis of the embedded assessment system asking the students to consider the interrelationships between two system components selected based on how their hypothesis compare with the expert model of the lake ecosystem. For instance, if a student included a plant and an herbivore in their hypothesis, a question prompt is dynamically generated by *Food Chain* to reinforce the requirement of having a primary producer in the lake ecosystem and guide the students further to think about what was needed for the plants to survive. An example of a dynamic question prompt is shown in Figure 5.

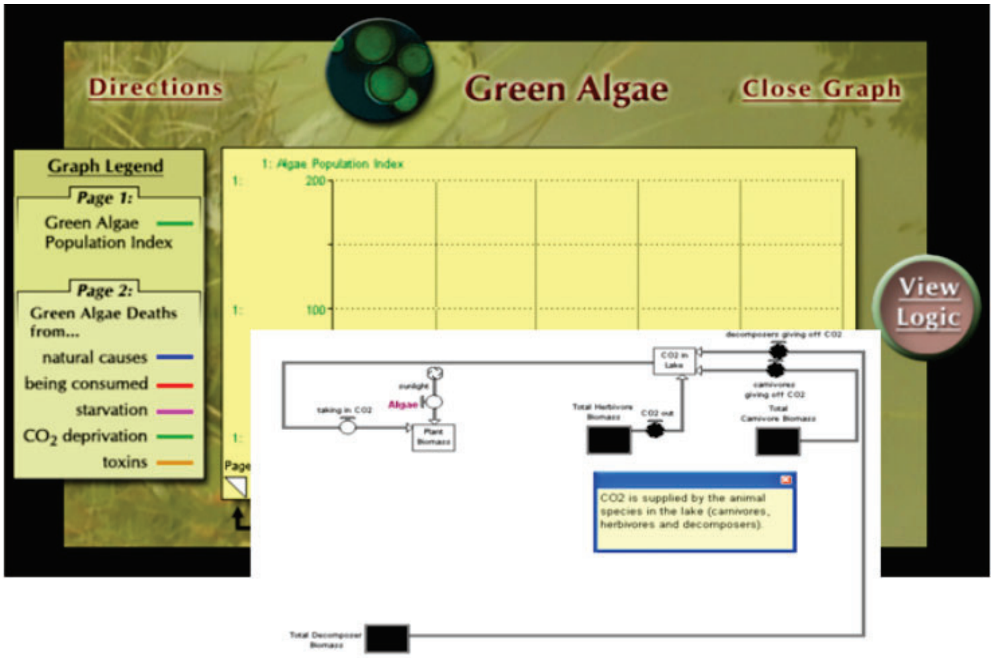


Nice try . . .
but the species you chose didn't survive. Before you try again, you should think about these two questions . . . .

1. What is the ultimate source of energy for any ecosystem?
2. What must be in the lake in order to capture some of that energy for use by residents of the ecosystem?

Figure 5. A sample question prompt received by participants

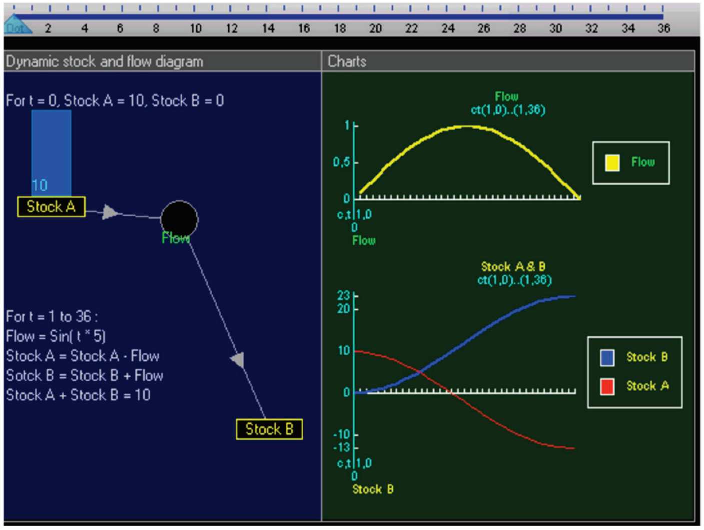
In the *learning-from-models* mode, following their hypothesis-testing while students are trying to interpret the results to construct a new hypothesis, they are given access to the stock-and-flow diagram that is similar to that of the domain experts underlying the system simulation. The stock-and-flow diagram is revealed piece-by-piece as students click on the *enter* tab on their keyboard; each piece is accompanied by an annotation that explains the relationships among the system constituents like telling a digital story of the complex ecosystem of Lake Mirabile (see Figure 6). In this way, the expert model underlying the simulation is made transparent to the students in an effort to support the cognitive regulation of the students during inquiry learning.



CO2 is supplied by the animal species in the lake (carnivores, herbivores and decomposers).

Figure 6. Dynamic model progression scaffold received by participants in the learning-from-models mode

In the *learning-by-modeling* mode, students go through the steps of the scientific inquiry as in the learning-from-models mode. However, they do not receive the dynamic model progression to scaffold their cognitive regulation. Instead, students engage in an iterative process of constructing and revising their own stock-and-flow models of the complex lake ecosystem to explain the phenomenon they are investigating during scientific inquiry and testing their model through the STELLA simulation interface (Figure 7). Therefore, the environment provides opportunities to enhance the transparency of the learning process. Hypotheses become visible as models or components of models, and students' predictions can be displayed as model outputs. Additionally, the validity of models is continuously and automatically evaluated using HIMATT analysis tools in comparison to domain-expert models (cf. Eseryel et al., 2013).



For t = 0, Stock A = 10, Stock B = 0

For t = 1 to 36 :
Flow = Sin(t * 5)
Stock A = Stock A - Flow
Sotck B = Stock B + Flow
Stock A + Stock B = 10

Figure 7. Model-building and testing interface using STELLA simulation interface

Procedure

Participants in both study conditions interacted with the *Food Chain* model-facilitated inquiry-learning environment for three weeks. In the first week, the participants were introduced to the Food Chain model-facilitated inquiry-learning and embedded assessment system and were asked to design their own experiments and test their hypotheses to answer a very simple complex problem (challenge# 1): Which two species (out of eight) can survive in Lake Mirabile by themselves for 90 days? This first week was intended to familiarize students with the simulation environment and to teach about scientific reasoning processes of hypothesis generation and testing. During this initial run, students in both conditions had access to domain-specific knowledge related to the species in Lake Mirabile but they did not engage in any model-based activity. In other words, students in the learning-from-models condition did not see the underlying stock-and-flow diagram depicting the dynamic interrelationships among the constituents of the lake ecosystem underlying the simulation; students in the learning-by-modeling condition did not engage in any system modeling activity. At the end of each inquiry-cycle, students only received a text-based verification feedback of whether or not their hypotheses were correct and the intelligent question-prompt to scaffold their cognitive regulation.

In the second week, participants tackled another problem scenario with mid-level complexity (challenge# 2), in which they were asked to identify the smallest number of species that will enable Sunfish to survive for 90 days in Lake Mirabile. Their entire hypotheses, experimental designs, and elaborated reports of the findings were collected in the system.

During the third week, participants tackled a very complex and ill-structured problem scenario (challenge# 3), which called for environmental policy-making. In this problem scenario, students were asked to play the role of an environmental scientist and evaluate the proposal to build 100 new houses on the shoreline at Lake Mirabile from an environmental impact standpoint. All of their hypothesis, results of their experiments, and their elaborated report of recommendations and assessment were collected in the system.

During the second and third weeks, the only difference between the two study conditions was the type of model-facilitated learning activity with which the students engaged. Students in the learning-from-models condition had access to the underlying stock-and-flow diagram depicting the dynamic interrelationships among the constituents of the lake ecosystem underlying the simulation; students in the learning-by-modeling condition engaged with constructing, testing, and revising their stock-and-flow models of the lake ecosystem under investigation.

Data Analysis

Stealth assessments within a learning environment are considered a viable method for drawing inferences about learners' behaviors (Chung & Baker, 2003). In the *Food Chain*, students engaged in inquiry-based learning to solve three progressively complex ecology problems. For each problem, they were asked to go through the inquiry process at least six times. Each inquiry cycle in Food Chain simulation involved the following steps. First, students individually developed a hypothesis using their prior knowledge and information given by the software. Then, they observed the results and analyzed the charts generated by the simulation environment for each variable. Finally, they explained the results. A computer-generated intelligent question prompt and an annotated dynamic model progression were provided to the students in the learning-from-models condition to interpret their findings and revise their hypotheses while the students in the learning-by-modeling condition only received computer-generated intelligent question-prompts and were engaged with constructing, testing, and revising their stock-and-flow models of the lake ecosystem under investigation.

All of students' activities and by-products of these activities were stored in the system and the HIMATT (Pirnay-Dummer, Ifenthaler, & Spector, 2010) analysis function was applied. The automated analysis function produces measures, which range from surface-oriented structural comparisons to integrated semantic similarity measures. Those measures include four *structural* (surface, graphical, structural, and gamma matching, also referred as SFM, GRM, STM, and GAM) and three *semantic* (concept, propositional, and balanced semantic matching, also referred as CCM, PPM, & BSM) indicators (Figure 8).

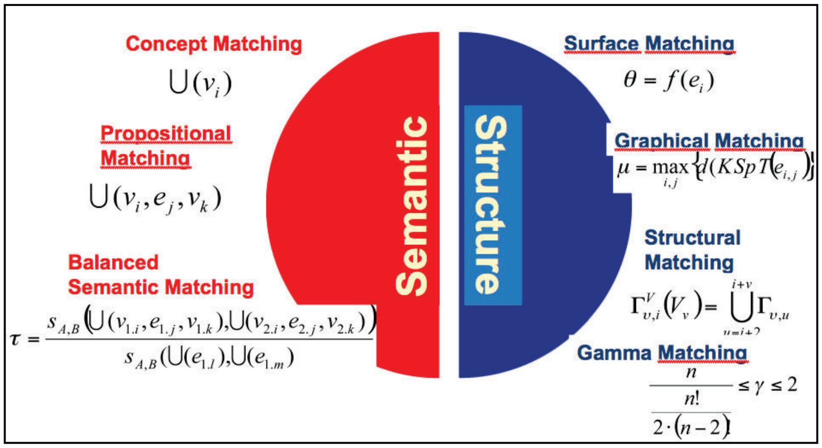


Figure 8. HIMATT measures for comparing structural and semantic similarity of student and expert models

Each of the participants' protocols was compared automatically against a reference solution. The reference solution included the stock-and-flow diagram developed by domain experts, which was used to guide the *Food Chain* model-facilitated inquiry learning environment. HIMATT uses specific automated comparison algorithms to calculate similarities between a given pair of frequencies or sets of properties. The similarity index s for each of the seven measures results in a measure of $0 \le s \le 1$, where $s = 0$ is complete exclusion and $s = 1$ is complete similarity. Reliability scores for the HIMATT measures range from $r = .79$ to $r = .94$ (Pirnay-Dummer, et al., 2010). Convergent validity scores lies between $r = .71$ and $r = .91$ for semantic comparison measures and between $r = .48$ and $r = .79$ for structural comparison measures (Pirnay-Dummer, et al., 2010).

Based on the HIMATT measures, the dependent variable *complex learning* was identified as a combination of semantic and structural properties: PREP = (CCM+PPM)+(SFM*GRM*STM). The variable is reported for the pre-test (PREP_{pre}) and post-test ($\text{PREP}_{\text{post}}$) results. Based on previous research using the HIMATT measures, the aggregation of structural and semantic measures best reflects individual's problem representation as it includes strong weights of semantic complexity; however, does not neglect the overall structural components (Eseryel et al., 2013).

Repeated measures ANOVA analysis was conducted to examine change on *complex learning* measure both after challenge 1, after challenge 2, and after challenge 3 in the *Food Chain*. All the analyses were conducted with the Statistical Package for the Social Sciences (SPSS).

Results

Table 1 summarizes the descriptive statistics for the dependent variable (complex learning) for both model-using (MU) and model-building (MB) groups along the three challenges: (T1) simpler complex problem where none of the groups engaged in model-using or model-building activities but only went through the scientific inquiry process supported by the *Food Chain* simulation and intelligent question-prompts; (T2) a problem with mid-level complexity, where MU and MB groups engaged with corresponding model-facilitated activities; and (T3) very complex, environmental policy analysis problem, where MU and MB groups engaged with corresponding model-facilitated activities.

Table 1. Descriptive statistics for the complex learning measure

	Model-Using (MU) Group Mean (SD)			*Model-Building (MB) Group* Mean (SD)		
	T1	T2	T3	T1	T2	T3
Complex Learning	1.44 (0.89)	3.45 (0.49)	4.33 (1.16)	1.64 (1.04)	3.75 (0.89)	6.92 (1.30)

As seen in Table 1, participants in the MB condition had slightly higher mean scores than the participants in the MU condition in complex learning after the first challenge in the Food Chain, during none of the groups engaged with in model-using or model-building activities. However, the repeated measures ANOVA analyses did not indicate significant differences between the two groups in this dependent variable. This result indicated that the participants in both conditions were comparably on equal basis when they completed the first challenge following only the simulation-based inquiry learning.

The results of the multiple repeated measures (three inquiry challenges) ANOVA with between-subject factors (2 groups) for complex learning scores revealed that there were statistically significant differences between the two groups throughout the three challenges [F (2, 237) = 42.66, $p < .01$, η2 = .53]. In the post-hoc analysis, the third challenge was attributed to the significant interaction effect for the time and group F (2, 237) = 49.86, $p < .01$, η2 = .57. As shown in Table 3, MB group scores were higher in the third challenge than the MU group.

These findings supported the hypothesis that the proposed AI-based model-facilitated learning and assessment framework is effective in facilitating the development of systems thinking skills. Both model-building and model-using were effective in facilitating deep understanding of a complex lake ecology system. However, in highly complex problem-solving tasks, like environmental policy decision-making, model-building was more effective than model-using.

Discussion

In this era of rapid technological progress, AI has become a pivotal force in transforming organizational dynamics, particularly in the field of leadership. Recent leadership reviews suggest that as artificial intelligence becomes an integral part of workflows, decisions can no longer be considered isolated managerial actions; hence, artificial intelligence is increasingly conceptualized as a force multiplier of system complexity, rather than just a decision-making tool (Aziz et al., 2025). Thus, leaders must comprehend how algorithmic outputs impact organizational, social, and ethical systems. In this context, AI shifts leadership practice from linear problem-solving to holistic system sensemaking (Carlucci & Skaržauskienė, 2010). In other words, artificial intelligence emphasizes the crucial necessity of systems thinking skills for effective leadership that can deal with complex problem-solving situations in today's world that increasingly calls for dynamic decision-making (Qudrat-Ullah, 2025). Systems thinking emphasizes interdependence, feedback loops, nonlinearity, and unintended consequences. AI systems function precisely through these dynamics, often operating at speeds and scales surpassing human cognitive capabilities.

How do leaders develop such crucial systems thinking skills? This paper proposed an AI-based model-facilitated learning and stealth assessment framework to address the gap in the leadership literature. We investigated the effects of two different approaches within the proposed framework: (1) learning from models (i.e., system model-using); and (2) learning by modeling (i.e., system model-building). The findings suggest that while both approaches are effective in facilitating the development of learners' systems thinking skill, as the problem complexity increases, model-building is more effective than model-using.

One possible explanation of this finding could stem from the important role of cognitive regulation skills during complex problem solving (cf. Azevedo, Guthrie, & Seibert, 2004; Eseryel & Law, 2012b; Pieschl, Stahl, & Bromme, 2008). In a previous study, we had found a cross-lagged association between learners' understanding of complex systems and their cognitive regulation (Eseryel & Law, 2012a); furthermore, task complexity was found as an important determinant of the level of cognitive regulation skill that was required of the learner: the more complex the task is the more demand is placed on the learner for higher-levels of cognitive regulation skills (Eseryel & Law, 2012a). Hence, the findings of this study suggest that model-building serve as a more effective cognitive regulation scaffold during learning of complex systems and supports learners' mental model transition as they piece-by-piece discovered the interdependencies among the different species in the complex lake ecosystem in Food Chain. On the other hand, learners in the model-using group may be so cognitively overloaded by the complexity of the problem that even though they had access to the expert model they had not internalized it as their own mental models to be able to effectively respond to the highly complex problem scenario. It may also be possible that learners with misconceptions kept rejecting all or the parts of the expert model since it cannot be readily accommodated into their existing mental models due to their existing misconceptions (Ifenthaler & Eseryel, 2013; Piaget, 1985; Posner, Strike, Hewsn, & Gertzog, 1982; Vosniadou, 1994). Future research studies should investigate these issues further.

References

Alessi, S. (2000). Building versus using simulations. In J. M. Spector & T. M. Anderson (Eds.), *Integrated and holistic perspectives on learning, instruction, and technology: Understanding complexity* (pp. 175-196). Dordrecht: Kluwer Academic Publishers.

Aziz, M. F., Rajesh, J. I., Jahan, F., McMurray, A., Ahmed, N., Narendran, R., & Harrison, C. (2025). AI-powered leadership: A systematic literature review. *Journal of Managerial Psychology*, 40(5), 604-630.

Azevedo, R., Guthrie, J. T., & Seibert, D. (2004). The role of self-regulated learning in fostering students' conceptual understanding of complex systems with hypermedia. *Journal of Educational Computing Research, 30*(1), 87-111. doi:10.2190/DVWX-GM1T-6THQ-5WC7

Berliner, D. C. (2002). Learning about and learning from expert teachers. *International Journal of Educational Research, 35*(5), 463–482. doi:10.1016/S0883-0355(02)00004-6

Carlucci, D., Lerro, A., Skaržauskienė, A. (2010). Managing complexity: Systems thinking as a catalyst of the organization performance. *Measuring Business Excellence*, 14(4), 49-64. doi: https://doi.org/10.1108/13683041011093758

Chung, G. K. W. K., & Baker, E. L. (2003). An exploratory study to examine the feasibility of measuring problem-solving processes using a click-through interface. *Journal of Technology, Learning and Assessment, 2*(2), 3-29.

Clement, J. J., & Rea-Ramirez, M. A. (Eds.). (2008). *Model based learning and instruction in science* (Vol. 2). NY: Springer.

Davidsen, P. I. (1996). Educational features of the system dynamics approach to modelling and simulation. *Journal of Structural Learning, 12*(4), 269-290.

de Jong, T. (2006a). Computer simulations: Technological advances in inquiry learning. *Science, 312*, 532-533.

de Jong, T. (2006b). Scaffolds for computer simulaiton based scientific discovery learning. In J. Elen & R. E. Clark (Eds.), *In dealing with complexity in learning environments* (pp. 107-128). London: Elseiver.

Dörner, D. (1980). On the difficulties people have in dealing with complexity. *Simulation & Games*(11), 87-106.

Dörner, D. (1989). *Die logik des misslingens. Strategisches denken in komplexen situationen*. Reinbek: Rowohlt.

Dörner, D. (1996). *The logic of failure: Why things go wrong and what can we do to make them right* (R. Kimber & R. Kimber, Trans.). New York: Holt.

Dörner, D., Kreuzig, H. W., Reither, F., & Stäudel, T. (1983). *Lohhausen. Vom Umgang mit Unbestimmtheit und Komplexität. [Lohhausen. On dealing with uncertainity and complexity]*. Bern: Huber.

Ellis, K., Gregory, A., Mears-Young, B.R. and Ragsdell, G. (1995). *Critical issues in systems theory and practice*. New York, NY: Plenum Press.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (revised ed.). Cambridge, MA: The MIT Press.

Eseryel, D. (2006). *Expert conceptualizations of the domain of instructional design: An investigative study on the DEEP assessment methodology for complex problem-solving outcomes.* (Doctoral Dissertation). Retrieved from Dissertations and Theses database. (UMI No. 3241853)

Eseryel, D. (2015). Causal influence diagrams. In J. M. Spector (Ed.), *Encyclopedia of educational technology*. Thousand Oaks, CA: Sage Publications, Inc.

Eseryel, D., Ifenthaler, D., & Ge, X. (2013). Validation study of a method for assessing complex ill-structured problem solving by using causal representations. *Educational Technology Research & Development, 61*(3), 443-463. doi:10.1007/s11423-013-9297-2

Eseryel, D., & Law, V. (2010). Promoting learning in complex systems: Effect of question prompts versus system dynamics model progressions as a cognitive-regulation scaffold in a simulation-based inquiry-learning environment. In *Proceedings of the 9th International Conference of the Learning Sciences*. Chicago, IL.

Eseryel, D., & Law, V. (2012a). *Complexity, cognitive regulation, and understanding of a complex science system*. Paper presented at the Annual Meeting of the American Educational Research Association, Vancouver, British Columbia.

Eseryel, D., & Law, V. (2012b). Effect of cognitive regulation in understanding complex science systems during simulation-based inquiry learning. *Technology, Instruction, Cognition, & Learning, 9*(1-2), 111-132.

Feltovich, P. J., Coulson, R. L., & Spiro, R. J. (2001). Learner's (mis)understanding of important and difficult concepts. In K. D. Forbus & P. J. Feltovich (Eds.), *Smart machines in education: The coming revolution in educational technology* (pp. 349-375). Menlo Park, CA: MIT Press.

Feltovich, P. J., Spiro, R. J., Coulson, R. L., & Feltovich, J. (1996). Collaboration with and among minds: Mastering complexity, individually and in groups. In T. Koschmann (Ed.), *CSCL: Theory and practice of an emerging paradigm* (pp. 25-44). Mahwah, NJ: Erlbaum.

Forrester, J. W. (1961). *Industrial dynamics*. Cambridge, MA: MIT Press.

Frensch, P. A., & Funke, J. (Eds.). (1995). *Complex problem solving: The European perspective*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Funke, J. (1989). *Komplexes problemlösen*. Berlin: Springer.

Funke, J. (1991). Solving complex problems: Exploration and control of complex systems. In R. J. Sternberg & P. A. Frensch (Eds.), *Complex problem solving: Principles and mechanisms* (pp. 185-222). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Gibbons, A. S. (2001). Model-centered instruction. *Journal of Structural Learning and Intelligent Systems, 14*(4), 511-540.

Gibbons, A. S. (2003). Model-centered learning and instruction: Comments on Seel (2003). *Technology, Instruction, Cognition, and Learning, 1*(3), 511-540.

Gomez, P. & Probst, G. B. J. (1987). *Vernetztes denken im management*. Bern: Die Orientierung.

Harel, I., & Papert, S. (1991). *Constructionism*. Norwood, NJ: Ablex.

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.

Ifenthaler, D. (2010a). Relational, structural, and semantic analysis of graphical representations and concept maps *Educational Technology Research and Development, 58*(1), 81-97.

Ifenthaler, D. (2010b). Scope of graphical indices in educational diagnostics. In D. Ifenthaler, P. Pirnay-Dummer, & N. M. Seel (Eds.), *Computer-based diagnostics and systematic analysis of knowledge* (pp. 213-234). New York: Springer.

Ifenthaler, D., & Eseryel, D. (2013). Facilitating complex learning by mobile augmented reality learning environments. In R. Huang, J. M. Spector, & Kinshuk (Eds.), *Reshaping learning: The frontiers of learning technologies in a global context* (pp. 415-438). New York: Springer.

Ifenthaler, D., Masduki, I., & Seel, N. M. (2011). The mystery of cognitive structure and how we can detect it. Tracking the development of cognitive structures over time. *Instructional Science, 39*(1), 41-61. doi:10.1007/s11251-009-9097-6

Jonassen, D. H. (2000). Toward a design theory of problem solving. *Educational Technology Research & Development, 48*(4), 63-85. doi:10.1007/BF02300500

Jonassen, D. H. (2004). *Learning to solve problems: An instructional design guide*. San Francisco, CA: Pfeiffer.

Jonassen, D. H., & Cho, Y. H. (2008). Externalizing mental models with mindtools. In D. Ifenthaler, P. Pirnay-Dummer, & J. M. Spector (Eds.), *Understanding models for learning and instruction. Essays in honor of Norbert M. Seel* (pp. 145-160). New York: Springer.

Jonassen, D. H., Strobel, J., & Gottdenker, J. (2005). Model building for conceptual change *Interactive Learning Environments, 13*(1-2), 15-37.

Jonassen, D. H., & Wang, S. (1993). Acquiring structural knowledge from semantically structured hypertext. *Journal of Computer-Based Instruction, 20*(1), 1-8.

Kafai, Y. B. (2006). Constructionism. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 35-47). Cambridge, UK: Cambridge University Press.

Kafai, Y. B., & Resnick, M. (1996). *Constructionism in practice: Designing, thinking, and learning in a digital world*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Milrad, M., Spector, J. M., & Davidsen, P. I. (2002). Model facilitated learning. In S. Naidu (Ed.), *Learning and teaching with technology: Principles and practices* (pp. 13-27). London: Kogan Page.

Narayanan, N. H., & Hegarty, M. (1998). On designing comprehensible interactive hypermedia manuals. *International Journal of Human-Computer Studies, 48*, 267–301.

Ossimitz, G. (1990). *Materialien zur systemdynamik*. Wien: HPT.

Ossimitz, G. (1996). *Stand und perspektiven der forschung zum systemischen denken*. Wien: Holder-Pichler-Tempsky.

Ossimitz, G. (2000). *Entwicklung systemischen denkens*. München: Profil.

Palaima, T. & Skaržauskienė, A. (2010). Systems thinking in a complex world. *Baltic Journal of Management*, 5(3), 330 – 355. http://dx.doi.org/10.1108/17465261011079749

Piaget, J. (1985). *The equilibrium of cognitive structures*. Chicago: University of Chicago Press.

Pieschl, S., Stahl, E., & Bromme, R. (2008). Epistemological beliefs and self-regulated learning with hypertext. *Metacognition & Learning, 3*, 17-37.

Pirnay-Dummer, P., & Ifenthaler, D. (2010). Automated knowledge visualization and assessment. In D. Ifenthaler, P. Pirnay-Dummer, & N. M. Seel (Eds.), *Computer-based diagnostics and systematic analysis of knowledge* (pp. 77-115). New York: Springer.

Pirnay-Dummer, P., Ifenthaler, D., & Spector, J. M. (2010). Highly integrated model assessment technology and tools. *Educational Technology Research and Development, 58*(1), 3-18. doi:10.1007/s11423-009-9119-8

Posner, G., Strike, K. A., Hewsn, P. W., & Gertzog, W. A. (1982). Accomodation of a scientific conception: Towards a theory of conceptual change. *Science Education, 66*(2), 211-227.

Qudrat-Ullah, H. (2025). Integrating Systems Thinking and AI. In H. Qudrat-Ullah (Ed.). *Navigating complexity: AI systems thinking for smarter decisions* (pp. 39-68). Cham: Springer. https://doi.org/10.1007/978-3-031-82742-6_3

Seel, N. M. (2003). Model-centered learning and instruction. *Technology, Instruction, Cognition, and Learning, 1*(1), 59-85.

Seel, N. M. (2006). Mental models and complex problem solving. In J. Elen & R. E. Clark (Eds.), *Handling complexity in learning environments: Theory and research* (pp. 43-66): Elseiver, Ltd.

Seel, N. M., Al-Diban, S., & Blumschein, P. (2000). Mental models and instructional planning. In J. M. Spector & T. M. Anderson (Eds.), *Integrated and holistic perspectives on learning, instruction and technology: Understanding complexity* (pp. 129-158). Dordrecht: Kluwer Academic Publishers.

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

Shute, V. L. (2015). Stealth assessent. In J. M. Spector (Ed.), *Encyclopedia of educational technology*. Thousand Oaks, CA: Sage Publications, Inc.

Snow, R. E. (1990). New approaches to cognitive and conative assessment in education. *International Journal of Educational Research, 14*(5), 455-473.

Spector, J. M. (2000). System dynamics and interactive learning environments: Lessons learned and implications for the future. *Simulation and Gaming, 31*(4), 509-516.

Spector, J. M. (2003). Problems with problem-based learning: Comments on model-centered learning and instruction in Seel (2003). *Technology, Instruction, Cognition, and Learning, 1*(4), 359-374.

Spector, J. M., Christensen, D. L., Sioutine, A. V., & McCormack, D. (2001). Models and simulations for learning in complex domains: Using causal loop diagrams for assessment and evaluation. *Computers in Human Behavior, 17*(5-6), 517-545.

Spector, J. M., & Koszalka, T. A. (2004). *The DEEP methodology for assessing learning in complex domains (Tech. Report No. NSF-03-542)* (NSF-03-542). Retrieved from Syracuse, NY:

Steed, M. (1992). STELLA, a simulation construction kit: Cognitive process and educational implications. *Journal of Computers in Mathematics and Science Teaching, 11*(1), 39-52.

Sterman, J. D. (1994). Learning in and about complex systems. *System dynamics review, 10*(2-3), 291-330.

van Joolingen, W. R., de Jong, T., Lazonder, A. W., Savelsbergh, E. R., & Manlove, S. (2005). Co-Lab: Research and development of an on-line learning environment for collaborative scientific discovery learning. *Computers in Human Behavior, 21*, 671-688.

Vosniadou, S. (1994). Capturing and modelling the process of conceptual change. *Learaning and Instruction, 4*, 45-69.

White, B. Y. (1993). ThinkerTools: Causal models, conceptual change, and science education. *Cognition and Instruction, 10*(1), 1-100.

World Economic Forum (2023). **The Future of Jobs Report 2023**. Geneva: World Economic Forum. https://www.weforum.org/publications/the-future-of-jobs-report-2023/