



# Using a three-dimensional thinking graph to support inquiry learning

## Citation

Chen, Juanjuan. No date. Using a three-dimensional thinking graph to support inquiry learning. Doctoral dissertation, Harvard University.

## Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:37678313>

## Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Open Access Policy Articles, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#OAP>

## Share Your Story

The Harvard community has made this article openly available.  
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

## **Using a Three-Dimensional Thinking Graph to Support Inquiry Learning**

Juanjuan Chen, Minhong Wang

Faculty of Education, The University of Hong Kong

Tina A. Grotzer, Chris Dede

Graduate School of Education, Harvard University

### **Abstract**

The use of external representations has a potential to facilitate inquiry learning, especially in hypothesis generation and scientific reasoning, which are typical difficulties encountered by students. This study proposes and investigates the effects of a Three-dimensional Thinking Graph (3DTG) that allows learners to combine in a single image, problem information, subject knowledge (key concepts and their relationships), and the hypothesizing and reasoning process involved in exploring a problem, to support inquiry learning. Two classes of eleventh grade students (97 in total) were randomly assigned to use either the 3DTG (i.e., experimental condition) or concept map (i.e., control condition) to complete a group-based inquiry task in an online environment. Data were collected from multiple data sources, including measures of group inquiry task performance, post knowledge scores, post questionnaires of student perceptions of consensus building, and open-ended survey. The analysis of group task performance revealed that participants in the experimental condition performed better in the inquiry task than their counterparts specifically in generating hypotheses and reasoning with data, but not in drawing conclusions. The findings show the 3DTG's benefits in facilitating exploring and justifying hypotheses, and also suggest that the 3DTG can share equal value with concept mapping in terms of drawing conclusions and consensus building. In addition, students using the 3DTG achieved higher scores in the post knowledge test, although no difference was found in their perceptions of consensus building. These findings were validated by students' responses to the survey. Implications of this study and future work are also discussed.

**KEYWORDS:** inquiry learning, collaborative learning, external representations, concept map, science education

## Introduction

Contemporary education emphasizes students' active involvement in learning, especially through inquiry and problem-solving experiences (Bransford, Brown, & Cocking, 1999; de Jong & van Joolingen, 1998; Lazonder & Harmsen, 2016). Originating in scientific inquiry practices, inquiry learning involves students in exploring phenomena or problems by asking questions, collecting and interpreting data, constructing evidence-based arguments, and forming conclusions (Bell, Urhahne, Schanze, & Ploetzner, 2010; Lazonder & Harmsen, 2016; Sandoval, 2003). Through collaborative inquiry activities, students acquire not only knowledge, but also discipline-related practices and reasoning skills (Hmelo-Silver, Duncan, & Chinn, 2007). Studies have indicated that inquiry learning is effective in fostering deeper and meaningful learning, improving academic achievement (Furtak, Seidel, Iverson, & Briggs, 2012), developing problem-solving skills (Geier et al., 2008; Hmelo-Silver et al., 2007), and enhancing intrinsic motivation (Albanese & Mitchell, 1993).

Information and communication (ICT) have been increasingly employed in inquiry learning by presenting problem contexts in vivid and interactive formats. This enables a variety of inquiry activities (such as data collection, data analysis, and the communication and discussion of results), thereby facilitating learners' understanding of the phenomena. Examples of technology-supported inquiry learning include computer simulation (van Joolingen, Savelsbergh, de Jong, Lazonder, & Manlove, 2005), Web-based Inquiry Science Environment (WISE) (Linn & Slotta, 2000), and immersive learning environments (Kamarainen, Metcalf, Grotzer, & Dede, 2015).

Nevertheless, whether in traditional or technology-enabled contexts, students often have difficulties in regulating the inquiry process and engaging in fruitful inquiry learning (Kollar, Fischer, & Slotta, 2007). In particular, many students do not know how to formulate hypotheses (de Jong & van Joolingen, 1998). Moreover, students' reasoning ability is usually

inadequate (Hogan & Maglienti, 2001; Zimmerman, Raghavan, & Sartoris, 2003); they have difficulties integrating evidence or data with subject knowledge and difficulties hypothesizing and reasoning with intertwined variables. For example, studies have pointed out that students are unable to adapt or revise their initial hypothesis when presented with conflicting evidence (Chinn & Brewer, 1993; Klahr & Dunbar, 1988; Kuhn, Black, Keselman, & Kaplan, 2000). Furthermore, they may reject hypotheses in the absence of disconfirming evidence and have a tendency to find confirming rather than disconfirming information (Zeineddin & Abd-El-Khalick, 2010).

Based on the foregoing, there is a great need to support inquiry learning to guide students through the complex inquiry process, and to help them become accomplished problem-solvers (Bransford et al., 1999; Kirschner, Sweller, & Clark, 2006; Quintana et al., 2004). Over time, an increased number of studies have investigated different kinds of support or guidance for inquiry learning. To facilitate the complex inquiry process, some task structuring strategies have been used to guide learners step by step through different stages of exploration (Hsu, Chiu, Lin, & Wang, 2015; Linn & Songer, 1991; Stratford, Krajcik, & Soloway, 1998). Furthermore, scripts, prompts, or hints (Davis, 2003; Kollar et al., 2007; Noroozi, Teasley, Biemans, Weinberger, & Mulder, 2013) have been provided to facilitate the inquiry process (i.e., what to do next or how to do it). In response to students' difficulties in the processes of hypothesizing and reasoning, additional cognitive supports have been investigated. For example, Shute and Glaser (1990) offered learners optional variables and linking words from which learners can pick to generate hypotheses. External representations have been employed to scaffold students' scientific reasoning, such as concept maps (Gijlers & de Jong, 2013), causal maps (Slof, Erkens, Kirschner, Janssen, & Jaspers, 2012), and evidence maps (Suthers, Vatrapu, Medina, Joseph, & Dwyer, 2008). These representational supports have been proved to be effective in facilitating hypothesis formulation or reasoning.

Inquiry learning with real-world problems involves complex processes of collecting a variety of information and data, integrating these data with various aspects of subject knowledge, and hypothesizing and reasoning with intertwined variables. Accordingly, there is a need for more studies investigating approaches to externalizing the complex, implicit aspects of inquiry tasks to facilitate deeper and more meaningful learning in authentic situations (Wang, Kirschner, & Bridges, 2016). This study proposes and investigates the effects of a Three-dimensional Thinking Graph (3DTG), which allows learners to combine in a single image, problem information, subject knowledge (key concepts and their relationships), and the hypothesizing and reasoning process involved in exploring a problem, to support inquiry learning. The premise underlying the design is that externalizing the complex aspects of the inquiry task makes it relatively easier for learners to successfully solve the problem (Janssen, Erkens, Kirschner, & Kanselaar, 2010). Representational guidance on reasoning also makes the formulation of hypotheses and logical reasoning salient (focusing student's attention on these activities) (Toth, Suthers, & Lesgold, 2002). This study therefore aims to explore the effects of the proposed 3DTG on inquiry learning in an online environment.

### **Theoretical framework**

#### **Inquiry learning**

Originating in scientific inquiry practices, inquiry learning engages students in exploring phenomena or problems by posing questions, collecting and interpreting information or data, constructing evidence-based arguments, and forming conclusions (Bell et al., 2010; Bransford et al., 1999; Lazonder & Harmsen, 2016; Sandoval, 2003). Inquiry activities may involve discussing an open-ended question, explaining a phenomenon, manipulating quantitative parameters to understand a scientific model, or solving a problem. Explanations, models, and theories are examples of inquiry learning artifacts. In inquiry learning, students

acquire not only content knowledge, but also discipline-related practices and reasoning skills through collaborative inquiry activities, especially in science teaching and learning (Hmelo-Silver et al., 2007).

The inquiry process often involves iterative cycles of gathering information and data through observations or experiments, generating hypotheses, reasoning with collected information and data, and drawing conclusions (Klahr & Dunbar, 1988; Njoo & De Jong, 1993a). Generating hypotheses is to formulate ideas about the relationships between variables. A hypothesis is constituted by a set of variables involved and the relations among them (Klahr & Dunbar, 1988; Van Joolingen & De Jong, 1997). There are several levels of hypotheses. Some researchers make a distinction between experimental and theoretical hypothesis. *Experimental hypothesis* is a prediction or anticipation of an event or of a relationship between causal and responding variables (e.g., iron is attracted by a magnet), which characterizes empirical regularities and it is likely to be acquired through observations of phenomena and repeated experiments (Germann & Aram, 1996; Wecker et al., 2013). *Theoretical hypothesis* explains observable phenomena on the theory level (e.g., the magnetic force), and generally refers to theoretical entities that cannot be directly observed (Germann & Aram, 1996). An example is “Iron transmits magnetic force”. In this study, the hypotheses focus on the identification of causes and effects (Kuhn et al., 2000). Thus, the term hypothesizing refers to formulating causal relations between variables.

Scientific reasoning is largely considered to be hypothetico-deductive in structure (Lawson, 2005). It involves questioning initial premises, seeking confirming or contradictory evidence for the hypothesis, revising initial ideas, building new understanding, and considering alternative hypotheses (Zeineddin & Abd-El-Khalick, 2010); learners try to interpret the gathered data to construct causal explanations based on evidence and logical argument (Hsu et al., 2015). Reasoning skills are believed to affect learners’ abilities to

develop scientific understandings and conduct scientific investigations (Lawson, Banks, & Logvin, 2007).

Students have difficulty reasoning about systems; particularly, they reveal inability to reason about causality in a systemic sense (Grotzer & Bell Basca, 2003). They tend to reason at a local, not a global, level, and miss the system-level interactions and the larger picture (Penner, 2000). Systemic explaining and reasoning characterizes scientific sense-making. It involves comprehending a variety of causal patterns such as domino-like, cyclic, or mutual patterns. Without a grasp of the underlying causal relationships, students would easily develop misconceptions and superficial understanding of the systems. As pointed out by Grotzer and Bell Basca (2003), local causal reasoning as opposed to domino-like reasoning is common in students' thinking about ecosystem relationships, making the root cause of a problem difficult to discern.

### **External representation**

External representations are graphical or diagrammatic representations of knowledge or information using maps, diagrams, tables, or pictures. (Cox, 1999; Novak, Bob Gowin, & Johansen, 1983; Roth & McGinn, 1998; Toth et al., 2002; Van Bruggen, Kirschner, & Jochems, 2002). They can either be constructed by the students themselves or created before hand by teachers or from textbooks. The learner's active construction of external representations related to a problem's solution has been found to promote deeper learning and improve learning processes and outcomes of inquiry learning and problem solving (Janssen et al., 2010; Suthers & Hundhausen, 2003).

The benefits of constructing an external representation to support inquiry learning can be explained on the cognitive, metacognitive, and social dimensions (Toth et al., 2002). In the cognitive dimension, it assists with problem solving by re-ordering information in useful ways, facilitating inferences, focusing students on the construction of knowledge, and helping



reduce cognitive load (Cox, 1999). In the metacognitive aspect, it tracks the progress of reasoning through the problem and directs attention to the unsolved part of the problem. In the social aspect, it serves as shared cognition or as a discussion anchor that coordinates the discourse between peers during collaborative inquiries (Roth & McGinn, 1998; Schwendimann & Linn, 2016). Additionally, constructing a shared external representation facilitates consensus building because the representation stimulates discussion and argumentation (Gijlers, Saab, Van Joolingen, De Jong, & Van Hout-Wolters, 2009; Weinberger, Stegmann, & Fischer, 2007). If learners have divergent ideas about a learning task, they must reach some kind of consensus in order to collaboratively solve it.

Among various external representations, concept map has been the most widely used in learning. Originally developed for science concept learning, a concept map is a graph consisting of nodes and labeled connecting lines, explicitly organizing and representing the knowledge of concepts and the relationships between them (Novak et al., 1983). It can be used to represent all kinds of concepts and relationships. Although most studies on concept mapping have used it as a conceptual learning tool (Markow & Lonning, 1998; Pedaste et al., 2013; Roth & Roychoudhury, 1993; Ruiz-Primo, Schultz, Li, & Shavelson, 2001; Schwendimann & Linn, 2016), several others deployed it to assist with problem-solving or inquiry learning. For example, Gijlers and de Jong (2013) suggested that students who constructed concept maps performed better on knowledge tests and were more engaged in consensus building. When concept mapping, students articulated, revisited, and reorganized their ideas (Schwendimann & Linn, 2016).

Causal mapping has been used to represent relations of cause and effect. Slof et al. (2012) examined its effects by asking students to construct a causal map from a given set of predefined concepts or solutions needed to solve a complex task. They found that students using causal maps justified their solution better (i.e., reasoning) and demonstrated more

cognitive and meta-cognitive activities reflected in online discussions than those who did not create causal maps. In addition, evidence maps have been developed to make the relationship between evidence (i.e., information or data) and a hypothesis more explicit, and are conceptually-oriented rather than problem-solving oriented (Suthers et al., 2008; Toth et al., 2002). Suthers and colleagues reported that students using evidence maps generated more hypotheses and were more likely to converge on the same conclusion. Conversely, the evidence mapping tool employed by Toth et al. (2002) had a significantly positive effects on reasoning, but no significant effect on hypothesis generation. Finally, Wang, Wu, Kinshuk, Chen, and Spector (2013) proposed a computer-based integrated cognitive map that enabled students to externalize the reasoning process on a reasoning map and the knowledge underlying the reasoning process on a concept map when they worked with diagnostic problems. The approach showed promising effects on diagnostic problem solving.

In summary, concept map is mainly used to visually represent the relationships between concepts in domain knowledge (Ruiz-Primo & Shavelson, 1996). However, concept mapping alone has been found to be inadequate in supporting complex problem-solving tasks, and particularly in eliciting and representing the process of applying knowledge to practice (Wang, Cheng, Chen, Mercer, & Kirschner, 2017). Causal map is a kind of concept map, which can be used to represent causal relationship, a specific type of relationship between concepts (Slof et al., 2012). Evidence map goes beyond conceptual knowledge by representing the reasoning process and building the relations between evidence and hypothesis (Suthers et al., 2008).

Given that inquiry and problem-solving tasks have usually mixed all kinds of information and data, concepts, and relationships, learning in such contexts often involves complex cognitive processes such as searching for information and data from multiple aspects, integrating information with domain knowledge, and generating a solution or

explanation to the problem by reasoning with data that are intertwined in sophisticated ways. Many students have found it hard and cognitively demanding. Therefore, there is a need for more studies to investigate how the complex cognitive processes involved in inquiry and problem-solving tasks can be externalized and facilitated towards desired task performance and learning outcomes. Here, we studied the use of the proposed 3DTG that allows learners to combine in a single image, problem information, subject knowledge (key concepts and their relationships), and the hypothesizing and reasoning process, to support inquiry learning. Externalizing the three cognitive aspects (i.e., searching information and data, recalling and applying domain knowledge, and hypothesizing and reasoning) in a holistic way is critical for successful inquiry learning.

More specifically, when students construct concept maps, relationships between intertwined concepts/variables will become clearer to them. Concept maps enable students to bring out important concepts/variables that they probably would not have noticed (Markow & Lonning, 1998). The learners will think about how different variables are related to each other. Second, compiling the problem information and the changes in a set of key variables can make important aspects of the problem salient. Table is an appropriate external representation of a series of information and data. Third, inspired by the design of evidence map, we designed a reasoning map to incorporate the processes of hypothesis generation and reasoning. The difference between evidence map used by Suthers and colleagues (2008) and our reasoning map is that evidence map only incorporates one hypothesis, whereas our reasoning map represents a logical series of hypotheses to find the root cause instead of superficial cause of a problem. Finally, integrating the three types of representations in a single image was inspired by Wang's et al. (2013) study on connecting problem-solving (i.e., reasoning map) and knowledge construction (i.e., concept mapping) in a single image, which has been proved to be promising in problem-solving contexts.

### Three-dimensional Thinking Graph

The proposed 3DTG is illustrated in Figure 1. It consists of three parts: a concept map, a data table, and a reasoning map. The concept map includes the concepts of subject knowledge underlying the problem and the relationships between the concepts. The data table records the problem information, reflected as a set of key variables and their changes during an observation period. The reasoning map is a representation of the evidential relationships between the hypotheses and the data or subject knowledge; each hypothesis is supported or rejected by evidences from the data or subject knowledge. A hypothesis can be further explained by other hypotheses explicating deeper causes of the problem. The three parts of the 3DTG can be drawn on one piece of paper. In carrying out the reasoning process, learners draw the reasoning map while observing the concept map and data table.

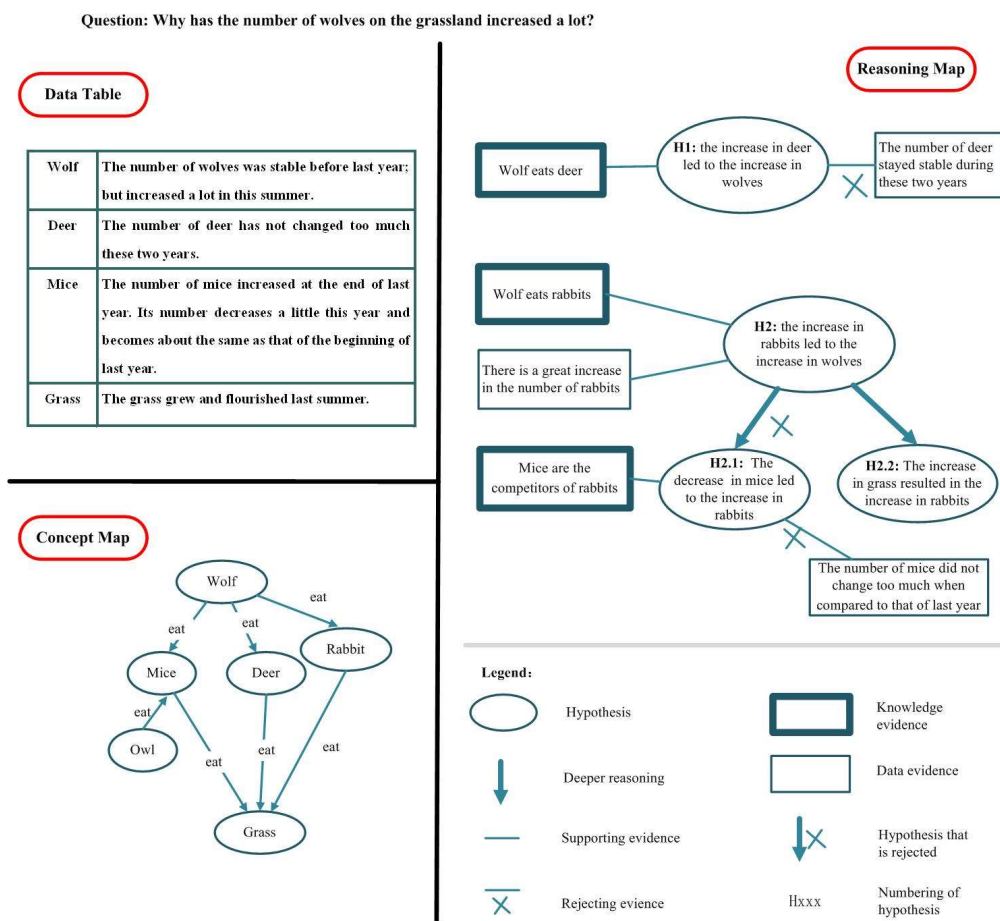


Figure 1. Three-dimensional Thinking Graph (3DTG)

## **Research Questions and Hypotheses**

This study aimed to investigate the effects of the proposed 3DTG on inquiry learning. More specifically, the effects on student inquiry performance and knowledge achievement are examined because inquiry learning is intended to promote knowledge acquisition as well as discipline-related inquiry skills (Hmelo-Silver et al., 2007). In addition, its effect on students' perceptions of consensus building is explored because consensus building is an important but rarely examined inquiry activity. Research has indicated that consensus building can be improved by student constructions of external representation (Gijlers & de Jong, 2013). Therefore, this study's hypotheses are stated as follows:

H1. Students using the 3DTG will perform better in the inquiry task in terms of formulating hypothesis, reasoning, and making a conclusion than those using only the concept map.

H2. Students using the 3DTG will achieve higher scores on a knowledge test than those using only the concept map.

H3. Students using the 3DTG will have more positive perceptions of consensus building than those using only the concept map.

## **Methodology**

### **Experimental Design**

This study employed a quasi-experimental design with two comparison conditions to explore the effects of using the 3DTG on inquiry learning in an online environment. Students in the experimental condition used the 3DTG, and their counterparts in the control condition used concept maps to facilitate the problem-solving process. All of the students were required to work in small groups to complete the same learning task of exploring a fish die-off problem in a biology course. The dependent variables (or learning outcomes) include group task performance, knowledge achievement, and perceptions of consensus building. These

dependent variables were analyzed to reveal whether the use of the 3DTG would lead to better learning outcomes than the use of concept map.

### **Participants**

Two 11th grade classes, from one high school, taught by the same biology teacher, were recruited and randomly assigned to the experimental condition or the control condition. One class with 48 students (24 males and 24 females) was assigned to the experimental condition using the proposed 3DTG, while the other class with 49 students (24 males and 25 females) was assigned to the control condition using a traditional concept mapping approach. In each condition, the students were divided into small groups of three or four each, based on their friendships. In total, there were 15 experimental groups and 16 control groups. The average age for both classes was 17 (ranging from 16 to 18). Before the experiment, the participants had already acquired basic knowledge about ecosystems.

### **Learning Environment and Materials**

Students in both conditions worked in small groups to explore a pollution problem within an ecosystem presented in an online environment. The online environment consists of two major modules: *problem context* and *learning support*.

The *problem context* module presented an authentic pollution problem in a pond ecosystem as a learning task that requires consideration of multiple perspectives and the use of evidence to reach an adequate solution. The problem-context was based on the EcoMUVE curriculum (Kamarainen et al., 2015), in which students explore a virtual pond and the surrounding watershed, observe simulated organisms over a number of virtual “days”, and collect relevant data in order to investigate why many of the fish had died off. The module contained *information collection* and *data observation* sub-modules, as shown in Figure 2. The sub-module of *information collection* provides a rich context, such as vivid descriptions and visualizations of the surroundings and background information on the pond ecosystem

necessary to problem solving. By selecting a specific day from a dropdown list of dates, the user could observe and collect information for that day, providing tacit clues or hints and guiding students in their observation of the phenomena. Students were able to relate the observation to their prior knowledge.

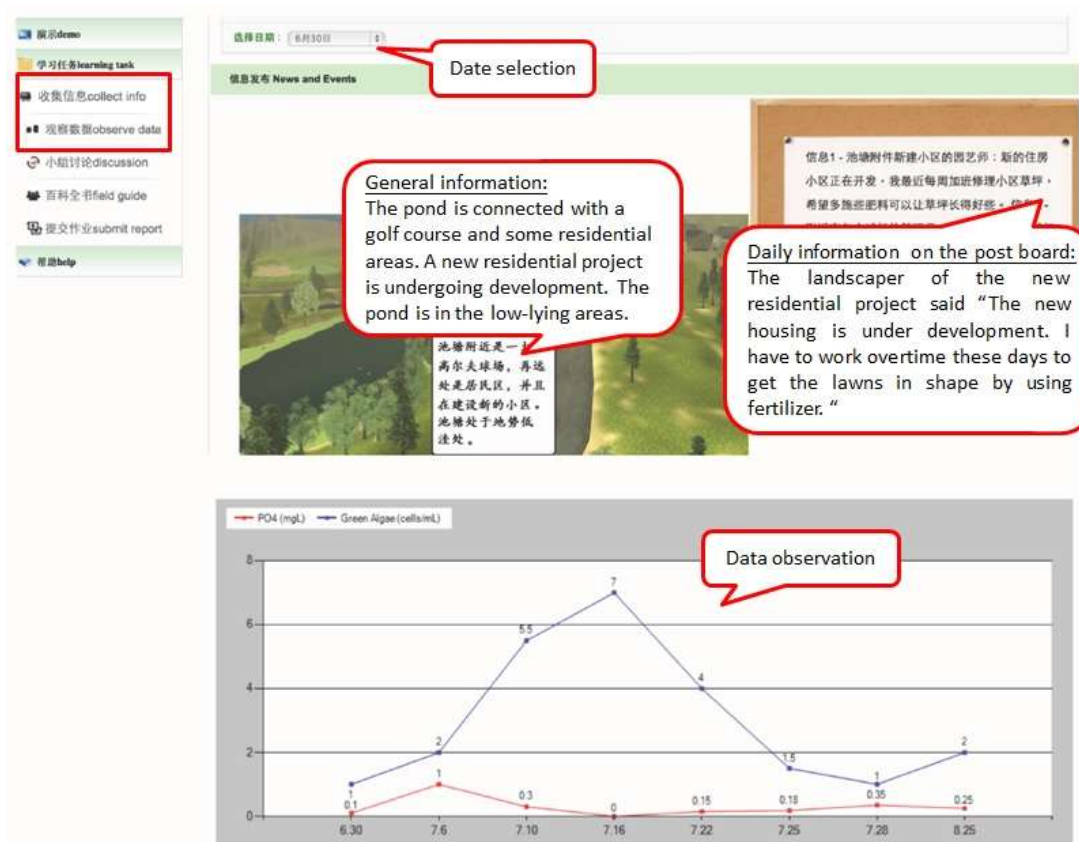


Figure 2. Information collection and data observation

The sub-module of *data observation* facilitates problem exploration by depicting data graphs of the key variables and the changes in variables over a period. For example, in the pond ecosystem, students could observe data on water conditions (e.g., water temperature, turbidity, concentration of  $\text{PO}_4$  and  $\text{NO}_3$ , dissolved oxygen), weather conditions (e.g., air temperature, wind speed, cloud cover), and the population of various organisms (e.g., bacteria, algae, bass). Moreover, the module enables the learner to construct a query by selecting variables to facilitate the analysis and comparison of variables.

The *learning support* module afforded some learning guidelines. First, it included domain-specific knowledge about ecosystems and ecological processes, and a list of key concepts regarding the problem context, such as the food web, the roles of biotic and abiotic factors, and the processes involved (e.g., photosynthesis). Previous research has shown the importance of prior knowledge in inquiry activities (Shin, Jonassen, & McGee, 2003) and indicated the necessity of having sufficient contextual concepts for productive problem solving (Gijbels, Dochy, Van den Bossche, & Segers, 2005). Without sufficient prior knowledge, students might not have made good interpretations of the data.

Second, this module incorporated the introduction of basic skills and the steps required for scientific inquiry, such as hypothesis formation and evidence-based reasoning. Instructional guidance on structuring the inquiry process has proved effective (Kirschner et al., 2006; Linn & Songer, 1991; Njoo & de Jong, 1993b; Quintana et al., 2004; White, 1993).

Third, the module provided general instructions on how to build concept maps, and how to conduct group discussions. Researchers have pointed out the necessity of providing learners with instructions and explicit guidance in the use of tools and strategies (Stoyanov & Kommers, 2006). Additionally, the environment for the experimental group offered guidelines for using the 3DTG. Lastly, an example is afforded to both groups to demonstrate how to use the learning system. Coaching also helped learners acquire a better understanding of the processes and principles of inquiry learning.

### **Learning Task**

The learning task involved performing causal reasoning and constructing logical and scientific explanations for the fish die-off problem, i.e., why so many large fish had died suddenly. By interacting with the online environment, students collected relevant background information and observed data changes over time for different variables. Meanwhile, they discussed and solved problem in small groups by evaluating and compiling the collected



information, formulating hypothesis, and reasoning. During the reasoning process, the students in the control condition were asked to create concept maps, whereas students in the experimental condition were asked to create 3DTGs. Finally, all groups in both conditions were required to submit an inquiry report presenting their solutions and how they formed their hypotheses together with evidence in relation to their reasoning and conclusions.

### **Procedure**

The experiment lasted for six sessions over two weeks (45 minutes for each session). During the first session consent forms were signed by the participants. A pre-test questionnaire was administered to collect information on the students' age, gender, and computer skills. Groups of three or four were then formed by students on the basis of their friendship.

At the beginning of the second session, the researcher provided a 20-minute training to both the experimental and control groups on how to perform inquiry learning in the online system. The training was conducted in the school's computer lab, where the researcher explicitly introduced the principles and procedures of inquiry learning in the form of an example. The principles of social interaction, together with the supporting materials were also briefly introduced during the training session. Additionally, students in the experimental condition were introduced to the approach to construct a 3DTG, and students in the control condition were introduced to the approach to create a concept map.

After the training, students began to collaboratively perform the learning task in the school's computer lab, with group members sitting together and each one accessing a networked computer. To support or reject their initial hypotheses, they made observations, collected information, and formulated hypotheses from multiple perspectives by brainstorming ideas based on the compiled evidence (i.e., information and data). During the group discussion and collaboration, each group member was able to share his/her findings

within the group and could be challenged by group peers. Finally, the group had to reach an agreement on the best explanations for the problem. At the end of the fifth session, each group was asked to submit an inquiry report synthesizing the group's inquiry process (including the collected information and data, generated hypothesis, reasoning, and conclusion). This lasted for three and a half sessions.

After the fifth session, a paper-based two-item open-ended survey was administered to ascertain all students' perceptions of the inquiry activities by requiring students to write the responses to two open-ended questions on the paper in about 15 minutes. Finally, in the sixth session, the participants were asked to individually take a knowledge test (30 minutes) and a Likert-scale questionnaire (15 minutes). The open-ended survey, the test, and the Likert-scale questionnaire took place in the participants' classroom.

### **Measures and Instruments**

Evaluation in problem solving should assess not only students' knowledge base, but also their problem-solving skills and competencies. Therefore, this study not only examined the students' knowledge by using a post knowledge test, but also investigated student inquiry performance as reflected in the inquiry report. In addition, Likert-scale questionnaire was used after the knowledge test to determine the student perceptions of the group consensus building activities. Finally, an open-ended survey was conducted to collect students' comments on the learning activities.

**Pre-test questionnaire.** The pre-test questionnaire asked students' gender, age, and self-assessment of their computer skills ("Please indicate your computer skills: 1. Very poor, 2. Poor, 3. Neither poor nor good, 4. Good, 5. Very good.")

**Pre and post knowledge test.** The previous semester's subject test of the school served as the pre-test to examine whether the two classes had equivalent prior knowledge before participating in the experiment. The pre-test was designed by two experts in biology

instruction. The post-test was designed by the researchers and an experienced biology teacher together. Both tests assessed students' knowledge of ecosystems addressing ideas regarding photosynthesis, respiration, and decomposition, containing 23 multiple-choice (with only one correct answer) questions, 2 fill-in-the-blank questions, and 2 short-answer essay questions, with a highest possible score of 100. The test items in both tests were adapted from the database of the National College Entrance Exam.

Two biology instructor experts evaluated the face and content validity to ensure that the tests aligned well with the learning objectives. In addition, the pre- and post-tests were pilot-tested with 73 grade 11 students who had similar characteristic with this study's participants. The two tests were administered in two consecutive days. The piloting results show that the Pearson correlation is  $r = 0.70$  ( $p < .01$ ) between the pre- and post-tests, indicating that the concurrent validity is quite acceptable. The difficulty level (i.e., the correct answer rate of the items) was 0.81 and 0.84 for the multiple-choice items of the pre- and post-tests, and 0.78 and 0.76 for non-multiple-choice items (fill-in-the-blank and short essay) of the pre- and post-tests, respectively, indicating comparable difficulty between the two tests. The means and standard deviations are as follows: Mean = 79.78 (SD = 10.12) for the pre-test, and Mean = 80.01 (SD = 7.51) for the post-test. These pilot results suggest that the two tests are equivalent.

With respect to the reliability of the tests, the internal consistency reliability for all the multiple-choice test items reached 0.71 for the pre-test and 0.76 for the post-test, 0.62 and 0.63 for non-multiple-choice items of the pre- and post-tests, respectively, showing that the two tests have high reliability. The inter-rater reliability for non-multiple-choice items was 0.81. Thus, the reliability of the two tests can be considered acceptable.

**Inquiry report.** The inquiry report was used to examine each group's inquiry task performances. It consisted of a summary and extraction of the group's problem-solving

processes. The solutions to ill-structured problems are generally difficult to assess because they are divergent and probabilistic. Their evaluation must consider both processes and the criteria for the group product (Jonassen, 1997). For the inquiry task presented in this study, the focus of the evaluation was on whether the students articulated the causal relations (i.e., explained why the pollution problem occurred) and whether the report incorporated important assumptions or hypotheses, and provided sufficient evidence (e.g., information and data), reliable arguments and reasoning.

Corresponding to the inquiry activities established by the literature (Bell et al., 2010; Gijlers & De Jong, 2005; Singer, Marx, Krajcik, & Clay Chambers, 2000; Windschitl, 2004), the assessment rubrics for the inquiry report in this study took into account five aspects: identification of critical information and data, formulation of the hypothesis, sufficiency of the hypotheses, reasoning and justification using collected information and data, and the conclusion. The total score for each aspect was 5. A mean score was attained for each aspect by averaging scores for all relevant elements.

The detailed assessment rubrics are presented in Table 1. It should be noted that rating relevance/importance and validity complemented each other. For example, it was possible for one piece of reasoning to receive a high rating score in terms of relevance/importance, but a low score in terms of validity due to invalid or flawed arguments.

The reason for evaluating the sufficiency of hypotheses was that the group might perform well in formulating one relevant and valid hypothesis, but not perform well in listing many relevant hypotheses; this could result in a situation in which groups generating fewer hypotheses received a higher score for the group report (Janssen et al., 2010). To overcome this possible bias, we evaluated the sufficiency (i.e., number of valid hypotheses) to better assess student ability to formulate hypotheses. The scores for sufficiency together with the hypotheses' importance and validity were assessed separately in the report. Finally, an inquiry

report received a score by averaging the five sub-scores for the five aspects listed in Table 1. Two of the researchers of the study designed the structure of the group report and assessment rubrics. Two domain experts were first trained and then applied the assessment rubrics to assess the inquiry report independently, and their scores were averaged. Cohen's Kappa suggested substantial agreement between the two raters, with all ratings exceeding 0.70 and being statistically significant (Cohen's Kappa ranging from 0.706 to 1). The scoring of group reports was reliable.

Table 1. Assessment rubrics for the inquiry report

Aspects	Rubrics	Description	Illustrative examples
1. Identification of critical information and data	Inclusion of critical information and data	The ratio of collected information and data to the total relevant information and data posted	<i>It rained heavily on July 6th.</i>
2. Formulation of hypothesis	2.1. Relevance or importance	Is the formulated hypothesis important to the cause analysis of the pollution problem? 1 – irrelevant or unimportant, 2 – a little relevant, 3 – relevant, 4 – very relevant or important, 5 – highly or particularly relevant/important	<i>The fish die-off might be caused by the deficiency of oxygen. E1</i> (This is a highly relevant/important hypothesis.)
	2.2. Validity	Is the formulated hypothesis valid? 1 – totally invalid, 2 – partially valid, 3 – seemingly valid to some extent, 4 – valid, 5 – absolutely valid	<i>The higher cloud cover led to the increase of NO<sub>3</sub>. E2</i> (This is an irrelevant hypothesis.) <i>The increase of PO<sub>4</sub> and NO<sub>3</sub> led to the increase of PH value. E3</i> (This hypothesis is totally invalid.)
3. Sufficiency of hypotheses	The number of valid hypotheses	How many valid hypotheses are there? 1 – less than two appropriate hypotheses, 2 – two hypotheses, 3 – three hypotheses, 4 – four to five hypotheses, 5 – more than five hypotheses	
4. Reasoning and justification	4.1. Relevance or importance	Is the reasoning and justification relevant to the analysis of the pollution problem's cause? How well does the report provide appropriate evidence, proof, or examples to support the hypothesis? 1 – irrelevant, 2	<i>The oxygen level reduced to an extremely low level (i.e., 4.1 mg/L), which caused the deficiency of oxygen needed by the fish. (One piece of highly relevant evidence supporting E1.)</i>

	– a little relevant, 3 – relevant, 4 – very relevant, 5 – highly/particularly relevant	
4.2. Correctness, reasonableness	Is the reasoning and justification correct or reasonable? 1 – little argument and evidence, 2 – argument with irrelevant evidence, 3 – correct argument with little evidence, 4 – correct argument supported by more evidence, 5 – correct argument with sufficient correct evidence	(The piece of evidence provided in the above row supporting <b>E1</b> is correct.)
5.1. Consistency	Is the conclusion consistent with hypothesis and reasoning? 1 – inconsistent, 2 – consistent to some extent, 3 – consistent to a large extent, 4 – consistent, 5 – extremely consistent	
5.2. Correctness	Does the conclusion explain the problem? 1 – totally no, 2 – explains a little, 3 – explains to some extent, 4 – explains most of the causes, 5 – explains perfectly	<i>On the one hand, the fish die-off was caused by the deficiency of oxygen; the increased green algae consumed lots of oxygen. On the other hand, the increased concentration of bacteria killed the fish. (This gets 2 scores as it explains a little.)</i>

*Note.* **E** in the example column is a label for the example.

**Post-test questionnaire.** A post-test questionnaire measured students' perception of group consensus building activities. It was based on a five-point Likert-scale ranging from 1 (strongly disagree) to 5 (strongly agree). Five items were developed from the construct of consensus building activity proposed by Gijlers and de Jong (2013). Example items are "During collaboration, our group built on the ideas of a partner or took over the perspective of a partner", and "During collaboration, our group integrated multiple ideas or viewpoints".

We performed the principal components of factor analysis with varimax rotation to ensure the construct validity of consensus building. As the value of Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy was .819 and Bartlett's test result was significant (p

< .001), the data were appropriate for factor analysis (i.e., if KMO value is less than 0.5 or Bartlett's test significant value is greater than 0.05, factor analysis should not be applied). In the varimax-rotated factors, one factor was extracted with an eigenvalue greater than 1 and it explained 71.035% of the total variance. All five items achieved factor loading above 0.5 (with minimum being 0.662). Thus, no items were removed. Regarding the reliability, Cronbach's alpha value for the consistency evaluation of the consensus building construct was .944, indicating credible internal consistency.

**Open-ended Survey.** Two-item open-ended survey were used to elicit and probe learners' perceptions in depth. The two items or questions were: (1) Did your group have any difficulty solving the problem (e.g., hypothesizing, reasoning)? (2) Was the 3DTG/concept map helpful to you in solving the problem? If yes, how did it help you?

Participants' written responses were qualitatively analyzed to generate patterns and themes that typified their perceptions. The analysis followed an iterative process of coding and generating themes and categories (Krahtwohl, 1998). An emergent theme was named as a category. Within each category, there may be sub-categories based on similar themes.

The first author analyzed the data. The second author conducted a blind round of analysis. The inter-rater reliability reached .87. Discrepancies in emergent themes or categories were discussed and reconciled by further consultation of the data. Based on the confirmed themes, the responses of all participants were coded by identifying the available themes in each response. Each response may include more than one theme.

### **Data Analysis Method**

This study adopted the following methods of data analysis:

- (1) Factor Analysis was conducted to ensure the construct validity of consensus building questionnaire items. Cronbach's  $\alpha$  was run to assess the internal consistency of the questionnaire.

- (2) Cohen's Kappa was run to determine the inter-rater reliability (or agreement) in scoring the inquiry report. The result exceeding 0.7 indicated substantial agreement, and 0.5 indicated moderate agreement.
- (3) One-way analysis of variance (ANOVA) was performed to evaluate statistical differences between the two conditions for the pre-test questionnaire, prior knowledge, and post-test questionnaire. Meanwhile, ANCOVA was used to explore the effects of the intervention on post-test scores.
- (4) Multivariate analysis of variance (MANOVA) was conducted to examine the statistical difference between the two conditions in the aspects of the group task performance.
- (5) A thematic content analysis of the written survey was performed to find common themes among students' responses to each open-ended question.

## Results

### Pre Knowledge Test and Computer Skills

Table 2. Descriptive statistics and ANOVA for pre-test and computer skills

	Condition	N	Mean	SD	SE	Levene's test ( <i>p</i> )	F(1, 94)	Sig.
Pre-test	E	48	63.19	8.486	1.225	.794	2.226	.139
	C	48	65.81	8.746	1.262			
Computer skills	E	41	3.34	.693	.108	.814	.396	.531
	C	49	3.34	.751	.107			

*Notes.* E = Experimental group. C = Control group.

Levene's test confirmed the equality of the variances of pre-test scores and computer skills in the two conditions. ANOVA was conducted on students' scores on pre-test and their self-perceived computer skills. The descriptive statistics and ANOVA results are shown in Table 2. It can be seen that there was no significant pre-existing difference between the two conditions with respect to pre-test scores ( $p = .139$ ) and computer skills ( $p = .531$ ).



### Group Task Performance

Although 15 experimental groups and 16 control groups performed the learning task, only 10 and 14 reports, respectively were submitted. The other groups failed to submit their reports due to computer or network problems. Levene's test confirmed the equality of variances of report scores in the two conditions, as shown in Table 3.

Table 3. MANOVA results of the group inquiry report and Levene's test for equality of variances

Aspect	df	<i>F</i> value	Significance	Partial eta-squared	$\eta_p^2$	Levene's test (p)
Information and data	1	2.695	.114	.105		.755
Formulation of Hypothesis	1	21.845	.000*	.487		.478
Sufficiency of Hypotheses	1	10.895	.003*	.321		.200
Reasoning	1	8.428	.008*	.268		.912
Conclusion	1	2.337	.140	.092		.601
Total	1	5.786	.025*	.201		.617

\*  $p < .05$ .

Descriptive statistics for group task performance are presented in Table 4. The experimental groups received higher mean scores in four out of five aspects of inquiry learning, excluding the collection of information and data. They also received higher total scores for their inquiry reports. Table 3 also presents the MANOVA results for the group inquiry report. It can be seen that the experimental groups performed significantly better in the formulation of their hypothesis, sufficiency of the hypotheses, reasoning, and the total quality of the report. In fact, the hypotheses and reasoning in the experimental groups' reports were generally logically organized, whereas there was some disorder in logic in the reports of control groups. For example, the experimental groups might formulate a logical series of hypotheses: H1. The fish die-off might be caused by the deficiency of oxygen; H1.1. The decrease in the amount of oxygen might be caused by the increase of bacteria; H1.1.1 The increase of bacteria might be caused by the increase of dead green algae. There is causal relationship between H1 and H1.1, and between H1.1 and H1.1.1. By contrast, the order of

the hypotheses in the comparison group's report might look like this: H1. The die-off of fish caused the increase of bacteria; H1.1. The death of green algae caused the increase in the concentration of PO<sub>4</sub>; H1.1.1. The construction of the nearby community had an effect on the fish die-off. There is not any causal relationship between H1 and H1.1, or between H1.1 and H1.1.1. A logical set of hypotheses would give right and clear direction for reasoning and finding a root cause of the problem, which is important for effective thinking and learning in inquiry contexts.

Thus, corresponding to H1, these results confirm the benefits of using the 3DTG to facilitate students' inquiry task performance, more specifically to formulate their hypotheses, the sufficiency of those hypotheses, and their reasoning using the data, but not in drawing conclusions. This shows that the 3DTG fostered students' skills in exploring and justifying their hypotheses.

Table 4. Descriptive statistics of the group inquiry report

Aspect	Condition	N	Mean	SD
Information and data	Experimental	11	3.62	.50
	Control	14	3.99	.60
Formulation of Hypothesis	Experimental	11	4.33	.49
	Control	14	3.41	.50
Sufficiency of Hypotheses	Experimental	11	4.00	.95
	Control	14	2.93	.68
Reasoning	Experimental	11	4.04	.60
	Control	14	3.39	.52
Conclusion	Experimental	11	3.95	.57
	Control	14	3.61	.56
Total	Experimental	11	3.93	.42
	Control	14	3.55	.37

### Post Knowledge Test

The descriptive statistics for the results of the post knowledge test are shown in Table 5. The Levene's test confirmed the equality of variances of the post-test scores in the two conditions ( $p = .777$ ). With pre-test score as the covariate, ANCOVA was then run to adjust

for pre-test score. The result revealed that there was statistically significant difference between the experimental and control conditions,  $F(1, 93) = 9.036, p = .003$ . Students in the experimental condition achieved higher scores in the post knowledge test. Moreover, Cohen's effect size index  $d$  (Cohen, 1992) was computed to illustrate the extent of practical difference between groups. Cohen's  $d$  values of 0.2, 0.5, 0.8 are interpreted as small, medium, and large effect sizes, respectively. The effect size reported in this study ( $d = 0.54$ ) showed a medium effect of the intervention. This demonstrates the effectiveness of the 3DTG in improving students' knowledge. Hence, H2 is accepted.

Table 5. Descriptive statistics and ANCOVA for post-test score

Condition	N	Mean	SD	Adjusted Mean	Cohen's $d$	$F(1, 93)$	Sig.
E	48	77.99	8.60	78.27	0.54	9.036	.003*
C	47	73.34	8.46	73.05			

Note. E = Experimental group. C = Control group.

\*  $p < .05$ .

### Post Questionnaire

Levene's test verified the equality of variances in the two conditions (Levene's value = .001,  $p = .980$ ). The means and standard deviations are as following:  $M = 4.27$  ( $SD = .633$ ) for the experimental condition, and  $M = 4.40$  ( $SD = .548$ ) for the control condition. The ANOVA results indicated no significant difference in students' perception of group consensus building activities between the two conditions ( $F(1, 79) = .953, p = .332$ ). Thus, H3 is rejected.

### Open-ended Survey

The analysis result of students' written responses to the two open-ended items are presented in Tables 6 and 7, respectively. The tables demonstrate the themes, illustrative examples of each theme, and the frequency of each theme appearing in students' responses. These examples are verbatim quotes selected from students' written responses. They illustrate the participants' perceptions of the difficulties encountered in the inquiry processes and the

benefits of drawing the maps. Regarding the difficulties, there were nine emergent categories (as shown in Table 6). Among them, *others* included three sub-categories, i.e., technical, time, and knowledge issues. There were about 10% of responses in both groups reporting no difficulties.

Table 6 Comparison of experimental and control group responses to the difficulties in solving the problem

Themes/Categories	Illustrative examples	Frequency	
		Experimental group (N = 68) K (%)	Control group (N = 67) K (%)
1. Difficulty in generating hypotheses	<i>Formulate hypotheses</i>	7 (10%)	6 (9%)
2. Difficulty in handling multiple hypotheses	<i>There were so many possibilities that we did not know where to start hypothesizing.</i>	5 (7%)	1 (1%)
3. Difficulty in investigating the root cause	<i>Could only find the surface cause; but could not find the root cause.</i>	2 (3%)	0
4. Difficulty in processing a lot of intertwined and complex data	<i>There were so much information and data that it was hard to process and synthesize them.</i>	14 (20%)	9 (13%)
5. Difficulty in reasoning with data and justifying hypotheses	<i>The compiled evidence could not prove the hypothesis.</i>	14 (20%)	26 (39%)
6. Difficulty in making a conclusion or summary	<i>Unable to draw conclusions from disorganized thinking</i>	8 (12%)	3 (4%)
7. Challenge in group work	<i>It was hard for the diverse ideas of group members to converge.</i>	3 (4%)	6 (9%)
8. Difficulty in drawing the map	<i>Draw the reasoning map</i> (Experimental group) <i>Drawing the concept map</i> (Control group)	3 (4%)	5 (7%)
9. Others			
9.1. Technical issue	<i>My computer did not work.</i>	3 (4%)	1 (1%)
9.2. Time issue	<i>Need more time to complete the task</i>	2 (3%)	0
9.3. Knowledge issue	<i>We had insufficient theoretical knowledge.</i>	1 (1%)	3 (4%)
10. No difficulty		6 (9%)	7 (10%)

Note. N = total number of responses (One participant might write more than one responses, or one response might convey more than one themes). K = number of responses to each theme/category. % = the percentage of responses.

As seen from the results in Table 6, the two main difficulties reported by both groups were processing a lot of intertwined and complex data, and reasoning and justifying hypotheses. Besides, both groups had problems with generating hypotheses. Challenges in

group work and drawing the map were also experienced by several students in both groups.

It seemed that more students in the control group reported difficulty in reasoning and justifying hypotheses, and coordinating diverse ideas among group members. Compared to the control group, experimental group mentioned more about the challenges in processing a lot of intertwined data, and making a conclusion or summarize the finding. Although both groups reported difficulties in hypothesizing, the responses of the experimental group involved more aspects of hypothesizing including handling multiple hypotheses and investigating the root cause, showing their engagement in high-order thinking when exploring hypotheses.

As seen in Table 7, regarding student responses to the second question on the benefits of the 3DTG (or concept map for the control condition), the most salient benefit was that the map made their thinking and reasoning visible. In addition, the two groups shared similar views that using the maps enabled them to hypothesize in a logical way and perform progressive reasoning.

However, there were some benefits reported only by the experimental group but absent in control group's responses, including visualizing the hypothesizing process in a global view, promoting reasoning globally, fostering reflection during the reasoning process, and facilitating conclusion-making and communication. Another difference was that 21% of the responses from the control group pointed out that concept mapping was not beneficial to their inquiry learning. In particular, one student in the control group mentioned that "*Although concept map facilitated the hypothesizing, it could not represent detailed information and evidence.*"

Table 7 Comparison of experimental and control group responses to the benefits of the mapping or thinking tool

Themes/Categories	Illustrative examples	Frequency	
		Experimental group (N = 47)	Control group (N = 33)

		K (%)	K (%)
1. Data collection			
1.1. Visualizing the data	<i>Making the data visible</i>	1 (2%)	0
1.2. Viewing the data in a global manner	<i>Facilitating the integration of the information</i>	1 (2%)	0
2. Hypothesis formulation			
2.1. Making multiple hypotheses and their relations visible	<i>Recording our hypotheses; visualizing the relations between the hypotheses</i>	4 (9%)	0
2.2. Analyzing multiple hypotheses in a big picture	<i>Facilitating our analysis of the fish die-off problem from multiple dimensions</i>	3 (6%)	0
2.3. Hypothesizing in a logical way	<i>Helping us clarify our thinking and find the right direction to explain the problem</i>	1 (2%)	1 (3%)
3. Data analysis and reasoning			
3.1. Making thinking visible	<i>We could see the causal relationships and easily justify our hypotheses.</i>	28 (60%)	24 (73%)
3.2. Reasoning globally	<i>Seeing the linkages between multiple events more clearly</i>	3 (6%)	
3.3. Progressive Reasoning	<i>The map guided us in making reasoning step-by-step and progressively.</i>	1 (2%)	1 (3%)
3.4. Reflection	<i>Helping us to figure out errors in previously formed hypotheses</i> <i>Helping notice easily-ignored data</i>	2 (4%)	
4. Making conclusion	<i>Made it easier to reach a conclusion from visualized relations.</i>	1 (2%)	
5. Communication	<i>Making others easily understand my ideas</i>	2 (4%)	
6. No benefits	<i>No.</i>	0	7 (21%)

## Discussion

The data analysis results for the overall inquiry task performance shows that the groups in the experimental condition performed better than their counterparts in the control condition. This was consistent with previous studies (Janssen et al., 2010; Slof, Erkens, Kirschner, & Helms-Lorenz, 2013; Suthers et al., 2008), which also proved the effectiveness of representational support for group performance in solving complex problems. The 3DTG provided learners with an overview of the entire inquiry task so that they could easily access the task from different perspectives, i.e., the problem's information and data, subject

knowledge underlying the problem, and the logic behind the reasoning. This made it relatively easier for learners to successfully solve the problem (Janssen et al., 2010).

More specifically, student groups using the 3DTG performed significantly better in almost all aspects of the inquiry process except for making conclusions. First, in formulating hypotheses, the experimental groups significantly outperformed their counterparts. One plausible explanation might be that the 3DTG enabled the learners to develop a logical set of hypotheses on the basis of the hypothetical reasoning processes articulated in the reasoning map, and the relevant data and knowledge represented in data table and concept map, respectively. Students' responses to the open-ended survey supported this explanation. While both groups reported difficulties in hypothesizing, the experimental group reported that they benefitted more from the 3DTG in visualizing multiple hypotheses and their relations in a big picture. In addition, the difficulties reported by students in the experimental group involved more aspects of hypothesizing including exploring multiple hypotheses and the root cause of the problem, showing their engagement in high-order thinking when exploring hypotheses. Such engagement enabled them to generate sufficient reasonable hypotheses in their inquiry report. Moreover, the 3DTG enabled students to review and revise previously formed hypotheses. The presence of rejected hypotheses meant that students could consider information or data that conflicted with their previously formed ideas. Conversely, students in the control condition expressed their confusions over when and how to reject a hypothesis. The finding confirmed the belief that well-designed external representations assist problem solving by re-ordering and refining the ideas in ways that are helpful to find the solution (Cox, 1999).

Second, the 3DTG significantly improved the experimental group's reasoning performance. This could be caused by the reasoning map in the 3TDG, which focused the students' attention on the reasoning activities (e.g., developing logical reasoning by finding

confirming and disconfirming evidence, thinking in an organized and logical way), thereby making the reasoning salient and easy to watch (Toth et al., 2002). These views were expressed in the experimental group's responses to the 3DTG's benefits. Research also has indicated that visual representations help learners explicitly clarify their thinking (Brna, Cox, & Good, 2001).

In particular, the 3DTG helped learners to clearly see the relationships between different pieces of information. Some students in the experimental condition said that drawing data table helped them notice easily-ignored data. Looking at the data table enabled the learners to summarize all possible pieces of evidence for or against a hypothesis without missing anything. This confirms the view that proper external representations help make information explicit (Cox, 1999). Studies have also indicated that the diagrams/graphs are quite useful for relating information (Suthers & Hundhausen, 2002). Further, based on the explicit information, the learners could easily reflect on the reasoning process, which was only mentioned in the experimental group's responses.

More importantly, requiring students to list the confirming and disconfirming evidence might have stimulated them to search for counterevidence (Janssen et al., 2010). Students have reported that finding counterargument is difficult in scientific reasoning (Kuhn et al., 2000). In this study, the control group students said concept mapping could not embody detailed information and evidence. In fact, some of their submitted reports failed to provide enough data or adequately explain the cause-and-effect relationship. The survey data also indicated that more students in the control group reported difficulty in reasoning and justifying hypotheses.

Although concept maps have the capability to make thinking visible, concept mapping alone is inadequate in eliciting and representing the process of applying knowledge to solve a problem (Wang et al., 2017). The 3DTG has greater potential to visualize and facilitate



learners' hypothetical reasoning and argumentation by integrating the concept map, data table, and reasoning map in a holistic way. The responses to the open-ended survey also showed that 3DTG promoted reasoning globally by making visible the required information, underlying concepts, and the reasoning process. All of these might better explain why the experimental groups received higher scores for reasoning and justification.

Third, with regard to the effects on drawing conclusions, there was no significant difference between the two conditions. Research has also suggested the effectiveness of concept mapping in facilitating the convergence of conclusions among group members (Gijlers & de Jong, 2013; Novak, 2002; Suthers et al., 2008). This indicated that the 3DTG and concept map were equally beneficial to drawing conclusions. Although the experimental groups did not perform significantly better in drawing conclusions, they got significantly higher scores for generating hypotheses and reasoning. This suggested that students using the 3DTG developed conclusions on the basis of more systematic thinking and reasoning and sufficient evidence, which is critical to inquiry learning. Hypothesizing and reasoning are fundamental inquiry skills (Toth et al., 2002). Drawing conclusions grounded in systematic thinking and reasoning has greater potential than jumping to correct conclusions with inadequate reasoning. Once students learn to formulate hypotheses and reason, their competence may transfer to other similar scientific activities.

Fourth, with respect to the students' perceptions of consensus building, the result indicated quite positive perceptions, despite the non-significant difference between the two conditions. This was consistent with the findings of Gijlers et al. (2009), which also indicated that encouraging learners to collaboratively build an external representation facilitated consensus building activities. The finding suggested that the 3DTG can share equal value in some aspects with concept map.

Finally, the better task performance of students in the experimental condition might lead

to their significantly higher scores in the post knowledge test than their counterparts. As students using the 3DTG deeply reflected on the relationships between the domain concepts and events, they had better understanding of the causal relationships in the ecosystems. Research has also suggested that reasoning and writing activities help students integrate new information with prior knowledge to develop deep, contextualized, and applicable knowledge (Keselman, Kaufman, Kramer, & Patel, 2007; Keys, 2000). On the other hand, reasoning ability influences achievement (Lawson et al., 2007); therefore, higher reasoning ability shown by the experimental group students may also explain their better achievement in the post-test. In addition, according to the survey results, slightly more students in the control group reported issues with knowledge.

### **Conclusion**

In science learning, students have often had difficulties engaging in fruitful inquiry learning. In addition to the difficulty involved with regulating the inquiry process, they have encountered problems generating hypotheses and carrying out scientific reasoning with intertwined variables. Despite the availability of different kinds of support or guidance (e.g., structuring tasks, using prompts and hints), learners might still have found it cognitively demanding to successfully complete inquiry and problem-solving tasks, which have typically combined all kinds of information and data, concepts, and relationships and have involved complex hypothesizing and reasoning processes using the data and knowledge. This study proposed and investigated the effects of the *3DTG* which allowed learners to articulate information, the relevant concepts and their relationships, and the hypothesizing and reasoning processes of exploring the problem in a holistic picture to support inquiry learning.

The findings show that the students using the 3DTG achieved significantly better knowledge and performed significantly better on the group inquiry task, suggesting the benefits of constructing the 3DTG to support inquiry learning and complex problem solving.

Specifically, the 3DTG provided learners with an overview of the inquiry task, and guides them in generating hypotheses step-by-step, developing evidence-based reasoning based on relevant data and knowledge, and knowing how and where to revise previously formed but unreasonable hypotheses or ideas. By incorporating a problem's data, subject knowledge, and the hypothesizing and reasoning process in a holistic visual representation, the proposed approach demonstrated its promising effects in supporting inquiry learning especially in facilitating the formulation of hypotheses and the reasoning process with complex problems. On the other hand, the 3DTG can share equal value with concept mapping in terms of drawing conclusions and consensus building.

### **Implications**

This study has several implications for designing support or guidance for inquiry learning in computer-supported environments. First, it is important to scaffold students when they engage in a complex inquiry task. Visually representing in a holistic picture the concepts of subject knowledge, problem information and data, and the progressive process of evidence-based reasoning and hypothesizing based on relevant data and knowledge can facilitate students' inquiry performance. Externalizing the key cognitive aspects (e.g., searching for information and data, constructing domain knowledge, and hypothesizing and reasoning) in a holistic way is critical for successful inquiry learning. Second, the choice of representational tools should align with the characteristics and demands of the problem or task (Cox, 1999; Slob, Erkens, Kirschner, & Jaspers, 2010). The 3DTG used in this study has shown promising effects in externalizing and facilitating complex hypothesizing of causal relations and reasoning on the basis of relevant data and subject knowledge. However, this representational facility might not automatically apply to all inquiry tasks without adaptation. Third, formative assessment of the learning process plays an important role in inquiry learning. The findings of this study indicate that students may reach correct conclusions that

are not grounded in systematic thinking and reasoning. Learning the inquiry process is a more important outcome than correctly solving the problem. Teachers may need to assess the students' thinking and reasoning processes to understand their inquiry performance better, find the specific difficulties encountered by them, and make further improvement to the scaffolding of inquiry learning.

### **Limitations and future work**

Some issues or limitations exist in the current study. The group consensus building was measured by a self-reported questionnaire, which might have been insufficient to evaluate consensus building activity. A more objective way would be to assess it during the learning process on the basis of social interaction or group discussion records, just as Gijlers and de Jong (2013) did or video record group discussion processes. Examining the discussion process might lead to a fuller understanding of the effects of constructing the 3DTG on group consensus building and drawing conclusions.

To improve the effectiveness of the 3DTG for drawing conclusions, some additional improvements may be needed. One possible enhancement is to explicitly remind students to critique and revisit their constructed maps. Research has indicated that combining constructing and critiquing concept maps can support the integration of ideas (Schwendimann & Linn, 2016). These issues will be addressed in future work. In addition, more research is needed to verify the effects of the 3DTG. As establishing the validity of an instrumentation tool is an ongoing process (Lederman, Abd-El-Khalick, Bell, & Schwartz, 2002), further validity tests of the questionnaire and rubrics for assessing the inquiry report would be needed.

### **References**

- Albanese, M. A., & Mitchell, S. (1993). Problem-based learning: a review of literature on its outcomes and implementation issues. *Academic Medicine*, 68(1), 52-81.
- Bell, T., Urhahne, D., Schanze, S., & Ploetzner, R. (2010). Collaborative inquiry learning: Models, tools, and challenges. *International Journal of Science Education*, 32(3),

- 349-377. doi:10.1080/09500690802582241
- Bransford, J., Brown, A. L., & Cocking, R. E. (1999). *How people learn : brain, mind, experience, and school*. Washington, D.C.: National Academy Press.
- Brna, P., Cox, R., & Good, J. (2001). Learning to think and communicate with diagrams: 14 questions to consider. *Artificial Intelligence Review*, 15(1-2), 115-134. doi:10.1023/A:1006584801959
- Chinn, C. A., & Brewer, W. F. (1993). The Role of Anomalous Data in Knowledge Acquisition: A Theoretical Framework and Implications for Science Instruction. *Review of Educational Research*, 63(1), 1-49. doi:10.3102/00346543063001001
- Cohen, J. (1992). Statistical Power Analysis. *Current Directions in Psychological Science*, 1(3), 98-101.
- Cox, R. (1999). Representation construction, externalised cognition and individual differences. *Learning and Instruction*, 9(4), 343-363.
- Davis, E. A. (2003). Prompting middle school science students for productive reflection: Generic and directed prompts. *Journal of the Learning Sciences*, 12(1), 91-142.
- de Jong, T., & van Joolingen, W. R. (1998). Scientific discovery learning with computer simulations of conceptual domains. *Review of Educational Research*, 68(2), 179-201.
- Furtak, E. M., Seidel, T., Iverson, H., & Briggs, D. C. (2012). Experimental and Quasi-Experimental Studies of Inquiry-Based Science Teaching: A Meta-Analysis. *Review of Educational Research*, 82(3), 300-329. doi:10.3102/0034654312457206
- Geier, R., Blumenfeld, P. C., Marx, R. W., Krajcik, J. S., Fishman, B., Soloway, E., & Clay-Chambers, J. (2008). Standardized test outcomes for students engaged in inquiry-based science curricula in the context of urban reform. *Journal of Research in Science Teaching*, 45(8), 922-939. doi:10.1002/tea.20248
- Germann, P. J., & Aram, R. J. (1996). Student Performances on the Science Processes of Recording Data, Analyzing Data, Drawing Conclusions, and Providing Evidence. *Journal of Research in Science Teaching*, 33(7), 773-798. doi:10.1002/(SICI)1098-2736(199609)33:7<773::AID-TEA5>3.0.CO;2-K
- Gijbels, D., Dochy, F., Van den Bossche, P., & Segers, M. (2005). Effects of Problem-Based Learning: A Meta-Analysis from the Angle of Assessment. *Review of Educational Research*, 75(1), 27-61.
- Gijlers, H., & De Jong, T. (2005). The relation between prior knowledge and students' collaborative discovery learning processes. *Journal of Research in Science Teaching*, 42(3), 264-282. doi:10.1002/tea.20056
- Gijlers, H., & de Jong, T. (2013). Using Concept Maps to Facilitate Collaborative Simulation-Based Inquiry Learning. *Journal of the Learning Sciences*, 22(3), 340-374. doi:10.1080/10508406.2012.748664
- Gijlers, H., Saab, N., Van Joolingen, W. R., De Jong, T., & Van Hout-Wolters, B. H. A. M. (2009). Interaction between tool and talk: How instruction and tools support consensus building in collaborative inquiry-learning environments. *Journal of Computer Assisted Learning*, 25(3), 252-267. doi:10.1111/j.1365-2729.2008.00302.x
- Grotzer, T. A., & Bell Basca, B. (2003). How does grasping the underlying causal structures of ecosystems impact students' understanding? *Journal of Biological Education*, 38(1), 16-29.
- Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry learning: A response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist*, 42(2), 99-107.
- Hogan, K., & Maglienti, M. (2001). Comparing the epistemological underpinnings of students' and scientists' reasoning about conclusions. *Journal of Research in Science Teaching*, 38(6), 663-687. doi:10.1002/tea.1025

- Hsu, C. C., Chiu, C. H., Lin, C. H., & Wang, T. I. (2015). Enhancing skill in constructing scientific explanations using a structured argumentation scaffold in scientific inquiry. *Computers and Education, 91*, 46-59. doi:10.1016/j.compedu.2015.09.009
- Janssen, J., Erkens, G., Kirschner, P. A., & Kanselaar, G. (2010). Effects of representational guidance during computer-supported collaborative learning. *Instructional Science, 38*(1), 59-88. doi:10.1007/s11251-008-9078-1
- Jonassen, D. H. (1997). Instructional design models for well-structured and III-structured problem-solving learning outcomes. *Educational Technology Research and Development, 45*(1), 65-94.
- Kamarainen, A. M., Metcalf, S., Grotzer, T., & Dede, C. (2015). Exploring Ecosystems from the Inside: How Immersive Multi-user Virtual Environments Can Support Development of Epistemologically Grounded Modeling Practices in Ecosystem Science Instruction. *Journal of Science Education and Technology, 24*(2-3), 148-167. doi:10.1007/s10956-014-9531-7
- Keselman, A., Kaufman, D. R., Kramer, S., & Patel, V. L. (2007). Fostering conceptual change and critical reasoning about HIV and AIDS. *Journal of Research in Science Teaching, 44*(6), 844-863. doi:10.1002/tea.20173
- Keys, C. W. (2000). Investigating the thinking processes of eighth grade writers during the composition of a scientific laboratory report. *Journal of Research in Science Teaching, 37*(7), 676-690.
- Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why Minimal Guidance During Instruction Does Not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching. *Educational Psychologist, 41*(2), 75-86. doi:10.1207/s15326985ep4102\_1
- Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science, 12*(1), 1-48. doi:10.1016/0364-0213(88)90007-9
- Kollar, I., Fischer, F., & Slotta, J. D. (2007). Internal and external scripts in computer-supported collaborative inquiry learning. *Learning and Instruction, 17*(6), 708-721. doi:10.1016/j.learninstruc.2007.09.021
- Krathwohl, D. R. (1998). *Methods of educational and social science research: An integrated approach* (2nd ed. ed.). New York: Longman.
- Kuhn, D., Black, J., Keselman, A., & Kaplan, D. (2000). The development of cognitive skills to support inquiry learning. *Cognition and Instruction, 18*(4), 495-523.
- Lawson, A. E. (2005). What is the role of induction and deduction in reasoning and scientific inquiry? *Journal of Research in Science Teaching, 42*(6), 716-740. doi:10.1002/tea.20067
- Lawson, A. E., Banks, D. L., & Logvin, M. (2007). Self-efficacy, reasoning ability, and achievement in college biology. *Journal of Research in Science Teaching, 44*(5), 706-724. doi:10.1002/tea.20172
- Lazonder, A. W., & Harmsen, R. (2016). Meta-Analysis of Inquiry-Based Learning: Effects of Guidance. *Review of Educational Research, 86*(3), 681-718. doi:10.3102/0034654315627366
- Lederman, N. G., Abd-El-Khalick, F., Bell, R. L., & Schwartz, R. S. (2002). Views of Nature of Science Questionnaire: Toward Valid and Meaningful Assessment of Learners' Conceptions of Nature of Science. *Journal of Research in Science Teaching, 39*(6), 497-521. doi:10.1002/tea.10034
- Linn, M. C., & Slotta, J. D. (2000). WISE science. *Educational Leadership, 58*(2), 29-32.
- Linn, M. C., & Songer, N. B. (1991). Teaching thermodynamics to middle school students: What are appropriate cognitive demands? *Journal of Research in Science Teaching, 28*(10), 885-918. doi:10.1002/tea.3660281003

- Markow, P. G., & Lonning, R. A. (1998). Usefulness of Concept Maps in College Chemistry Laboratories: Students' Perceptions and Effects on Achievement. *Journal of Research in Science Teaching*, 35(9), 1015-1029.
- Njoo, M., & De Jong, T. (1993a). Exploratory learning with a computer simulation for control theory: Learning processes and instructional support. *Journal of Research in Science Teaching*, 30(8), 821-844. doi:10.1002/tea.3660300803
- Njoo, M., & de Jong, T. (1993b). Supporting exploratory learning by offering structured overviews of hypotheses. In T. d. J. D. Towne, & H. Spada (Ed.), *Simulation-based experiential learning* (pp. 207-225). Berlin, Germany: Springer-Verlag.
- Noroozi, O., Teasley, S. D., Biemans, H. J. A., Weinberger, A., & Mulder, M. (2013). Facilitating learning in multidisciplinary groups with transactive CSCL scripts. *International Journal of Computer-Supported Collaborative Learning*, 8(2), 189-223. doi:10.1007/s11412-012-9162-z
- Novak, J. D. (2002). Meaningful learning: The essential factor for conceptual change in limited or inappropriate propositional hierarchies leading to empowerment of learners. *Science Education*, 86(4), 548-571.
- Novak, J. D., Bob Gowin, D., & Johansen, G. T. (1983). The use of concept mapping and knowledge vee mapping with junior high school science students. *Science Education*, 67(5), 625-645. doi:10.1002/sce.3730670511
- Pedaste, M., de Jong, T., Sarapuu, T., Piksööt, J., van Joolingen, W. R., & Giemza, A. (2013). Investigating ecosystems as a blended learning experience. *Science*, 340(6140), 1537-1538.
- Penner, D. E. (2000). Explaining systems: Investigating middle school students' understanding of emergent phenomena. *Journal of Research in Science Teaching*, 37(8), 784-806.
- Quintana, C., Reiser, B., Davis, E. A., Krajcik, J., Fretz, E., Duncan, R. G., . . . Soloway, E. (2004). A scaffolding design framework for software to support science inquiry. *Journal Of The Learning Sciences*, 13(3), 337-386.
- Roth, W. M., & McGinn, M. K. (1998). Inscriptions: Toward a theory of representing as social practice. *Review of Educational Research*, 68(1), 35-59.
- Roth, W. M., & Roychoudhury, A. (1993). The concept map as a tool for the collaborative construction of knowledge: A microanalysis of high school physics students. *Journal of Research in Science Teaching*, 30(5), 503-534. doi:10.1002/tea.3660300508
- Ruiz-Primo, M. A., Schultz, S. E., Li, M., & Shavelson, R. J. (2001). Comparison of the reliability and validity of scores from two concept-mapping techniques. *Journal of Research in Science Teaching*, 38(2), 260-278.
- Ruiz-Primo, M. A., & Shavelson, R. J. (1996). Problems and Issues in the Use of Concept Maps in Science Assessment. *Journal of Research in Science Teaching*, 33(6), 569-600.
- Sandoval, W. A. (2003). Conceptual and epistemic aspects of students' scientific explanations. *Journal of the Learning Sciences*, 12(1), 5-51.
- Schwendimann, B. A., & Linn, M. C. (2016). Comparing two forms of concept map critique activities to facilitate knowledge integration processes in evolution education. *Journal of Research in Science Teaching*, 53(1), 70-94. doi:10.1002/tea.21244
- Shin, N., Jonassen, D. H., & McGee, S. (2003). Predictors of well - structured and ill - structured problem solving in an astronomy simulation. *Journal of Research in Science Teaching*, 40(1), 6-33.
- Shute, V. J., & Glaser, R. (1990). A large-scale evaluation of an intelligent discovery world: Smithtown. *Interactive Learning Environments*, 1(1), 51-77.
- Singer, J., Marx, R. W., Krajcik, J., & Clay Chambers, J. (2000). Constructing extended

- inquiry projects: Curriculum materials for science education reform. *Educational Psychologist*, 35(3), 165-178.
- Slof, B., Erkens, G., Kirschner, P. A., & Helms-Lorenz, M. (2013). The effects of inspecting and constructing part-task-specific visualizations on team and individual learning. *Computers and Education*, 60(1), 221-233. doi:10.1016/j.compedu.2012.07.019
- Slof, B., Erkens, G., Kirschner, P. A., Janssen, J., & Jaspers, J. G. M. (2012). Successfully carrying out complex learning-tasks through guiding teams' qualitative and quantitative reasoning. *Instructional Science*, 40(3), 623-643. doi:10.1007/s11251-011-9185-2
- Slof, B., Erkens, G., Kirschner, P. A., & Jaspers, J. G. M. (2010). Design and effects of representational scripting on group performance. *Educational Technology Research and Development*, 58(5), 589-608. doi:10.1007/s11423-010-9148-3
- Stoyanov, S., & Kommers, P. (2006). WWW-intensive concept mapping for metacognition in solving ill-structured problems. *International Journal of Continuing Engineering Education and Life Long Learning*, 16(3), 297-316.
- Stratford, S. J., Krajcik, J., & Soloway, E. (1998). Secondary students' dynamic modeling processes: Analyzing, reasoning about, synthesizing, and testing models of stream ecosystems. *Journal of Science Education and Technology*, 7(3), 215-234.
- Suthers, D. D., & Hundhausen, C. D. (2002). *The effects of representation on students' elaborations in collaborative inquiry*. Paper presented at the Proceedings of the Conference on Computer Support for Collaborative Learning: Foundations for a CSCL Community.
- Suthers, D. D., & Hundhausen, C. D. (2003). An experimental study of the effects of representational guidance on collaborative learning processes. *Journal of the Learning Sciences*, 12(2), 183-218.
- Suthers, D. D., Vatrupu, R., Medina, R., Joseph, S., & Dwyer, N. (2008). Beyond threaded discussion: Representational guidance in asynchronous collaborative learning environments. *Computers and Education*, 50(4), 1103-1127. doi:10.1016/j.compedu.2006.10.007
- Toth, E. E., Suthers, D. D., & Lesgold, A. M. (2002). "Mapping to Know": The Effects of Representational Guidance and Reflective Assessment on Scientific Inquiry. *Science Education*, 86(2), 264-286. doi:10.1002/sci.10004
- Van Bruggen, J. M., Kirschner, P. A., & Jochems, W. (2002). External representation of argumentation in CSCL and the management of cognitive load. *Learning and Instruction*, 12(1), 121-138. doi:10.1016/S0959-4752(01)00019-6
- Van Joolingen, W. R., & De Jong, T. (1997). An extended dual search space model of scientific discovery learning. *Instructional Science*, 25(5), 307-346.
- van Joolingen, W. R., Savelsbergh, E. R., de Jong, T., Lazonder, A. W., & Manlove, S. (2005). Co-Lab: Research and development of an online learning environment for collaborative scientific discovery learning. *Computers in Human Behavior*, 21(4), 671-688. doi:10.1016/j.chb.2004.10.039
- Wang, M., Cheng, B., Chen, J., Mercer, N., & Kirschner, P. A. (2017). The use of web-based collaborative concept mapping to support group learning and interaction in an online environment. *Internet and Higher Education*, 34, 28-40. doi:10.1016/j.iheduc.2017.04.003
- Wang, M., Wu, B., Kinshuk, Chen, N. S., & Spector, J. M. (2013). Connecting problem-solving and knowledge-construction processes in a visualization-based learning environment. *Computers and Education*, 68, 293-306. doi:10.1016/j.compedu.2013.05.004
- Wecker, C., Rachel, A., Heran-Dörr, E., Waltner, C., Wiesner, H., & Fischer, F. (2013).



- Presenting theoretical ideas prior to inquiry activities fosters theory-level knowledge. *Journal of Research in Science Teaching*, 50(10), 1180-1206. doi:10.1002/tea.21106
- Weinberger, A., Stegmann, K., & Fischer, F. (2007). Knowledge convergence in collaborative learning: Concepts and assessment. *Learning and Instruction*, 17(4), 416-426. doi:10.1016/j.learninstruc.2007.03.007
- White, B. Y. (1993). ThinkerTools: Causal models, conceptual change, and science education. *Cognition and instruction*, 10(1), 1-100.
- Windschitl, M. (2004). Folk theories of "inquiry:" How preservice teachers reproduce the discourse and practices of an atheoretical scientific method. *Journal of Research in Science Teaching*, 41(5), 481-512.
- Zeineddin, A., & Abd-El-Khalick, F. (2010). Scientific reasoning and epistemological commitments: Coordination of theory and evidence among college science students. *Journal of Research in Science Teaching*, 47(9), 1064-1093. doi:10.1002/tea.20368
- Zimmerman, C., Raghavan, K., & Sartoris, M. L. (2003). The impact of the MARS curriculum on students' ability to coordinate theory and evidence. *International Journal of Science Education*, 25(10), 1247-1271. doi:10.1080/0950069022000038303