Education Tech Research Dev
https://doi.org/10.1007/s11423-024-10433-2

RESEARCH ARTICLE

From experience to explanation: an analysis of students' use of a wildfire simulation

Trudi Lord[1] · **Paul Horwitz**[1] · **Amy Pallant**[1] · **Christopher Lore**[1]

Accepted: 1 November 2024

Abstract

This study employs the Experiential Learning Theory framework to investigate students' use of a wildfire simulation. We analyzed log files automatically generated by middle and high school students ($n = 1515$) as they used a wildfire simulation and answered associated prompts in three simulation-based tasks. We first analyzed students' log files to determine which, if any, measure of simulation use–quantity of runs, variation in runs, or quality of experimentation setup–predicted their scores on responses to observation and explanation questions that followed the simulation experience. Of the three measures, only the quality of students' simulation use was significantly correlated to their written explanation scores in all three tasks. Further investigation into the sequence in which students used the simulation and answered the questions revealed two common patterns in between two-thirds and three-quarters of the students in each task: (1) students ran the simulation and then answered the observation and explanation questions in that order or (2) students ran the simulation, answered the observation question, ran the simulation again, and then answered the explanation question. While there was no clear relationship between these two patterns and students' scores on the explanation question, this finding has resulted in an updated experiential learning framework specific to simulation use. Implications for the design of scaffolding for future simulation-based learning experiences around natural hazards are discussed.

Keywords Simulations · Modeling · Experiential learning · Trace data · Natural hazards

Introduction

Wildfires rage. More and more acres are burning each year, coming dangerously close to highly populated areas across the globe (Hoover & Hanson, 2023; Turco et al., 2023). Wildfire hazards, such as polluted air and decreased visibility from smoke, are now impacting areas that were previously too far from fuel sources to be affected (Warneke et al., 2023). As natural hazards such as wildfires increase due to climate change (Melia et al., 2022), the future demands scientists who are well-versed in the environmental factors

✉ Trudi Lord
tlord@concord.org

1 Concord Consortium, 25 Love Lane, Concord, MA 01742, USA

Published online: 27 November 2024

involved in wildfire spread and propagation as well as prevention and mitigation techniques to reduce risks and impacts on communities and ecosystems (Alderson et al., 2022). Students in Earth and environmental science classrooms today will be the scientists that face these challenges tomorrow.

Wildfire season now comprises the entire year with reports of devastating fires in Greece in August, Canada in October, and Australia in December. Other related natural disasters such as flooding and mudslides that often follow after wildfires have destroyed forests and compromised soils (Ren et al., 2011; Tiwari et al., 2020). These are only a few of the many natural hazards that have been exacerbated by a changing climate; hurricanes, floods, and droughts have also become more severe (Alifu et al., 2022; Garner, 2023; Hoell et al., 2022). Consequently, the United Nations has declared that the public must have access to science-based education on natural disasters and has suggested that this education be provided via technology-based interventions (Center 2015). Global interest in teaching about natural hazards, risks, and resilience has produced a number of computer-based tools and curricula to teach about hurricanes (Luo et al., 2008), coastal flooding (Taillandier & Adam, 2018), volcanic eruptions (Lore et al., 2023; Mani et al., 2016), earthquakes (Lore & Seevers, 2022; Moutinho et al., 2017), and wildfire risk and impacts (Lord et al., 2024). Previous research on wildfire education has focused primarily on fire safety and prevention programs (Ballard et al., 2016; McCaffrey, 2015; Monroe et al., 2016).

This study investigated the relationship between students' interactions with a wildfire simulation and their ability to make meaning from their experience. To do this, we analyzed the logs generated by students' interactions with the simulation in three inquiry-based, scaffolded tasks. We also analyzed students' observations of simulation outputs (via multiple-choice questions) and their written explanations (via open-response questions) to assess whether or not they understood the concepts presented in specific simulation experiences. Finally, we examined the sequence in which students used the simulation and answered the questions to determine if that sequence predicted their conceptual understanding.

Literature review

Simulations can be used to teach students about environmental science (Kukkonen et al., 2014; Petersson et al., 2013) including natural disasters (Taillandier & Adam, 2018), which typically cannot be taught in the field. Simulations provide an alternate way for students to experiment with natural phenomena that cannot be observed in the classroom due to time scales and variables that are difficult or dangerous to manipulate (Feurtzeig & Roberts, 1999; Horwitz, 1996, 1999; Horwitz & Christie, 2000; Pallant & Lee, 2015, 2017; Pallant et al., 2023; Petersen et al., 2020; Quellmalz et al., 2012). Through the use of a simulation, a student can have "realistic experiences from which to gain and manipulate knowledge to understand better the relationship between the concepts being investigated" (Widiyatmoko, 2018, p. 38). The ability to experiment by adjusting initial simulation inputs and observing different simulation outputs gives students agency in their learning process (Windschitl & Andre, 1998). Furthermore, virtual experiments that include simulations can both help students build a mental model of scientific phenomena and motivate them to learn (Dede et al., 2005). Holzinger et al. (2024) advocate the use of simulations in forestry education for improving safety and enhancing communication between stakeholders and the public.

In recent years, multiple studies have shown that students' understanding of science concepts can be fostered by using simulations (Abdullah & Shariff, 2008; Plass et al., 2012; Ramasundaram et al., 2005; Sarabando et al., 2014; Scalise et al., 2011; Widiyatmoko, 2018). A small group of researchers who study educational simulations have attempted to analyze the clickstream data generated by students' use of simulations and other online tools (Buckley et al., 2006; Gobert et al., 2012; McElhaney & Linn, 2011; Wilson et al., 2018). Various data mining techniques have been developed that look closely at trace data (Baker & Yacef, 2009; Gobert et al., 2013; Martin & Sherin, 2013). More recently, researchers have mined the data generated by the use of online simulations to answer research questions around a broad variety of topics, such as exploring students' problem-solving strategies (Greiff et al., 2018), training intelligent tutoring systems (Henderson et al., 2020), providing students with automated feedback (Lee et al., 2021), and assessing student learning (Horwitz et al., 2023; Lord et al., 2024).

While it is not possible for researchers to simultaneously and directly observe how every student in the classroom is using a simulation, timestamped logs of students' mouse clicks can be used to measure their interactions. Such logs can be used to measure dosage effect. For example, Wilson et al. (2018) found that the more students used a simulation, the better they learned concepts, as demonstrated by pre- to post-test scores. In contrast, McElhaney and Linn (2011) found that the number of simulation runs and variation in runs did not predict performance. Logs can also be used to determine the quality of students' interactions with a simulation. For example, in a study by Buckley et al. (2006), researchers examined the actions of students who ran a genetics simulation in a few specific tasks within a larger module. They found that students who ran the simulation in certain tasks multiple times systematically (i.e., changing one variable at a time) had higher pre- to post-test gains than students who ran the simulation multiple times unsystematically (i.e., not changing any variables over multiple runs). Similarly, Gobert et al. (2012) found that students who ran single-variable experiments with a simulation had higher learning gains than students who experimented unsystematically.

A previous study by Lord et al. (2024) also found that students' systematic use of 10 simulations included in a wildfire unit was correlated to gains from pre- to post-test; however, that study was limited as it did not dive into the specific simulation tasks to, for example, investigate the correlation between simulation use and written explanations of scientific phenomena. The current study builds upon that work and the work of others who have investigated the correlation between model interactions and students' ability to write explanations. For example, Scalise and Clarke-Midura (2018) studied log data of students' investigations in a virtual world, connecting students' inquiry process with their ability to write scientific explanations. Lee et al. (2021)'s study of students' use of a groundwater simulation parsed log files in real time and compared student actions in a simulation with their written arguments for the purpose of providing automated feedback.

Theoretical framework

A combination of multiple learning theories, Kolb's (1984) Experiential Learning Theory (ELT) reflects prior work by Dewey, Piaget, Vygotsky, and others that focus on student-driven learning experiences. ELT is a cycle that includes four parts. The student must have a *concrete experience to start.* Then, they must make *observations* of their experience, which can be descriptive or reflective. In the third step, students *explain* what they observed and pull concepts together. In the final step, students engage in further *experimentation* to test out their ideas. Traditionally, the concrete experiences that start the cycle have been hands-on, but any novel task, including those presented virtually, can initiate the cycle of observation and contemplation (Enns, 1993). Computer simulations give students the opportunity to experience a phenomenon as well as to "exercise higher order capabilities such as reflective thinking and abstract conceptualisation" (Falloon, 2019, p. 138). The modified ELT framework shown in Fig. 1 was developed to describe students' use of a wildfire simulation (Lord et al., 2024).

This paper reports a detailed examination of this modified ELT framework by fine-grained observation of students' experiences with simulation-based tasks embedded in an online wildfire unit. In each task, students started a simulation experience by setting variables and then running an experiment. As the simulation ran, students were able to observe what happens as fire spreads on the landscape, reflect on what they see, and answer a multiple-choice question based on their observations. In the next step, students were prompted to make sense of the simulation experience and demonstrate their understanding by writing an explanation of the phenomenon illustrated in the simulation. The green arrow in the diagram, which bypasses step four, represents the path through the steps for those students who did not extend their experimentation in a particular simulation task, but moved on to the next task.

Experiential learning is a cyclical process and can include multiple experiences (Kolb & Kolb, 2018). This study dives into three such experiences that required students to set up and run a wildfire simulation that illustrated certain phenomena associated with wildfire propagation. Along with each simulation experience, the tasks prompted students to make descriptive observations of surface-level simulation output as well as reflective observations that required them to consider variation in burn outcomes under specific environmental conditions. Following this, students were asked to compose a written explanation that integrated their simulation experience and observations, to reason about the environmental conditions that affect wildfires. As they worked through

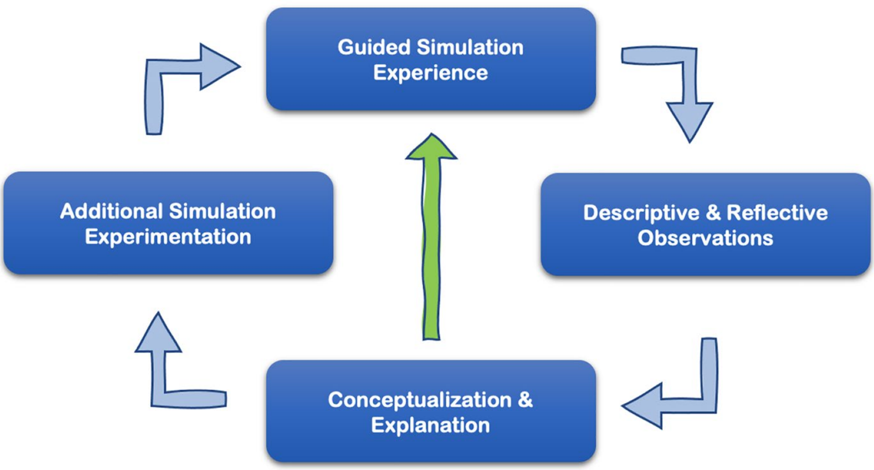

Fig. 1 Modified Experiential Learning Theory framework (Lord et al., 2024) for use with a single guided simulation task. In the additional simulation experimentation phase, students may perform new experiments to test their ideas

the first three steps of the ELT cycle, students acquired growing knowledge of wildfires and their behavior under varying environmental conditions. In the fourth step, students could choose to run additional experiments with the simulation to test and refine their ideas.

This paper focuses on two research questions that look closely at how students engage in the ELT cycle in three simulation-based tasks:

1. How are three measures of students' simulation experience (quantity, variation, and quality of simulation use) related to their answers to multiple-choice and written explanations of wildfire phenomena?
2. To what extent does the sequence of students' simulation use and question responses predict their observation and written explanation scores?

To answer the first research question, we compared three measures of students' simulation experience to their scores on the specific description and explanation questions associated with each task. The three measures are quantity, variation, and quality. *Quantity* is measured by how many times a student ran the simulation. *Variation* is measured by the number of *unique runs,* i.e., simulation runs differing in at least one input parameter. The number of runs and number of unique runs, which we refer to as "dosage measures," may provide researchers with interesting results but they do not consider how the student set up an experiment in order to answer a specific question, which is especially important when they are using an open-ended simulation (Gobert et al., 2013).

Variance in the third measure, the *quality* of the simulation experience, can only be detected using more fine-grained analytics. In a study by Lee et al. (2021), researchers parsed log files to find patterns in students' actions while using a water simulation (e.g., looking at water availability in confined vs. unconfined aquifers) and compared them to the quality of students' written arguments in the claim-evidence-reasoning format (McNeill & Krajcik, 2008). Lee's team used a decision tree to examine the combination of variables and student actions that resulted in the best simulation experience to produce the highest quality explanations. Pallant et al. (2023) examined a different artifact of students' work with a simulation–their graphical snapshots of a plate tectonics simulation–and compared those to their explanations. In that case, students whose snapshot showed the simulation was configured to produce the necessary evidence provided explanations that evinced a significantly higher level of reasoning than that demonstrated by students whose snapshots showed that they failed to set up the simulation experience correctly. Finally, Lord et al. (2024) reported that students' pre-to-post-test gains were correlated with the quality of their simulation experience; however, that study aggregated students' model use across a module with 10 simulation-based tasks and did not look closely at simulation use in the context of specific tasks.

Based on the ELT framework shown in Fig. 1, we hypothesized that students would complete the cycle of simulation experience, observation, explanation, and then, perhaps, further experimentation. To verify this pattern, this study examined students' actions as they worked through three simulation-based tasks. To answer our second research question, we analyzed the logs of students' clickstreams to determine not only their interactions with the simulation but also with the questions that followed. We used the ELT framework to design and develop the entire wildfire curriculum. Each task scaffolds students through a simulation-based experiment and then asks both observation and explanation questions.

We explored whether or not the different paths followed by students through the ELT cycle were predictive of their scores on the subsequent questions.

Research context

The wildfire simulation was embedded in a five-activity curriculum unit that gave students the ability to experiment with several factors that affect the propagation of wildfires. The first activity introduces students to the natural phenomenon of wildfires and the impacts of these fires on people in the United States over the past 100 years. In addition, students are given historical data on the frequency of such fires and the total area burned, which has trended upward over time. The second activity provides students with a unique interactive simulation tool called the Wildfire Explorer, which they use to conduct investigations on three factors that influence wildfire spread: amount of moisture, wind, and slope of terrain. In the third activity, students use the simulation to explore differences in the spread and intensity of wildfires based on various types of vegetation as fuel. Students are also introduced to the concept of fire suppression and the unintended effect of putting out fires, which allows dead, dry vegetation to build on the forest floor and create increased fuel for the next fire. In the fourth activity, students are provided with mitigation tools to help reduce the risks of wildfires to communities represented in the simulation. Finally, in the fifth activity, students are asked to synthesize their knowledge of complex factors such as the effect of climate change on wildfire spread as well as past wildfire suppression efforts, and to apply their knowledge to more complex simulation-based tasks.

The unit includes 10 simulation-based tasks. This study focuses on the first three of these, in which students are asked to run experiments that focus on a single variable. The unit includes instructions for adjusting the initial parameters of the simulation (e.g., drought level, location of sparks that start the fire, vegetation type), but students are free to run the simulation as many times as they like with whatever settings they choose. They are then asked to investigate the effect of three parameters—drought, terrain, and vegetation—on fire propagation. Each task is accompanied by two questions, a multiple-choice question that calls for a description of what occurred in the simulation and an open-response question that asks the student to explain why the simulation acts in that way.

Drought task

In this task, students investigate the effect of moisture on wildfire spread. The simulation presents two zones that can be set differently. Students can set the drought level (no drought, mild drought, and medium drought) to be different in the two zones, place a spark in each zone, and compare the spread of the fire, which they can do visually and also by using a graph generated by the model. A multiple-choice question prompts them to observe the output of the simulation and select the drought level that resulted in the fastest spread. An open-response question requires students to explain why they think the wildfire burned fastest under the drought condition they selected.

Terrain task

Students explore how wildfires spread in different terrain (see Fig. 2). They are asked to place two sparks in the mountainous zone, one at the base of a mountain slope and one at

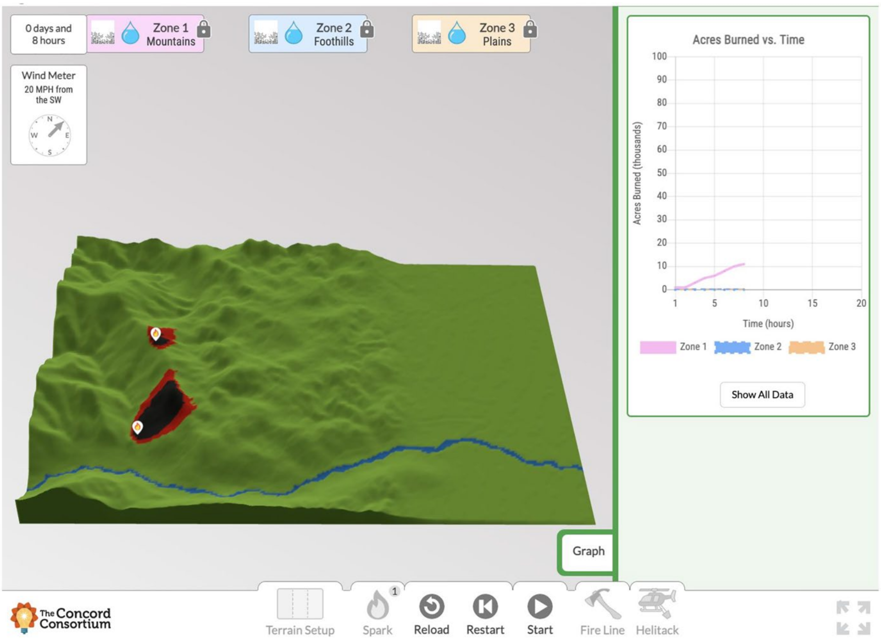

Fig. 2 The Terrain Task asks students to place sparks at the bottom and the top of a slope in a mountainous area and observe the rate at which the wildfire spreads in both scenarios

the top. Students are instructed to observe the wildfires as they burn. A multiple-choice question asks them to compare the speed in which the fire spread when started in different areas. They are then asked to explain why fire moves differently depending on where the spark was located.

Vegetation task

In this task students explore the spread of wildfires fueled by different types of vegetation: grass, shrub, and forest. They are instructed to run an experiment by changing the vegetation type in each zone, though they are not explicitly advised to hold all other parameters constant. The multiple-choice question for this task asks students to select the vegetation that caused the fire to spread the fastest, and the open-response question asks them to explain why a wildfire spreads more quickly with the type of vegetation they chose.

Methods

Subjects

This study involved 1515 students who participated in a field test of the wildfire unit and experienced at least one of the three simulation-based tasks in the spring of 2022. Among the students, 40.6% were male, 38.2% were female, and 21.2% were non-binary or preferred not to answer; 42.2% spoke English as a second language; 72.8% were white. Students were taught by 24 teachers in 15 high schools and 9 middle schools across the United States. Five of the schools were located in CA; two schools each were located in KY, PA, MI, NJ, TX, and WA; and one each was in FL, IA, MA, MD, MO, NC, and RI. The schools were located in a mix of suburban (14), rural (5), and urban (5) areas.

Data collection

As students used the online wildfire unit, their work was automatically saved. Every action that students made, including their typed answers to questions, was logged and timestamped. The Wildfire Explorer simulation was also instrumented to track students' mouse clicks. Every time a student clicked a button, altered a setting in the simulation, or placed a spark in a given location to start a wildfire, the event was saved and timestamped. Students' answers to questions and their actions within the simulation accumulated over the multiple days that they used the unit and were saved into one log file per student. These log files were then downloaded and parsed by researchers.

Simulation experience scoring

Our first research question is concerned with understanding the students' experiences using the simulation by looking at trace data. We gathered data from three tasks: Drought Task, Terrain Task, and Vegetation Task. There are multiple ways to measure student actions via their simulation clicks. For each task, we computed the total number of times the student ran the model, or the *quantity*, as well as the *variation* in runs, which was the number of unique runs (i.e., runs with different starting parameters) students made. The third measure of simulation use measured the *quality* of the experimental setup of each simulation run. The quality score was computed using distinct rubrics customized for each task (see Tables 1 and 2). The rubrics were designed to give the highest score to a student who ran the simulation in such a way that the output would show the target phenomenon. For example, in the Terrain Task, the optimal simulation run requires that students place one spark at the base of a mountain and one at the top and then run the simulation. When students added a spark to a region on the landscape, the elevation at that point was logged, enabling researchers to evaluate how well the student met the requirement to observe the fire moving differently when started at the top versus the bottom of a slope. For the Terrain Task,

Table 1 Rubric for scoring the quality of the Terrain Task simulation experience

Score	Criteria
0	The student never went to the task page
1	The student went to the task page but never ran the simulation
2	The student ran the simulation but never with two sparks in the mountain zone
3	The student ran the simulation with two sparks in the mountain zone, but those sparks were not placed with one at the bottom and one at the top of a mountain
4	The student ran the simulation with at least one spark at the bottom and one at the top of a mountain

The simulation will not run unless at least one spark is placed to start a fire

Table 2 Rubric for scoring the quality of the drought and vegetation tasks simulation experience

Score	Criteria
0	The student never went to the task page
1	The student went to the task page but never ran the simulation
2	The student ran the simulation but not with a spark in both zones
3	The student ran the simulation with sparks in both zones but both regions were the same with respect to the target variable (drought or vegetation)
4	The student ran the simulation with a spark in each zone and each zone was different with respect to the target variable but also different with respect to some other variable (not a controlled experiment)
5	The student ran the simulation with a spark in each zone and each zone was different with respect to the target variable and the same with respect to all other variables (a controlled experiment)

The simulation will not run unless at least one spark is placed to start a fire

students received a score of 0 if they never visited the page with the simulation, a 1 if they visited the page but never ran the simulation, a 2 if they ran the simulation but never with two or more sparks in the mountain zone, a 3 if they ran the model but not with at least one spark at the top of a mountain (within the highest elevation range) and one at the bottom (within the lowest elevation range), and a 4 if they ran the model with at least one spark at the top and one at the bottom.

The Drought and Vegetation Tasks were scored similarly. The Drought Task asked students to investigate the spread of fires under different drought conditions. In that case, the students set each zone to a different level of drought, added a spark to each zone, and watched the fire spread. The Vegetation Task asked students to compare the spread of fires in three different vegetations: grass, shrubs, and forest. Students' performance on each of these tasks produced simulation quality scores that ranged from 0 to 5. Students received a zero for never visiting the page and a 1 for visiting the page but not running the simulation. Students received a 2 if they ran the simulation with only one spark in one zone. Students who placed a spark in both zones but did not compare two variations of the target variable (drought or vegetation) received a 3. If the student varied the target variable as well as an additional variable, they received a 4. Finally, students received a 5 if they conducted a controlled experiment with a spark in each zone where each zone was different with respect to the target variable and the same with respect to all other variables.

Observation scoring

We scored students' responses to a multiple-choice question that asked them to examine the simulation output and make an observation. Answers to these questions were scored automatically by the system as a 0 for incorrect and a 1 for correct. For example, in the Terrain Task, students were asked the following question, "Add a spark to the top of one mountain and the bottom of a different mountain. Then, run the simulation. Did the wildfire move more quickly as it moved up or down the mountain?" Students chose from the following three responses, "The fire spread more quickly as it moved up the mountain," "The fire spread more quickly as it moved down the mountain," and "The fire spread at the same rate through the mountains."

Explanation scoring

In each task, students were asked to formulate an explanation for the phenomena they observed in the simulation. For example, in the Terrain Task, students were asked, “Why do you think the fire behaves differently if it begins at the base of the mountain versus at the top of a mountain?” Students’ responses to these open-ended questions were scored on a scale from 0 to 4, using a knowledge integration rubric, which measured the number of salient ideas that students included in their written response as well as links between these ideas (see Fig. 3). Knowledge integration rubrics are more sensitive to inquiry-based

Terrain Task Observation Prompt: Add a spark to the top of one mountain and the bottom of a different mountain. Then, run the model.
Did the wildfire move more quickly as it moved up or down the mountain?

(1) The fire spread more quickly as it moved up the mountain.
(2) The fire spread more quickly as it moved down the mountain.
(3) The fire spread at the same rate through the mountains.

Terrain Task Explanation Prompt: Why do you think the fire behaves differently if it begins at the base of the mountain versus at the top of a mountain?

Important features to include in explanation:

- (spread rate) Flames go up, moving fire uphill, wind pushes fire up
- (vegetation) Fire spreads from tree to tree/branch to branch
- (moisture level) Fire below dries out (or preheats) the trees above
- (fuel amount) More fuel/trees (at bottom compared to top), the larger/hotter fire

Important links to include in explanation:

- Spread rate/moisture level link: Drier fuel causes wildfire to spread more quickly.
- Spread rate/vegetation link: If vegetation is dry, it burns more easily.
- Spread rate/fuel amount: Fires need fuel to burn; the more fuel, the larger the fire.

Scoring Rubric

Knowledge Integration level (score)	Criteria	Example student response
Irrelevant Off-task (0)	Left the answer blank or gave an unrelated answer	I don’t know.
No-link Non-normative (1)	Gave a scientifically invalid answer or restated multiple choice	Wildfires happen randomly.
Partial-link Normative ideas (2)	Elicited one of the features above	In a fire, flames point upward.
Full-link Single link between two normative ideas (3)	Made a link with two of the features above (speed, vegetation, moisture, or fuel amount)	Flames from lower trees can reach the trees above and that makes them hotter and dries them out.
Complex link At least two links between three or more normative ideas (4)	Made two or more links with the features above	When a fire starts at the bottom of a mountain, the flames can reach all the vegetation above. If there are a lot of trees on the mountain, the fire will become huge. It will quickly move up and burn all the fuel in its path.

Fig. 3 Prompts and scoring rubric for the Terrain Task

learning questions as they can better discern the level of students' understanding of complex science concepts than more simple rubrics (Lee et al., 2011). Students who received a score of 0 or 1 wrote irrelevant or non-normative answers, such as "I have no idea," or answers not pertaining to the scientific factor being studied, such as "The landscape was plains so it spread faster and there was no vegetation." Students who made partial links to important features, such as "Because flames point uphill," received a score of 2. Students received a score of 3 when they were able to make a full link between two ideas. In one case, a student wrote, "Fire spreads up the mountain quicker than it spreads down it because when moving up the mountain, there is more vegetation right in front or above the flame and so it catches fire easily and the fire spreads." To receive a score of 4, student responses had to include a complex link or multiple links, such as "The fire at the base of the mountain is burning upward. As it's traveling upward brush is able to catch more quickly as there's more available dead vegetation. At the top of the mountain it has to travel downwards which is an unnatural direction for fire to burn as it faces upward." This response connects the speed of the fire spread to the direction of the flames and the availability of additional fuel. A score of 2 or above was considered a high score as it indicates that the student was able to at least partially explain the phenomenon. Two members of the project team scored the students' responses to the explanation questions and a third researcher resolved any discrepancies. The quadratic weighted kappa was 0.930 for the Drought Task, 0.926 for the Terrain Task, and 0.894 for the Vegetation Task.

Results

Simulation experience, observations, and explanations

We calculated the first two measures of students' simulation experience by counting the quantity and variation of runs. There was a wide range of runs for each task (see Table 3). Some students ran the model an extreme number of times–one student ran the Vegetation Task simulation 65 times with 35 different distinct experiments.

We then calculated the simulation quality scores for each student run. For the purpose of this analysis, if students ran the same simulation multiple times, we used their highest score. So, if a student experimented with the simulation, running it several times with less-than-ideal initial parameters but ran one time with the maximum score of 5 for the Drought and Vegetation Tasks or a score of 4 for the Terrain task, they received the higher score. Extra runs were not always helpful. For example, in the second task, one student ran 16 times with 13 different initial conditions. These extra experiments did not help the student, who answered the multiple-choice question incorrectly (observation score of 0), indicating that wildfire moves downhill more quickly than uphill (the opposite is true and can be seen in the simulation). This student also had a low explanation score of 1, answering simply,

Table 3 Summary of quantity of runs and variation in runs counts for each task

Task	Quantity		Variation	
	Range	Mean	Range	Mean
Drought	0–61	5.47	0–25	3.58
Terrain	0–36	3.13	0–15	2.13
Vegetation	0–64	2.43	0–31	1.76

Table 4 Contingency table for the Drought Task showing the number of students who scored in each category

		Answer type			
		Observation		Explanation	
		Incorrect	Correct	Low	High
Quality score	Low	261	347	174	434
	High	236	671	168	739
Quantity score	Low	340	650	233	757
	High	157	368	109	416
Variation score	Low	332	618	234	716
	High	165	400	108	457

Table 5 Contingency table for the Terrain Task showing the number of students who scored in each category

		Answer type			
		Observation		Explanation	
		Incorrect	Correct	Low	High
Quality score	Low	390	979	805	564
	High	35	111	70	76
Quantity score	Low	379	931	760	550
	High	46	159	115	90
Variation score	Low	368	912	740	540
	High	57	178	135	100

"the fire spreads faster down a mountain." In comparison, another student received a 3 for their explanation score, having run the model for this task just once under the nominal initial conditions. This student's explanation was, "Fire spreads up the mountain quicker than it spreads down it because when moving up the mountain, there is more vegetation right in front or above the flame and so it catches fire easily and the fire spreads."

We aggregated the open-response scores into two categories as follows: scores lower than 2 were scored as "low" and those 2 or greater were scored as "high." We then aggregated the simulation scores. The simulation quality was scored as "low" for scores less than or equal to 3 and "high" for scores 4 or higher. For the quantity and variation counts, we did something similar, scoring the number of runs as "low" for runs less than or equal to 5 and "high" for a number of runs greater than 5; unique runs were scored similarly but with a cutoff of 3. Although admittedly arbitrary, these cutoff points resulted in comparably sized cohorts in the low and high categories for each variable.

Having reduced each of our five variables to two levels, we proceeded to compute contingency tables containing counts for every pair of variables for each of the three tasks. The contingency tables for each task show the number of students who scored low on the simulation scores and also answered the multiple-choice question incorrectly on that page, the number who scored low but answered correctly, the number who scored high and answered incorrectly and the number who scored high and answered correctly (see Tables 4, 5, and 6).

We computed the probability that these contingency counts could have arisen by chance using a chi-squared distribution including the Yates continuity correction (Yates, 1934)

Table 6 Contingency table for the Vegetation Task showing the number of students who scored in each category

		Answer type			
		Observation		Explanation	
		Incorrect	Correct	Low	High
Quality score	Low	216	308	245	279
	High	99	892	378	613
Quantity score	Low	283	1098	562	819
	High	32	102	61	73
Variation score	Low	272	1077	544	805
	High	43	123	79	87

Table 7 This table shows the relationships between the three simulation measures and the observation and explanation scores for each task

Task	Simulation score	Answer type	Chi-squared	Probability
Drought	Quality	Observation	46.4	< .001*
		Explanation	20.7	< .001*
	Quantity	Observation	2.97	0.090
		Explanation	1.36	0.244
	Variation	Observation	5.05	0.016***
		Explanation	5.86	0.016***
Terrain	Quality	Observation	1.12	0.290
		Explanation	5.94	0.015***
	Quantity	Observation	3.39	0.066
		Explanation	0.194	0.659
	Variation	Observation	1.771	1.83
		Explanation	0.001	0.974
Vegetation	Quality	Observation	201	< .001*
		Explanation	10.1	.001**
	Quantity	Observation	0.658	0.417
		Explanation	0.985	0.321
	Variation	Observation	2.617	0.106
		Explanation	2.928	0.087

*$p < .001$

**$p < .005$

***$p < .05$

(see Table 7). For the Drought Task, the quality and variation scores were predictive of both the observation and explanation scores at the $p < 0.001$ level for the quality score and the $p < 0.05$ level for the variation score. The quantity score was not a significant predictor of either the observation or the explanation scores. For the Terrain Task, the only significant prediction was between the quality and explanation scores whereas for the Vegetation Task, the quality score predicted both. Notably, the simulation quality score was correlated with the explanation scores in all three tasks ($p < 0.001$ for Drought, $p < 0.05$ for Terrain, and $p < 0.005$ for Vegetation).

Simulation and question answering sequence

Our original ELT framework for simulations assumed the typical path would be simulation, description question, explanation question, and then optional additional experimentation with the simulation. However, by analyzing logs of students' actions on each task page, we uncovered 14 different sequences of actions that students took when running the simulation and answering the observation and explanation questions (see Table 8). In fact, we found that a small number of students never ran the simulation but still answered the description and/or explanation questions. Students who never visited the task page are not included in these counts. For all three tasks, the majority of students fall into two sequence categories. The first category includes those students who followed the expected path through the task in a linear fashion. These students, hereafter called the Linear Group, ran the simulation (one or more times), answered the observation question, and then completed the explanation. The second category includes those students who returned to the simulation between answering the two questions. These students, hereafter called the Return Group, ran the simulation (one or more times), answered the observation question, and then returned to run the simulation again (one or more times) before completing the explanation. In the Drought Task, 67% of students fell into these two categories, in the Terrain Task it was 76% and in the Vegetation Task it was 77%.

To determine if the sequence that students used during each task was correlated with their explanation scores on that task, we calculated probabilities by once again dividing the students into two groups by explanation score and comparing these groups by their sequence type. In only one of the three tasks, the Drought Task, did the sequence of students' actions significantly correlate with their explanation scores ($p < 0.005$). In the other two tasks, students' use of either of these two patterns did not predict their scores

Table 8 Analysis of the log files revealed 14 different path sequences that students took through each task

	Drought		Terrain		Vegetation	
	(n)	%	(n)	%	(n)	%
Observation (only)	15	1.0	6	0.4	7	0.5
Observation, explanation	65	4.5	99	7.1	93	6.8
Observation, explanation, simulation	14	1.0	14	1.0	18	1.3
Observation, simulation, explanation	69	4.8	52	3.7	88	6.4
Explanation (only)	8	0.6	7	0.5	6	0.4
Explanation, observation, simulation	4	0.3	0	0	3	0.2
Explanation, simulation, observation	48	3.3	30	2.1	23	1.7
Simulation (only)	21	1.5	9	0.6	7	0.5
Simulation, observation	159	11.1	13	0.9	15	1.1
Simulation, observation, explanation*	502	34.9	799	56.9	734	53.8
Simulation, observation, simulation, explanation**	458	31.8	268	19.1	310	22.7
Simulation, explanation	19	1.3	30	2.1	25	1.8
Simulation, explanation, observation	46	3.2	44	3.1	30	2.2
Simulation, explanation, simulation, observation	10	0.7	32	2.3	6	0.4

*Linear Group

**Return Group

Table 9 Comparison of sequence group and explanation scores for each task

Task	Sequence group	Explanation score low	Explanation score high	Chi-squared	Probability
Drought	Linear group	129	373	10.697	0.001*
	Return group	77	381		
Terrain	Linear group	476	323	1.654	0.198
	Return group	147	121		
Vegetation	Linear group	311	423	0.123	0.726
	Return group	127	183		

*$p < .005$

on the explanation questions (see Table 9). The observation question was answered prior to the bifurcation in the sequence; therefore, it was not affected by the return to the simulation and was not included in the analysis.

Discussion

Using a fine-grained analysis of students' actions in three wildfire tasks, this study yielded insights into the relationship between the simulation experience and the other steps of the ELT cycle. To answer the first research question, we investigated students' experience with the Wildfire Explorer simulation in three different tasks. The first two measures of simulation use quantified the number of total simulation runs and different simulation experiments that students conducted. While the range of these two measures was wide for each of the three tasks, the means of both of these scores decreased from the first to the third task, suggesting that students conducted less experimentation as they moved through the module. Only in the first Drought Task did a higher number of variations in runs predict both a correct score on the observation question and a higher score on the explanation question. This may be because students were excited to experiment with the different variable settings at the beginning of the module, which were not restricted to those required by the task prompt. From log data analysis, it is clear that many students spent time experimenting with different simulation inputs. For all three Wildfire Simulation tasks, running the simulation many times did not affect the students' observation or explanation scores, which is in line with the results of the McElhaney and Linn (2011) study. This suggests that designing introductory tasks to allow for open-ended experiences does not necessarily detract from student learning of the target phenomenon.

The most notable impacts to students' ability to make observations and write explanations on the three tasks came from the third measure of performance, the simulation quality score, which measures the extent to which the students ran a controlled experiment that would result in a display of the target phenomenon of the task. Previous research on the Wildfire Explorer (Lord et al., 2024) showed that students' score on this measure across 10 tasks in the whole wildfire module correlated to higher post-test scores. In this study, we focused on the simulation experience in relation to the next two steps of the ELT cycle—the ability to make descriptive and reflective observations and then to conceptualize and explain the phenomenon. In two of the three tasks, the quality score correlated with students' ability to make correct observations of the target phenomenon. And in all three tasks, the quality score predicted the explanation score. In each of the three tasks, our analysis shows that when students have set up a simulation in a way that is likely to illustrate the desired phenomenon, they are better able to make sense of what they see and consequently write a higher-level explanation.

This finding has implications for the design of future simulation-based tasks and the potential use of real-time simulation scoring and feedback. Extra experience with the simulation, whether it is testing multiple variables or running multiple times, is less impactful than students' quality use of the simulation to make sense of the science content. When students conduct an appropriately controlled experiment using the simulation, they are able to experience the target phenomena, make observations of model output, and write thoughtful explanations. The results of this study suggest that simulation-based learning could be further enhanced via the development of an intelligent tutoring system built to analyze students' simulation actions and provide feedback in real-time. This approach could significantly improve students' understanding of the factors that contribute to wildfires. Given the growing risk of forest fires due to climate change, simulations with real-time feedback have the potential to reach students in related fields like forestry, agriculture, and environmental sciences, as well as the general public (Holzinger et al., 2024).

Previous work on simulation feedback has also relied on hand scoring written responses (Lee et al., 2021), which is labor intensive. Future versions of the wildfire module could score simulation runs in real time and give students feedback on their experimental setup prior to their answering the explanation questions, avoiding the time and expense of incorporating natural language processing tools to evaluate students' responses, which require large amounts of data for training the model. The teacher could be alerted when students are not using the model to run controlled experiments. The data analysis from this study could be used to train an AI system that evaluates students' simulation experience.

Our second research question looked more closely at the sequence in which students experimented with the simulation and answered the question prompts. Counter to our original hypothesis, which allowed for two paths through the ELT cycle, we found that students moved through the ELT cycle in 14 different ways. No student, in any of the three tasks, went through the four steps of the cycle as shown in the original modified ELT framework for simulations. The two most common paths—the Linear Group and the Return Group in our analysis—forced us to rethink our modified ELT cycle. While the Linear Group comprised the most students for each task and fits the prior version of the ELT cycle (e.g., completing the first three steps in order and omitting the fourth step), the prevalence of the Return Group prompted us to revise the ELT cycle for simulations. We collapsed the initial simulation experience and additional experimentation into one step as students frequently ran multiple and varied experiments prior to moving to the observation step. While not every permutation was included in this simplified model, the ELT cycle (see Fig. 4) shows that (1) the simulation experience includes not only the experiment to fulfill the task requirements but also experimentation and (2) the return to the simulation prior to answering the explanation question.

Again, the Drought Task stood out as the only task with a significant difference in the explanation scores between the Linear Group and Return Group where returning to the simulation prior to answering the explanation question was correlated with a higher score. This may be because students were more engaged in this first task and were excited to use

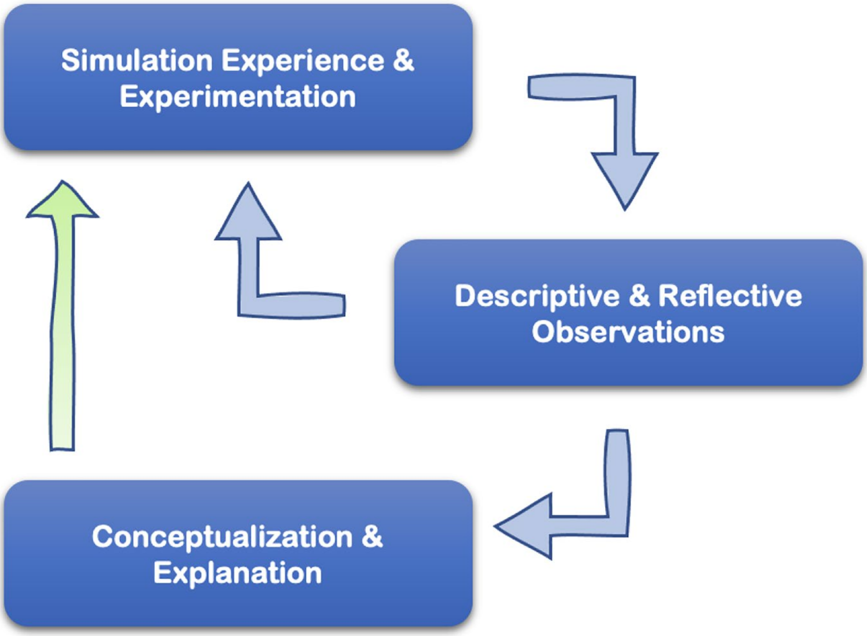

Fig. 4 Modified ELT framework for simulation-based tasks. The blue boxes and arrows show the two most common paths through a single task. The initial simulation experience and further experimentation have been combined into one step. The green arrow represents the flow to the next task; each time the student is presented with a new task, this cycle is repeated

the simulation. In providing an open-end simulation environment that enables students to explore by manipulating variables and testing their mental models, we gave them an opportunity to employ a variety of problem-solving strategies, none of which proved to be significantly correlated to higher performance on the task (Greiff et al., 2018). While students may have strayed from the scaffolded task prior to writing their explanations, allowing students agency in how they worked through the task did not deter from student learning; indeed, in some cases, it enhanced their conceptual understanding.

Limitations

The implications of this study are limited. While the wildfire module includes 10 simulation-based tasks, we focused on just three of the tasks. Students' responses to the remaining tasks' explanation scores were not evaluated nor were their paths through the ELT cycle on the next seven tasks. Future analysis could be performed on the remaining tasks. The results could also be related to students' pre-to-post-test gains. In addition, while this study focused on wildfire education, the same ELT framework could be used to study other natural hazard online resources that use computational models, simulations, or game-based curricula.

Conclusion

This study illustrates the importance of designing open-ended simulation-based learning tasks to support students' ability to turn experiences into explanations. We show that the most important measure of the simulation experience is setting up and running a controlled experiment that creates the phenomenon under investigation so that students can interpret the system output, make observations, and write high-quality explanations. However, there is diversity in the approach that students take when going through a simulation-based task. The paths that students take to come to their explanations are sometimes winding. Simulations do not need to be locked down to allow access only to the variable in question because access to other variables can encourage further experimentation and, in some cases, result in deeper learning.

Acknowledgements This material is based upon work supported by the National Science Foundation under Grant No. 1812362.

Author contributions All authors contributed to the study conception and design. Material preparation, data collection, and analysis were performed by Trudi Lord, Paul Horwitz, Amy Pallant, and Christopher Lore. The first draft of the manuscript was written by Trudi Lord and all authors commented on versions of the manuscript. All authors read and approved the final manuscript.

Funding This material is based upon work supported by the National Science Foundation under Grant No. 1812362.

Data availability The data that support the findings of this study are available from the corresponding author upon request. Materials are available at https://learn.concord.org/geo-wildfire.

Declarations

Competing interest The authors have no conflict of interest to declare.

Ethical approval The research done in this study was approved by Ethical & Independent Review Services (approval number 18176-05) and with the 1964 Helsinki Declaration.

Informed consent Informed consent was obtained from all individual participants included in the study.

Research involving human participants and/or animals The research done in this study was approved by Ethical & Independent Review Services (approval number 18176-05) and with the 1964 Helsinki Declaration.

Consent to participate Informed consent was obtained from all individual participants included in the study.

Consent to publish Informed consent was obtained from all individual participants included in the study.

References

Abdullah, S., & Shariff, A. (2008). The effects of inquiry-based computer simulation with cooperative learning on scientific thinking and conceptual understanding of gas laws. *Eurasia Journal of Mathematics, Science and Technology Education, 4*(4), 387–398.

Alderson, D. L., Darken, R. P., Eisenberg, D. A., & Seager, T. P. (2022). Surprise is inevitable: How do we train and prepare to make our critical infrastructure more resilient? *International Journal of Disaster Risk Reduction, 72*, 102800. https://doi.org/10.1016/j.ijdrr.2022.102800

Alifu, H., Hirabayashi, Y., Imada, Y., & Shiogama, H. (2022). Enhancement of river flooding due to global warming. *Scientific Reports, 12*(1), 20687.

Baker, R. S., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining, 1*(1), 3–17.
Ballard, H. L., Dixon, C. G. H., & Harris, E. M. (2016). Youth-focused citizen science: Examining the role of environmental science learning and agency for conservation. *Biological Conservation, 208*, 65–75. https://doi.org/10.1016/j.biocon.2016.05.024
Buckley, B. C., Gobert, J. D., & Horwitz, P. (2006, June). Using log files to track students' model-based inquiry. In *Proceedings of the 7th international conference on learning sciences* (pp. 57–63).
Center, A. D. R. (2015). Sendai framework for disaster risk reduction 2015–2030. *United Nations Office for Disaster Risk Reduction: Geneva, Switzerland.*
Dede, C., Clarke, J., Ketelhut, D. J., Nelson, B., & Bowman, C. (2005, April). Students' motivation and learning of science in a multi-user virtual environment. In *American educational research association conference, Montreal, Canada* (pp. 1–8).
Enns, C. Z. (1993). Integrating separate and connected knowing: The experiential learning model. *Teaching of Psychology, 20*(1), 7–13. https://doi.org/10.1207/s15328023top2001_2
Falloon, G. (2019). Using simulations to teach young students science concepts: An experiential learning theoretical analysis. *Computers & Education, 135*, 138–159. https://doi.org/10.1016/j.compedu.2019.03.001
Feurtzeig, W., & Roberts, N. (1999). *Modeling and simulations in science and mathematics education*. Springer.
Garner, A. J. (2023). Observed increases in North Atlantic tropical cyclone peak intensification rates. *Scientific Reports, 13*(1), 16299.
Gobert, J. D., Pedro, M. A. S., & Montalvo, O. (2012). Leveraging educational data mining for real-time performance assessment of scientific inquiry skills within microworlds. *4*.
Gobert, J. D., Sao Pedro, M., Raziuddin, J., & Baker, R. S. (2013). From log files to assessment metrics: Measuring students' science inquiry skills using educational data mining. *Journal of the Learning Sciences, 22*(4), 521–563.
Greiff, S., Molnár, G., Martin, R., Zimmermann, J., & Csapó, B. (2018). Students' exploration strategies in computer-simulated complex problem environments: A latent class approach. *Computers & Education, 126*, 248–263.
Henderson, N., Kumaran, V., Min, W., Mott, B., Wu, Z., Boulden, D., Lord, T., Reichsman, F., Dorsey, C., Wiebe, E., & Lester, J. (2020). Enhancing student competency models for game-based learning with a hybrid stealth assessment framework. *International Educational Data Mining Society (EDM 2020)* (pp. 92-103).
Hoell, A., Quan, X.-W., Hoerling, M., Diaz, H. F., Fu, R., He, C., Lisonbee, J. R., Mankin, J. S., Seager, R., Sheffield, A., Simpson, I. R., & Wahl, E. R. (2022). Water year 2021 compound precipitation and temperature extremes in California and Nevada. *Bulletin of the American Meteorological Society, 103*(12), E2905–E2911.
Holzinger, A., Schweier, J., Gollob, C., Nothdurft, A., Hasenauer, H., Kirisits, T., Häggström, C., Visser, R., Cavalli, R., Spinelli, R., & Stampfer, K. (2024). From industry 5.0 to forestry 5.0: Bridging the gap with human-centered artificial intelligence. *Current Forestry Reports, 10*(4), 442–455. https://doi.org/10.1007/s40725-024-00231-7
Hoover, K., & Hanson, L. A. (2023). Wildfire statistics (IF10244; In Focus, p. 3). Congressional Research Service. Retrieved from https://sgp.fas.org/crs/misc/IF10244.pdf
Horwitz, P. (1996). Linking models to data: Hypermodels for science education. *The High School Journal, 79*(2), 148–156.
Horwitz, P. (1999). Designing computer models that teach. In W. Feurzeig & N. Roberts (Eds.), *Modeling and simulation in science and mathematics education* (pp. 179–196). Springer.
Horwitz, P., & Christie, M. A. (2000). Hypermodels: Embedding curriculum and assessment in computer-basedmanipulatives. In International Conference on Mathematics/Science Education andTechnology (pp. 207–213). Association for the Advancement of Computing in Education (AACE).
Horwitz, P., Reichsman, F., Lord, T., Dorsey, C., Wiebe, E., & Lester, J. (2023). If we build it, will they learn?An analysis of students' understanding in an interactive game during and after a researchproject. *Technology, Knowledge and Learning, 28*(4), 1825–1839.
Hyogo Framework for Action 2005–2015: Building the resilience of nations and communities to disasters—full text | United Nations Office for Disaster Risk Reduction. (2007, Nov 15). Retrieved from http://www.undrr.org/publication/hyogo-framework-action-2005-2015-building-resilience-nations-and-communities-disasters
Kolb, A., & Kolb, D. (2018). Eight important things to know about the experiential learning cycle. *Australian Educational Leader, 40*(3), 8–14.

Kolb, D. A. (1984). *Experiential learning: Experience as a source of learning and development*. Prentice-Hall.

Kukkonen, J. E., Kärkkäinen, S., Dillon, P., & Keinonen, T. (2014). The effects of scaffolded simulation-based inquiry learning on fifth-graders' representations of the greenhouse effect. *International Journal of Science Education, 36*(3), 406–424. https://doi.org/10.1080/09500693.2013.782452

Lee, H. S., Gweon, G. H., Lord, T., Paessel, N., Pallant, A., & Pryputniewicz, S. (2021). Machine learning-enabled automated feedback: Supporting students' revision of scientific arguments based on data drawn from simulation. *Journal of Science Education and Technology, 30*(2), 168–192.

Lee, H.-S., Liu, O. L., & Linn, M. C. (2011). Validating measurement of knowledge integration in science using multiple-choice and explanation items. *Applied Measurement in Education, 24*(2), 115–136.

Luo, J., Makwana, A. P., Liao, D., & Kincaid, J. P. (2008). Hurricane!—A simulation-based program for science education. 2008 Winter Simulation Conference (pp. 2543–2548). IEEE. https://doi.org/10.1109/WSC.2008.4736365

Lore, C., Lee, H.-S., Pallant, A., Connor, C., & Chao, J. (2023). Integrating computational thinking into-geoscientific inquiry about volcanic eruption hazards and risks. *International Journal of Scienceand Mathematics Education*. https://doi.org/10.1007/s10763-023-10426-2

Lore, C., & Seevers, S. (2022). Using block coding to model gps motion, land deformation, andearthquake risk. *The Earth Scientist, 39*(1), 11–15.

Lord, T., Horwitz, P., Lee, H. S., Pallant, A., & Lore, C. (2024). Using simulations to support students' conceptual development related to wildfire hazards and risks from an experiential learning perspective. *Journal of Science Education and Technology,* 1–15.

Mani, L., Cole, P. D., & Stewart, I. (2016). Using video games for volcanic hazard education and communication: An assessment of the method and preliminary results. *Natural Hazards and Earth System Sciences, 16*(7), 1673–1689. https://doi.org/10.5194/nhess-16-1673-2016

Martin, T., & Sherin, B. (2013). Learning analytics and computational techniques for detecting and evaluating patterns in learning: An introduction to the special issue. *Journal of the Learning Sciences, 22*(4), 511–520.

McCaffrey, S. (2015). Community wildfire preparedness: A global state-of-the-knowledge summary of social science research. *Current Forestry Reports, 1*(2), 81–90. https://doi.org/10.1007/s40725-015-0015-7

McElhaney, K. W., & Linn, M. C. (2011). Investigations of a complex, realistic task: Intentional, unsystematic, and exhaustive experimenters. *Journal of Research in Science Teaching, 48*(7), 745–770. https://doi.org/10.1002/tea.20423

McNeill, K. L., & Krajcik, J. (2008). Scientific explanations: Characterizing and evaluating the effects of teachers' instructional practices on student learning. *Journal of Research in Science Teaching: THe Official Journal of the National Association for Research in Science Teaching, 45*(1), 53–78.

Melia, N., Dean, S. M., Pearce, H. G., Harrington, L. J., Frame, D. J., & Strand, T. (2022). Aotearoa New Zealand's 21st-century wildfire climate. *Earth's Future*. https://doi.org/10.1029/2022ef002853

Monroe, M. C., Ballard, H. L., Oxarart, A., Sturtevant, V. E., Jakes, P. J., & Evans, E. R. (2016). Agencies, educators, communities and wildfire: Partnerships to enhance environmental education for youth. *Environmental Education Research, 22*(8), 1098–1114. https://doi.org/10.1080/13504622.2015.1057555

Moutinho, S., Moura, R., & Vasconcelos, C. (2017). Contributions of model-based learning to the restructuring of graduation students' mental models on natural hazards. *Eurasia Journal of Mathematics, Science and Technology Education, 13*(7), 3043–3068. https://doi.org/10.12973/eurasia.2017.00704a

Pallant, A., & Lee, H.-S. (2015). Constructing scientific arguments using evidence from dynamiccomputational climate models. *Journal of Science Education and Technology, 24*(2), 378–395.

Pallant, A., & Lee, H-S. (2017). Teaching sustainability through systems dynamics: Exploring stocks and-flows embedded in dynamic computer models of an agricultural land management system. *Journal of Geoscience Education, 157*, 146–157. https://doi.org/10.5408/16-169.1

Pallant, A., Pryputniewicz, S., & Lee, H-S. (2023). Developing geo-sequential reasoning about tecton-icprocesses using computational simulations. *International Journal of Science Education*. https://doi.org/10.1080/09500693.2023.2217471

Petersen, G. B., Klingenberg, S., Mayer, R. E., & Makransky, G. (2020). The virtual field trip: Investigating how to optimize immersive virtual learning in climate change education. *British Journal of Educational Technology, 51*(6), 2099–2115. https://doi.org/10.1111/bjet.12991

Petersson, E., Andersson, A. L., & Säljö, R. (2013). Exploring nature through virtual experimentation. *Nordic Journal of Digital Literacy, 8*(3), 139–155.

Plass, J. L., Homer, B. D., Hayward, E. O., Frye, J., Huang, T.-T., Biles, M., Stein, M., & Perlin, K. (2012). The effect of learning mechanics design on learning outcomes in a computer-based geometry game. In S. Göbel, W. Müller, B. Urban, & J. Wiemeyer (Eds.), *E-Learning and games for training, education, health and sports* (pp. 65–71). Springer.

Quellmalz, E. S., Timms, M. J., Silberglitt, M. D., & Buckley, B. C. (2012). Science assessments for all: Integrating science simulations into balanced state science assessment systems. *Journal of Research in Science Teaching, 49*(3), 363–393.

Ramasundaram, V., Grunwald, S., Mangeot, A., Comerford, N. B., & Bliss, C. M. (2005). Development of an environmental virtual field laboratory. *Computers & Education, 45*(1), 21–34. https://doi.org/10.1016/j.compedu.2004.03.002

Ren, D., Fu, R., Leslie, L. M., & Dickinson, R. E. (2011). Modeling the mudslide aftermath of the 2007 Southern California Wildfires. *Natural Hazards, 57*, 327–343.

Sarabando, C., Cravino, J. P., & Soares, A. A. (2014). Contribution of a computer simulation to students' learning of the physics concepts of weight and mass. *Procedia Technology, 13*, 112–121. https://doi.org/10.1016/j.protcy.2014.02.015

Scalise, K., & Clarke-Midura, J. (2018). The many faces of scientific inquiry: Effectively measuring what students do and not only what they say. *Journal of Research in Science Teaching, 55*(10), 1469–1496.

Scalise, K., Timms, M., Moorjani, A., Clark, L., Holtermann, K., & Irvin, P. S. (2011). Student learning in science simulations: Design features that promote learning gains. *Journal of Research in Science Teaching, 48*(9), 1050–1078. https://doi.org/10.1002/tea.20437

Taillandier, F., & Adam, C. (2018). Games ready to use: A serious game for teaching natural risk management. *Simulation & Gaming, 49*(4), 441–470. https://doi.org/10.1177/1046878118770217

Tiwari, B., Ajmera, B., Gonzalez, A., & Sonbol, H. (2020). Impact of wildfire on triggering mudslides—A case study of 2018 Montecito debris flows. *Geo-congress 2020* (pp. 40–49). American Society of Civil Engineers.

Turco, M., Abatzoglou, J. T., Herrera, S., Zhuang, Y., Jerez, S., Lucas, D. D., AghaKouchak, A., & Cvijanovic, I. (2023). Anthropogenic climate change impacts exacerbate summer forest fires in California. *Proceedings of the National Academy of Sciences, 120*(25), e2213815120. https://doi.org/10.1073/pnas.2213815120

Warneke, C., Schwarz, J. P., Dibb, J., Kalashnikova, O., Frost, G., Al-Saad, J., Brown, S. S., Brewer, W. A., Soja, A., Seidel, F. C., Washenfelder, R. A., Wiggins, E. B., Moore, R. H., Anderson, B. E., Jordan, C., Yacovitch, T. I., Herndon, S. C., Liu, S., & Crawford, J. (2023). Fire influence on regional to global environments and air quality (FIREX-AQ). *Journal of Geophysical Research: Atmospheres, 128*(2), e2022JD037758. https://doi.org/10.1029/2022JD037758

Widiyatmoko, A. (2018). The effectiveness of simulation in science learning on conceptual understanding: A literature review. *Journal of International Development and Cooperation, 24*(1), 35–43.

Wilson, C. D., Reichsman, F., Mutch-Jones, K., Gardner, A., Marchi, L., Kowalski, S., Lord, T., & Dorsey, C. (2018). Teacher implementation and the impact of game-based science curriculum materials. *Journal of Science Education and Technology, 27*(4), 285–305. https://doi.org/10.1007/s10956-017-9724-y

Windschitl, M., & Andre, T. (1998). Using computer simulations to enhance conceptual change: The roles of constructivist instruction and student epistemological beliefs. *Journal of Research in Science Teaching, 35*(2), 145–160.

Yates, F. (1934). Contingency tables involving small numbers and the χ 2 test. *Supplement to the Journal of the Royal Statistical Society, 1*(2), 217–235.

Trudi Lord is a senior project manager at The Concord Consortium. Her research focuses on teaching and learning science with educational technology as well as user interface design and trace data analysis.

Paul Horwitz is a senior scientist emeritus at The Concord Consortium. While a physicist by training, his educational research spans over 30 years in multiple fields including genetics, evolution, mathematics, and data analytics.

Amy Pallant is a senior research scientist at The Concord Consortium. She's contributed to research across Earth and environmental science topics, with a particular interest in geology and systems modeling.

Christopher Lore is a research associate at The Concord Consortium interested in geology and environmental science. His recent work includes research on computation thinking and multiple representations.