



Two comparative studies of computer simulations and experiments as learning tools in school and out-of-school education

Johanna T. Krüger¹ · Tim N. Höffler¹ · Martin Wahl² · Katrin Knickmeier¹ · Ilka Parchmann¹

Received: 24 September 2020 / Accepted: 12 November 2021 / Published online: 9 March 2022
© The Author(s) 2022

Abstract

Interactive computer simulations and hands-on experiments are important teaching methods in modern science education. Especially for the communication of complex current topics with social relevance (socioscientific issues), suitable methods in science education are of great importance. However, previous studies could not sufficiently clarify the educational advantages and disadvantages of both methods and often lack adequate comparability. This paper presents two studies of direct comparisons of hands-on experiments and interactive computer simulations as learning tools in science education for secondary school students in two different learning locations (Study I: school; Study II: student laboratory). Using a simple experimental research design with type of learning location as between-subjects factor ($N_{Study\ I}=443$, $N_{Study\ II}=367$), these studies compare working on computer simulations versus experiments in terms of knowledge achievement, development of situational interest and cognitive load. Independent of the learning location, the results showed higher learning success for students working on computer simulations than while working on experiments, despite higher cognitive load. However, working on experiments promoted situational interest more than computer simulations (especially the episodic and value-related component). We stated that simulations might be particularly suitable for teaching complex topics. The findings reviewed in this paper moreover imply that working with one method may complement and supplement the weaknesses of the other. We conclude that the most effective way to communicate complex current research topics might be a combination of both methods. These conclusions derive a contribution to successful modern science education in school and out-of-school learning contexts.

Keywords Science education · Hands-on experiments · Computer simulation · Learning · Interest

✉ Johanna T. Krüger
jkrueger@leibniz.de

¹ Leibniz Institute for Science and Mathematics Education, Olshausenstr. 62, 24118 Kiel, Germany

² GEOMAR Helmholtz Centre for Ocean Research Kiel, Wischhofstr. 1-3, 24148 Kiel, Germany

Introduction

Many highly relevant but difficult to communicate topics with social relevance (socio-scientific issues) exist in science education. It is a great challenge for learners to grasp these complex processes and concepts, especially when several aspects require simultaneous consideration, for example regarding the intricate topic of climate change. Therefore, context-based approaches are often used in science education to promote more positive attitudes toward science and to provide a solid basis for scientific understanding of such complex topics (Bennett & Jennings, 2011). Considering contexts and applications of science as starting points for the development of scientific ideas (context-based teaching approaches) requires appropriate methods. For instance, interactive computer-based methods might help promote the acquisition of knowledge and understanding in such contexts in a comprehensive manner. Interactive computer simulations (hereafter only referred to as computer simulation or simulation) are becoming an increasingly important element of science education not only as an assessment tool (Bennett, Persky, Weiss, & Jenkins, 2010), but also for learning in subjects such as chemistry, biology or physics (Quellmalz et al., 2008; Rutten, van Joolingen, & van der Veen, 2012). Compared to simple animations, computer simulations allow learners to actively interact with the simulated scenario by changing given parameters according to their own ideas and subsequently receive direct feedback from the system. The underlying mathematical model determines how the computer simulation reacts to the parameter changes of the learner, which shows the influence of a certain aspect on the process and the outcomes (Develaki, 2017; Lin et al., 2015). The fact that computer simulations represent a model of a system (natural or artificial) or a process with all its determining parameters (Jong & van Joolingen, 1998) enables the learner to safely experiment and simulate in an artificial learning environment (Lin et al., 2015). Indications already suggest that they are particularly suitable for communicating complex issues (Smetana & Bell, 2012) and are able to improve overall science skills (Siahaan et al., 2017). Unfortunately, computer simulations as learning tools often lead to high cognitive load (Paas et al., 2003a) and in some cases do not produce the expected learning success (Jong, 2010; Köck, 2018). The increased time required for the introduction of such a method, the lack of know-how as well as financial and material resources of the educational institution are factors limiting the use of computer simulations in school and out-of-school education (Hanekamp, 2014). Furthermore, appropriate simulations for various topics are not always available or accessible at the educational institutions.

Instead, so-called hands-on experiments are often used in science education as methods for teaching scientific processes and concepts (Di Fuccia, Witteck, Markic, & Eilks, 2012; National Research Council (NRC), 2012; National Science Teachers Association (NSTA), 2013). Students work on a scientific question alone or in small groups developing their own hypotheses, testing these with experiments and interpreting the results in relation to their hypotheses. The experiments used in our investigations are theory-based and model experiments. This means an experiment with a material model that differs from the represented original in at least one characteristic for reasons of availability, complexity, hazard potential, accessibility, or cost. An example for a model experiment is the representation of ocean acidification processes with candles (=CO₂ production), aquarium (=atmosphere), ocean water (experimental variable) and distilled (control variable). In particular, basic processes and concepts can vividly be conveyed with laboratory experiments (Köck, 2018) as they usually represent the effects of one influencing variable on one single aspect (e.g. the influence of increased atmospheric CO₂ concentration on the pH value of seawater).

However, they fall short in conveying complex concepts, as it is usually not possible to represent several influencing variables in a comprehensible way through model experiments (Braund & Reiss, 2006). In addition, since they often represent past knowledge processes, they often do not adequately reflect current questions and investigations.

Shouldn't educators, therefore, use simulations instead of traditional hands-on experiments especially when communicating complex and current topics like socioscientific issues? Which advantages and disadvantages do the two methods offer in comparison to each other? This paper presents two studies of direct comparisons of hands-on experiments and interactive computer simulations for learning complex marine ecology issues. We investigated the effects of the methods on knowledge gain, cognitive load, and the development of situational interest of secondary school students. For this purpose, we next describe the known educational effects of the two methods regarding content knowledge acquisition, cognitive load, and interest in the following.

Content Knowledge

Findings regarding the influence of hands-on experiments or simulations on students' learning outcomes are inconclusive so far. On the one hand, empirical studies indicate that computer simulations foster students' involvement in the observation and investigation of phenomena, which supported students' conceptual change in science (Rutten et al., 2012; Trundle & Bell, 2010). For instance Park (2019) found that after working on a computer simulation on physical concepts, the students predicted and explained the given scientific phenomena with more valid scientific ideas. Especially for the communication of complex processes several studies point to the potentially highly effective suitability of simulations (Sarabando, Cravino, & Soares, 2016; Smetana & Bell, 2012). Quellmalz et al. (2008) summarize that "Simulations are well-suited to investigations of interactions among multiple variables in models of complex systems (e.g., ecosystems, weather systems, wave interactions) and to experiments with dynamic interactions exploring spatial and causal relationships" (Quellmalz et al., 2008, p. 193). Experiments are also considered very important in science education in terms of knowledge acquisition (Hofstein & Lunetta, 2004). Hands-on activities improve students' academic achievement and understanding of scientific concepts through the manipulation of real objects, as abstract knowledge can be communicated more concretely and clearly (Ekwueme et al., 2015). Working on hands-on activities increases the students' ability to interpret data and think critically (Tunnicliffe, 2017). According to Hattie (2012) a medium effect for learning with experiments could be found ($d=0.42$).

On the other hand, empirical studies report several cognitive and metacognitive difficulties for students learning with computer simulations (Jong, 2010; Jong & van Joolingen, 1998; Köck, 2018; Mayer, 2004). This is mostly due to the high cognitive load that results from working on these complex systems (Jong, 2010). In particular a lack of interactivity can lead to learning only rudimentary content (Linn, Chang, Chiu, Zhang, & McElhaney, 2010). In addition, the meta-study of Hattie (2012) showed only low effects regarding the influence of computer simulations and simulation games on the learning success ($d=0.33$). Likewise, not all authors agree with the positive influence of experiments on learning outcomes. Students often have major problems working systematically and strategically (Jong & van Joolingen, 1998). Additionally, school experiments often present reconstructions and simplifications of past knowledge processes and might not adequately reflect today's issues

and investigations (Braund & Reiss, 2006). Thus, realistic conceptions about research often cannot be developed.

Cognitive load

Cognitive load theory (see van Merriënboer and Sweller (2005) for an overview) claims that successful learning can only occur when the (limited) cognitive capacities of learners' working memory system are not overburdened. A cognitive overload is likely to occur especially in cases of very difficult tasks (high intrinsic cognitive load) in combination with inappropriate and overly complex instructional designs (high extraneous cognitive load), which results in low learning outcomes (Mayer et al., 2005). The cognitive effort required when working in multimedia learning environments (like a computer simulation) is usually high (Anmarkrud et al., 2019; Mutlu-Bayraktar, Cosgun, & Altan, 2019). In a simulation, students must consider several aspects simultaneously. This requires learners to understand the nonlinear structure of the learning environment, which "costs" some cognitive resources (high intrinsic cognitive load). Understanding the complex underlying systems and processes of a computer simulation is often a great challenge especially for learners with low prior knowledge (Jong, 2010; Jong & van Joolingen, 1998). What is more, by encouraging learners to manipulate too many variables, simulations might generate split-attention effects (having to keep in mind many elements or having to observe several changes on different places on the screen) (Kalyuga, 2007). Especially simulations without guidance/instructional support tend to cause learners' disorientation and thus lead to extraneous cognitive load. These high cognitive demands of simulations often lead to a cognitive overload (Paas et al., 2003a) which results in a more difficult understanding and low learning outcomes (Mayer et al., 2005). Therefore several studies have shown the effectiveness of instructional support for learning with computer simulations to prevent cognitive overload (Eckhardt et al., 2018; Jong & van Joolingen, 1998; Rutten et al., 2012). With regard to performing hands-on activities, previous studies (e.g., Scharfenberg et al., 2007; Winberg & Berg, 2007) indicated that students might also experience increased extraneous load during. This is thought to be caused by students' hands-on tasks, such as reading instructions, handling equipment, and interacting with peers within the workgroup. In some cases, this extraneous load seemed to be so high that it had a negative impact on learning outcomes.

Interest

In educational research two types of interest can be differentiated situational and individual (Krapp et al., 1992). Individual interest is a relatively permanent predisposition attending objects, events or ideas and deals with certain contents (Renninger, 2000). Situational interest is mainly caused by situation-specific environmental stimulations like novel or conspicuous activities (Lin et al., 2013). Both types of interest are characterized by three aspects: an emotional, a value-related and a cognitive component (e.g. Engeln, 2004; Krapp, 2007). In our study, the cognitive component of situational interest is mainly concerned with a person's need to increase competence, knowledge and skills in relation to the object of interest. Therefore, we will call this aspect epistemic component in the following. Extensive research exists on changing motivation and interest by using computer simulations as learning tools. Hence, in several studies increased interest (e.g. Jain & Getis, 2003) as well as the motivating effect of computer simulations (Rutten et al., 2012) were observed. Both

a high level of interactivity and the possibility of exercising control over the learning environment can increase motivation for the subject matter. Attractive learning environments, such as out-of school student laboratories in particular, are said to have a motivating effect on students (Glowinski & Bayrhuber, 2011). With less attention than in research on simulations, an increased interest in processing experiments was also observed (Palmer, 2009).

Previous comparative studies

Despite a few studies already directly comparing digital and hands-on methods, a clear trend with regard to educational advantages and disadvantages cannot yet be determined. Some studies show advantages of experiments regarding learning outcomes (Kiroğlu et al., 2019; Marshall & Young, 2006), while other studies found advantages regarding learning effects in digital simulations (Finkelstein et al., 2005; Lichti & Roth, 2018; Paul & John, 2020; Scheuring & Roth, 2017). Berger (2018) for example, observed a higher motivation of the students when working on the computer-based physics experiment compared to a hands-on experiment.

Other studies, in turn, could not find a clear trend of cognitive benefits in either direction (Chini, Madsen, Gire, Rebello, & Puntambekar, 2012; Evangelou & Kotsis, 2019; Lamb et al., 2018; Madden et al., 2020; Stull & Hegarty, 2016). For example, Zandler and Greiner (2020) showed no general difference, in a direct comparison of experiments and simulations in chemistry education, in students' learning outcomes on reactions of metals. However, indications exist that students can learn different aspects better through the different methods. For example, a recent study by (Puntambekar et al., 2020) showed that students who carried out physical labs engaged in discussions related to setting up equipment, making measurements and calculating results. In contrast, students who conducted virtual labs engaged in more discussions related to making predictions, understanding relationships between variables, as well as interpreting scientific phenomena.

The present study

This paper presents findings of two studies offering direct comparisons of hands-on experiments and interactive computer simulations as learning tools for students from Grade 10 to Grade 13. The two studies investigated the effects of working on these learning methods focusing on content knowledge achievement, cognitive load and changes in situational interest. Due to the fact that the investigations were identical in terms of design and procedure, they show a high degree of comparability. The first study took place in school and the second study in a student laboratory to confirm the independence of the results from the learning location. A student lab can also be described as a science center outreach lab. There, students gain an insight into current science by working on experiments in inquiry-based learning environments (Itzek-Greulich et al., 2015). Climatic and anthropogenic changes in marine ecosystems served as subject matter.

Why another comparative study?

Most of the existing comparative studies in this research field show the difference between teaching with simulations vs. teaching without simulations (e.g., as presented in the detailed review by Rutten et al., 2012). Thus, the many different factors of instruction, such

as reading textbooks, completing tasks, classroom discussions, or different group sizes, can hinder a valid measurement of the effects of the methods. This leads to a lack of comparability. Our study focused on the direct comparison of the methods, independent of the lessons' context. The external factors are therefore stable: same introduction, structure of the instruction unit and same topics, only the learning modalities were different. This allows us to make conclusions about the effects of the methods itself. Other studies examine the comparison between working in virtual vs. physical labs (Darrah et al., 2014; Puntambekar et al., 2020). These studies often compare at the level of the entire learning environment (e.g., different laboratories) rather than at the level of the individual modality (experiments or simulation). Furthermore, some studies exist with a direct comparison of the two methods primarily in the fields of physics or chemistry education, focusing on the comparison of virtual vs. analogous experiments of low complexity (e.g., Evangelou & Kotsis, 2019; Madden et al., 2020). Our study presents a direct comparison of differently performed experiments with the same content and high complexity. Compared to the existing comparative studies, our studies include a large sample size ($N_{Study I}=443$; $N_{Study II}=367$; $N_{total}=810$). For example, the 60 studies presented in the detailed review by Rutten et al. (2012) had an average sample size of 160 participants. Finally, current findings on learning with interactive computer simulations and hands-on experiments do not yet provide sufficient insights into cognitive and motivational processes of learning on socio-scientific issues with high complexity. The studies presented here address this desideratum and aim to close this gap.

Research Questions and Hypotheses

The research questions and corresponding hypotheses of this project are presented in the following.

- (a) To what extent do interactive computer simulations / hands-on experiments contribute to an increased content knowledge? How do both methods differ?

We expected that both methods would improve students' content knowledge, since the topic was new to most of them and both experiments (Ekwueme et al., 2015; Tunnicliffe, 2017) and simulations (Rutten et al., 2012; Smetana & Bell, 2012) have been shown to promote learning. If we manage to keep the effects of cognitive load low by providing appropriate instructional support (Mayer et al., 2005), we expect computer simulations to achieve the same or even better learning success for students (Evangelou & Kotsis, 2019; Lichti & Roth, 2018; Madden et al., 2020).

- (b) To what extent do interactive computer simulations / hands-on experiments affect cognitive load while completing the tasks? How do both methods differ?

We expected a medium students' cognitive load for both methods. Both methods had instructional support, which has been proven to reduce learners' extraneous cognitive load (Eckhardt et al., 2018). We tried to adapt the level of content knowledge to the participants' skills in order to reduce intrinsic load. However, we expected a low level of prior knowledge of the students, since the topic addressed was not a mandatory topic in the German curriculum, which could have a negative impact on the intrinsic cognitive load (Jong, 2010; Jong & van Joolingen, 1998). Furthermore, we expected that the participants had different cognitive skills, since they came from different schools and grades and would also not be used to work on computer-based learning environments

since this educational method is still not common in German science classes. This can lead to an overburdening with the complex structures, especially for the simulation (Jong, 2010). Therefore, we expected higher cognitive load when working with the computer simulations compared to working with the experiments.

- (c) To what extent do interactive computer simulations / hands-on experiments increase situational interest of the students? How do both methods differ?

We expected an at least medium level of situational interest for both methods since both provide evidence of a general motivating character (Palmer, 2009; Rutten et al., 2012). Through the work on computer simulations we expected a higher increased situational interest because of the motivation character of a digital medium. These hypotheses applied to all subscales of the situational interest (emotional, epistemic, value-related). With regard to the comparison of the different learning locations, a higher interest can be expected through the work in the student laboratory (Glowinski & Bayrhuber, 2011).

Methods and design

The two investigations (Study I and Study II) presented here were similar in structure, overall topic, objectives, and evaluation design but differed in duration, learning location and complexity of the topic (for a comparison of the similarities and differences between the two studies, see Table 1). In each case, scientists of marine ecology, media psychology and education developed the experiments and simulations together using co-design. The methods were based on real scientific data. The experiments were hypothesis-driven model experiments, and the simulations were interactive and instructionally supported. We conducted both studies using a simple experimental research design (Fig. 1) with type of

Table 1 Comparison of the two presented studies

	Study I	Study II
Location	School	Student laboratory
Topic	Ocean acidification and impacts on calcifying organisms (one-dimensional)	Warming, eutrophication, salinity and ocean acidification of the Baltic Sea and impacts on three organisms, their interactions and challenges on the ecosystem/social level (multi-dimensional)
Duration	90 min	180 min
Design	1. Pre-test (20 min) 2. Introduction (20 min) 3. Practical work: exp./sim. (30 min) 4. Post-test (20 min)	1. Pre-test (15 min) 2. Introduction (30 min) 3. Practical work: exp./sim. (120 min) 4. Post-test (15 min)
Participants	$N=443$ ($N_{exp}=221$, $N_{sim}=222$) Grade: 10–13 Average age: 17.58 years	$N=367$ ($N_{exp}=198$, $N_{sim}=169$) Grade: 10–13 Average age: 17.02 years
Instruments	Same measurements for both studies: Personal data (age, gender, marks in science subjects) Individual interest in biology & chemistry (adapted from Frey et al., 2009) Prior knowledge (self-developed instrument) Content knowledge (self-developed instrument) Cognitive load (Paas et al., 2003a, 2003b) Situational interest (adapted from Engeln, 2004)	

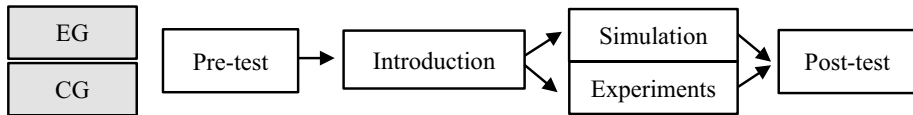


Fig. 1 Study design

learning location as between-subjects factor. Thereby the investigation of the effects of the simulation represented the experimental group (EG) and the investigation of the effects of the experiments represented the control group (CG).

However, when we speak of "experiment" in the following, we always mean the experiment as a learning method, which we have used in school or in the student laboratory. We used a pre-test to assess students' personal data as well as their prior knowledge, marks in science subjects and interest in biology and chemistry (independent variables). The post-test analyzed the effectiveness of simulation-based learning compared to experiment-based learning regarding content knowledge, situational interest, and cognitive load (dependent variables).

Design of Study I

The first study was a 90-min intervention at school. The purpose of this intervention was to convey the process and the effects of ocean acidification on a global and local level. In an initial introduction, students received general information related to the testing and the course of the day. Subsequently, a 20-min paper–pencil-test took place. The introductory lecture provided an overview of the topic of ocean acidification and highlighted resulting problems for marine ecosystems. Afterwards, participants were randomly assigned into two groups. We separated the groups into different rooms to avoid any interference while conducting the experiment or the simulation, respectively. Students had to work on their assigned method for 30 min. The experiments were conducted in small groups with a supervisor who guided the experiment and was available for questions. The supervisor handed out the scripts and, if necessary, organized the structural procedure (for example, he explained the materials, he provided assistance in answering the questions in the script, and he paid attention to the avoidance of typical mistakes in performing the experiment). For the simulation, each student used their own laptop, but they worked in small groups or teams. A supervisor was available for technical as well as content related questions. After handing out the scripts, he showed the students the most important functions of the simulation (operation of the necessary buttons), he helped to answer the questions in the script and explained, if necessary, incomprehensible contexts. The same supervisors were present during the entire test period. Finally, we administered a 20-min post-test.

Materials of Study I

Both methods conveyed the same content: pH-value, reasons for and extent of the increase in global atmospheric carbon dioxide (CO₂) concentration, equilibrium reaction of CO₂ in seawater and the effect of acidification on calcifying organisms. In both treatments, students used the same research questions to investigate the process and effects of ocean acidification. The activities were supported by a script, in which the results of the simulation

or experimental work had to be summarized by leading questions. The paper scripts were handed out at the beginning of the method phase and had to be completed during the work. The questions were based on the tasks of the methods and were the same for both methods. An example of this is: "Explain why calcification is more difficult with increased CO₂ concentration in the atmosphere. Take into account the previously established reaction equations for the equilibrium system." The supervisor assisted the students both in performing the experiments or working through the simulation and in answering the questions.

Hands-on experiment

First, students investigated the extent of the increasing CO₂ concentration in the atmosphere and the characteristics of the pH-value in the script. This included analyzing a Keeling Curve and matching different liquids to the corresponding pH-value. Then they conducted an experiment that mainly focused on the consequences of an increased atmospheric CO₂ content for seawater with special regard on the buffering effect. An experimental setup was hence chosen that represents the ocean (Baltic Sea water), the combustion of fossil fuels (candles), and the atmosphere (upside down aquarium). Distilled water was used as a control. In both water samples, a digital pH meter measured the changes of the pH-value. After 10 min of testing, the students observed a considerably faster decrease in pH-value in the distilled water than in the water of the Baltic Sea. The students could take the set-up and procedure of the experiment from the script and had to implement this as autonomously as possible. After documenting the results, the students set up the equilibrium reaction equation of CO₂ in seawater and thus addressed the buffering effect of seawater as an explanation for the different changes of the pH-value in the two water samples. Support cards were available to them for this purpose. Finally, they analyzed the resulting effects on calcifying organisms such as mussels, which were present as an illustrative object and discussed possible protective measures. The experiment used here is part of the BIOACID-project (Biological Impacts of Ocean Acidification) and was developed by scientists researching ocean acidification as part of their public relations work (BIOACID, 2012).

Interactive computer simulation

The simulation is a learning environment created at Stanford University, for which permission had been granted for use on this research project. As an introduction, the digital learning tool illustrated the reasons for and the extent of the increasing atmospheric CO₂ concentration by an animated Keeling Curve. Then the computer simulation offered the possibility to interactively examine the characteristics of the pH-value by matching the acidity of different liquids on a pH scale. In the main section, students were able to control the increase in atmospheric CO₂ concentration and observe the resulting effects on an animated equilibrium reaction of seawater. For this purpose, the students simulated the CO₂ concentration of the atmosphere under three different scenarios using a slider from the year 1865 to 2090. In real time they could observe changes in the schematically presented equilibrium reaction as well as changes in the exact concentrations of carbon dioxide (CO₂), hydrogen carbonate (HCO₃⁻), and carbonate (CO₃²⁻) (see Fig. 2). Supporting elements were implemented such as toolboxes which, when clicked on, explain how to read the Keeling curve or explanations that appear when swiping over various reaction arrows. In the next part, learners had to interactively differentiate

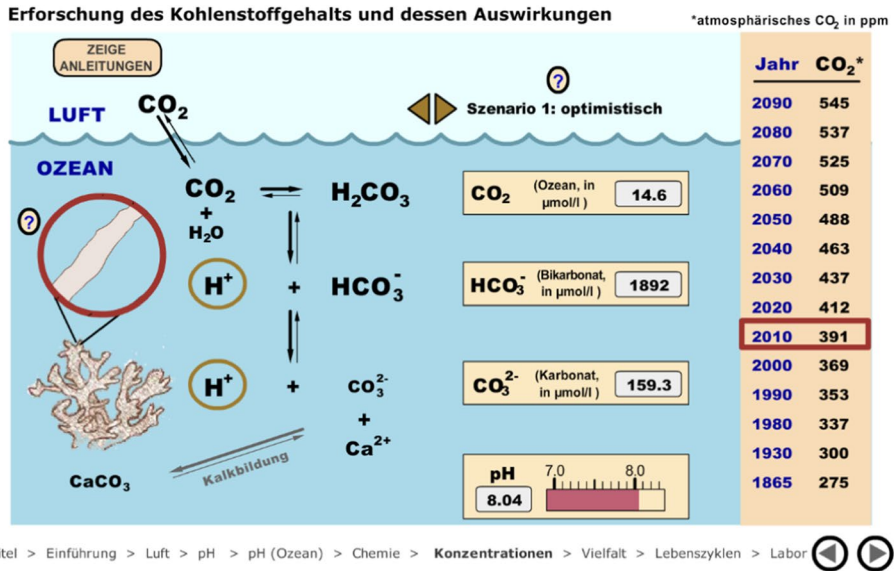


Fig. 2 Screenshot of the simulation from Study I (<http://i2sea.stanford.edu/AcidOcean/AcidOcean.htm>)

between calcifying and non-calcifying organisms. Finally, they discussed in the group the negative consequences of acidification on the calcification process and developed protective measures.

Design of Study II

The second study was part of a full-day student lab day. The study presented here consisted of a 180-min intervention for students of grades 10–13 on the topic "Future of the Baltic Sea". The purpose of this intervention was to convey the processes of major environmental impacts of the Baltic Sea (warming, eutrophication, acidification and salinity changes) and its effects on representative organisms of the ecosystem (gammarids, bladderwrack, and epiphytes) as well as the resulting challenges for the whole system (water quality, fishing, tourism). The 15-min pre-test took place after receiving general information about the testing and the course of the day. The students carried out all testing using a digital questionnaire on tablet computers. Subsequently, a 30-min introductory lecture provided information on the causes and processes of global changes in the oceans and gave an insight into the resulting effects on marine ecosystems. Afterwards we randomly assigned the participants into two groups and separated them into different rooms. The experiment group was further divided into two subgroups, each conducting four experiments of 30 min. A supervisor guided the experiment and was available for questions during the procedure for each group. He handed out the scripts, presented the available materials and supported the students to answer the questions in the script and if necessary explained unclear issues. The simulation group was also divided into two subgroups, each working at a group table. Although students had their own tablet, they worked in teams or together as a group. A supervisor was available for technical

as well as content related questions and he supported the group discussions. The learning period consisted of two hours for all groups and was followed by the 15-min digital post-test.

Materials of Study II

Both methods conveyed the same content. The experiments each conveyed a single process of change (warming, acidification, eutrophication, or salinity changes) with the effects on one organism (gammarids, bladderwrack, or epiphytes) and the resulting ecosystem and societal changes (water quality, fishing, or tourism). Thus, the students learned about the different changes and their effects as well as their interactions one after the other. In the simulation, it was possible to simulate all changes simultaneously and to observe the effects on all three organisms together, taking into account their interactions. Furthermore, the adjusted parameters represented directly possible effects on the entire system (water quality, fishing, and tourism). To avoid disorientation and cognitive overload especially by working on the simulation, students had to work on a supporting script in which they had to summarize the results of their work by the same guiding questions. An example of this is: "Describe the effect of elevated water temperature on bladderwrack fitness". The script for the experiments was in paper form and contained the respective instructions, explanations and questions for each experiment. The script for the simulation was digital and embedded in the simulation website. It contained instructions for the operation of the simulation as well as basic information about all parameters as well as the questions.

Hands-on experiments

The students investigated the processes and effects of (a) increased water temperature on the fitness of bladderwrack, (b) over-fertilization on the growth rate of epiphytes, (c) changed salinity on the fitness of gammarids, and (d) ecosystem changes caused by increasing acidification. The activities in the four experiments were different, but the procedure was always the same. First, the students informed themselves about the parameters of the station (one change and one organism each) through available materials such as texts or illustrations (e.g., salinity and gammarids). They then derived the research question (e.g., "What is the effect of lower Baltic Sea salinity on the population of gammarids?"). With the explanation in the script, they conducted the various experiments and described the explanations for the observed effects in the script. Examples of the activities were (a) preparing three bladderwrack samples of the same size in different water temperatures and observing the oxygen production via a digital oxygen sensor, (b) photometrical examining differently eutrophied water samples using a chlorophyll-a measurement, (c) counting gammarids and determining their mortality rate in different salinities of Baltic Sea water, or (d) performing a titration of differently acidified seawater samples. Finally, the students discussed in the group what impact the effects could have on other organisms in the system and what measures could help to protect them (e.g., the reduced feeding pressure of the smaller gammarid population on the epiphytes and the resulting negative effects on the bladderwrack, as well as the reduction of the anthropogenic greenhouse effect as a measure).

Interactive computer simulation

The overall goal of the computer simulation was to enable learners to understand the effects of the combination of different ecosystem changes on all three organisms in interaction. Before the students could interact with the simulation, they listened to short informative audio files on each variable of the simulation for basic information. Afterwards the interactive mode was unlocked and the students could then interact with the simulation in a self-determined manner. The interface allowed users to select and manipulate the different changes by moving a slider (see Fig. 3). The students could observe shifts in the population sizes of the three organisms in real time. Of course, students could (and should) study only one change at a time — just as in the experiments — by moving one slider on predefined settings. The difference to the experiments was that the effects on the organisms were presented in interaction. This means that indirect effects are also visible (for example, a lower salinity leads directly to a lower population of gammarids, which in turn leads to an increase in epiphytes and thus has a negative effect on the development of bladderwrack). Further effects on water quality, for example, can be observed simultaneously with a three-level smiley system at the bottom of the screen. For more information about these effects and the other parameters of the simulation, we implemented supporting elements such as toolboxes. Students first had to learn about the individual parameters of the simulation (changes and organisms) guided by the script (texts, illustrations and videos were available for this purpose). Then they had to change each individual impact on the Baltic Sea (warm-

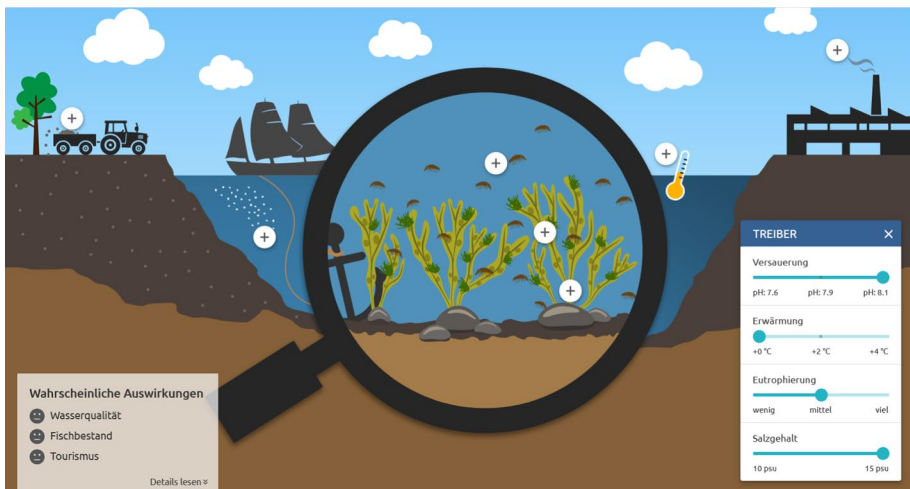


Fig. 3 Screenshot of the simulation from Study II (URL: <https://ostsee-der-zukunft.experience-science.de/simulation.html>)

ing, eutrophication, acidification and salinity changes) guided by research questions and observe, describe and explain the effects on the organisms. Then the students had to change several parameters at the same time and investigate and describe the effects on the level of organisms and further impacts. At the end, they had to discuss and describe possible measures to protect the Baltic Sea (Fig. 3).

Samples of Study I + II

The participants in Study I were 443 students from 19 German secondary schools with an average age of 17.58 years ($SD = 1.41$), 56.88% students were female and 43.12% students were male. The random dividing of the classes into the methods resulted in a distribution of 221 students who worked on the experiments and 222 students who worked on the simulation.

The participants in Study II were 367 students from 21 German schools. The average age was 17.02 years ($SD = 1.20$), 54% of the students were female and 46% were male. The experiments were conducted by 198 students and the simulation by 169 students.

Measures of Study I + II

In the pre-test we assessed sociodemographic data like age, gender and grade for a better impression of the sample. In addition, we asked for the students' marks in science subjects (biology, chemistry, physics) as well as prior knowledge and individual interest in biology and chemistry to better understand the pre-existing cognitive and motivational competencies of the participants. In the post-test, we applied questionnaires on content knowledge, cognitive load and situational interest. The questions for Study I were presented as paper-pencil-test; the questionnaires of the second study were digital and integrated into the learning environment of the simulation.

Content knowledge

We developed different questionnaires for pre- and post-testing to investigate content knowledge. Since the topics of both studies were not part of the German standard curriculum, we expected only little prior knowledge. Therefore, an increase in content knowledge through the interventions was highly expected. We administered only a short pre-test, to avoid frustration as well as a testing effect, i.e., a learning effect because of answering the same questions twice. The content knowledge pre-test is only about verifying possible differences in prior knowledge between the later randomly divided test groups. We developed a detailed post-test because we focused on the difference between the two methods. Several experts in the field developed the tasks in accordance to the topics of the simulation and experiments.

In the first study, the pre-test consisted of three multiple-choice items on basic principles of ocean acidification (Appendix, Table 8). Two items had four different answering options and were scored with one point; one item had five different answering options and was scored with two points. Students could achieve four points in total. The post-test consisted of nine questions. The structure of the questions was based on the content areas of the methods and thus reflect the topics covered in the experiments and simulation: pH-value (3 questions), reasons for and extent of the increase in global atmospheric carbon dioxide (CO_2) concentration (2 questions), equilibrium reaction of CO_2 in seawater (3 questions) and the effect of acidification on calcifying organisms (1 question). See Appendix, Table 9 for example items for the different content knowledge areas. Five multiple-choice items with four different answering options were developed. There was one correct answer for each question, which was scored with one point. There were also three multiple-choice questions with five answering options and two correct answer possibilities. These items were scored with two points.

Questions were only asked about the topics covered in both methods. Questions about the practical implementation (e.g., set-up of the experiment) were explicitly not asked in order to really measure the learning gains through the methods in a comparable way. The multiple-choice items were developed by experts and subsequently validated. In addition, we developed a task where students should mark the increase or decrease of the oceans' concentrations of carbon dioxide, hydrogen carbonate and carbonate in a table with an arrow. Three points could be achieved for this. The participants were able to score up to 15 points in total. Both tests were developed on the basis of existing knowledge tests in chemistry (Höft et al., 2019).

The knowledge pre-test in the second study included three open-ended questions that tested the general state of knowledge about changes and challenges of the Baltic Sea ecosystem (Appendix, Table 8). On the one hand, we decided to ask which threats of the Baltic Sea the students are generally aware of (Question 1). Second, we asked about two concepts that are most likely to be addressed in science education as benchmarks for knowledge about ecological changes such as climate change and anthropogenic impacts (Question 2 and Question 3). For the scoring of the open questions we worked out an expectation framework according to which 6 possible points could be given for each question. For the question about the threats to the Baltic Sea, three overarching categories (climate change, overuse, pollution) could be found with two sub-points each. For questions 2 and 3, students were able to score two points each for causes, process, and effects of the change. Thus, students could score a total of 18 points. The post-test included 15 multiple-choice questions and three open-ended questions. The multiple-choice questions were also structured according to the methods' content: treated organisms and their interactions (4 questions), processes of the changes (4 questions), effects of a single change (4 questions) and effects of combined changes (3 questions). See Appendix, Table 10 for example items of the different content knowledge areas. For the last three open questions, students had to (a) describe the effects of all four predicted changes in combination for the year 2100 at the organism level, (b) the resulting impacts on the ecosystem as well as (c) on the societal level. The multiple-choice questions provided four different answer options with one correct answer each, which were developed and validated by several experts of the field. For the three open-ended questions, we developed a horizon of expectations with three possible points for each question: (a) more epiphytes, less bladderwrack, fewer amphipods; (b) impacts on biodiversity, water quality, coastal protection; (c) threats to tourism, fisheries, health. Each multiple-choice item was scored with one point; the open questions were scored with three points. In total, the participants could score up to 24 points.

Cognitive load

We decided to measure cognitive load considering that working on interactive computer simulations and conducting hands-on experiments demand a certain amount of learners' cognitive resources. Thus, we implemented the cognitive load test by Paas et al. (1994) in the post-test phase as student self-report (Table 2). Therefore, students had to evaluate their perceived cognitive load within two questions on a 7-point rating scale (1 = made no effort at all; 7 = made a real effort). The items were provided by Paas et al. We chose this short scale because it is the most widely used measure

Table 2 Summary of reliabilities (Cronbach's α) of used scales in the pilot study ($N=44$); SI = Situational interest

Scale	Cronbach's α	No. of items	Example item
Interest in biology	0.931	5	I am interested in learning new things in biology.
Interest in chemistry	0.920	5	I am interested in learning new things in chemistry.
Cognitive load	0.797	2	How difficult was the experiment/simulation to understand? When working on the experiment/simulation, my mental effort was...
SI emotional	0.749	4	I felt comfortable while working on the experiment/simulation.
SI epistemic	0.519	5	I would like to learn more about the things we worked on in the simulation/experiment.
SI value-related	0.392	3	It is important to me personally that we did the experiments/simulations today.

of cognitive load due to its reliability and sensitivity and, furthermore, its ease of use (Paas et al., 2003b). Furthermore, since we did not focus on the distinction between intrinsic and extrinsic load in our study, this time-saving two-item scale was particularly suitable.

Interest

We examined interest in biology and chemistry to control differences in individual interest as an influence variable on situational interest. These two disciplines were chosen because they represent the fundamental areas of the topics addressed in both studies. For this purpose, we adapted the test “pleasure and interest in science” for general interest in science to interest in biology and chemistry. The test was originally developed and validated for the PISA study (Frey et al., 2009). The students had to assess five statements each on their individual interest in chemistry and biology using a 4-point rating scale (1 = completely disagree; 4 = completely agree).

Participants’ situational interest (SI) in the simulation and experiments was assessed via the questionnaire developed and validated by Engeln (2004). The test used 12 items to measure the emotional, epistemic and value-related components of situational interest. Participants could indicate their answers on a 4-point rating scale (1 = completely disagree; 4 = completely agree). See Table 2 for example items.

Verifying the instruments’ quality

We piloted the instruments for interest and cognitive load with a group of $N=44$ students from three different classes of Grade 13 (Table 2).

The analysis of the reliability of the value-related situational interest resulted in a Cronbach’s alpha value of .510 if one of the items was not considered. We attributed the low reliability of this component to linguistic inconsistency and improved this in the following. All other scales showed sufficiently high reliabilities and could be used without modifications for the main studies. We developed both content knowledge tests in collaboration with teachers and marine biology scientists to ensure a high content validity. The content knowledge questionnaires were validated by the same students of the pilot study for Study I and by a group of different 23 students from Grade 12 for Study II. The students conducted the interventions as planned, but answered the questions in an open plenary session with us. This collaborative validation enabled us to eliminate

Table 3 Summary of reliabilities (Cronbach’s α) of used scales for Study I+II

Scale	Cronbach’s α	
	Study I	Study I
Interest in biology	0.917	0.900
Interest in chemistry	0.919	0.891
Cognitive load	0.595	0.586
Situational interest emotional	0.825	0.801
Situational interest epistemic	0.621	0.748
Situational interest value-related	0.555	0.763

Table 4 Means (standard deviations) of independent variables of both groups and corresponding values of the *t*-tests for Study I

Pre-test Study I	Mean Value Experiments (SD)	Mean Value Simulation (SD)	<i>df</i>	<i>T</i>	<i>p</i>
Prior knowledge	1.30 (1.04)	1.29 (1.06)	440	0.059	0.95
Marks in science subjects	2.59 (0.85)	2.50 (0.83)	424	1.041	0.30
Interest in biology	2.89 (0.70)	2.92 (0.74)	438	- 0.434	0.66
Interest in chemistry	2.25 (0.80)	2.37 (0.77)	438	- 1.601	0.11

Table 5 Means (standard deviations) of independent variables of both groups and corresponding values of the *t*-tests for Study II

Pre-test Study II	Mean Value Experiments (SD)	Mean Value Simulation (SD)	<i>df</i>	<i>T</i>	<i>p</i>
Prior knowledge	3.16 (2.44)	3.63 (2.47)	363	1.807	0.07
Marks in science subjects	2.95 (1.10)	3.00 (1.04)	361	0.527	0.60
Interest in biology	2.96 (0.66)	2.93 (0.64)	362	0.435	0.66
Interest in chemistry	2.47 (0.67)	2.47 (0.70)	364	- 0.027	0.98

Table 6 Means (standard deviations) of dependent variables and corresponding values of the *t*-tests and effect sizes (Cohen's *d*) for Study I, all *p*-values are adjusted for alpha-inflation using Benjamini & Hochberg (FDR) correction

Post-test Study I	Mean Value Experiments (SD)	Mean Value Simulation (SD)	<i>df</i>	<i>T</i>	<i>p</i>	<i>d</i>
Content knowledge	7.90 (3.06)	8.51 (3.09)	441	- 2.08	0.048	- 0.19
Cognitive load	2.96 (0.97)	3.17 (1.08)	441	2.15	0.048	0.19
Situational interest emotional	3.04 (0.83)	3.07 (0.64)	441	0.34	0.735	-
Situational interest epistemic	2.38 (0.57)	2.19 (0.56)	441	- 3.43	0.003	- 0.33
Situational interest value-related	2.7 (0.57)	2.49 (0.59)	438	- 3.68	< 0.001	- 0.35

ambiguities in formulation and specialist terminology. After the adjustments of the pilot studies, we could determine the following internal consistency for the instruments used in Study I and Study II (Table 3).

Results

All results presented here were calculated with the statistics program SPSS. We used *t*-tests to examine mean value comparisons. The *t*-test indicated no significant differences in prior knowledge, average marks in the science subjects (biology, chemistry, physics), nor in interest in biology or chemistry between the later random assignment to the groups

Table 7 Means (standard deviations) of dependent variables and corresponding values of the *t*-tests and effect sizes (Cohen’s *d*) for Study II all *p*-values are adjusted for alpha-inflation using Benjamini & Hochberg (FDR) correction

Post-test Study II	Mean Value Experiments (SD)	Mean Value Simulation (SD)	<i>df</i>	<i>T</i>	<i>p</i>	<i>d</i>
Content knowledge	8.58 (3.81)	13.33 (4.49)	365	- 10.95	<.0001	- 1.05
Cognitive load	2.75 (0.95)	2.99 (1.04)	365	2.28	0.029	0.21
Situational interest emotional	2.41 (0.62)	2.33 (0.57)	364	1.28	0.203	-
Situational interest epistemic	3.22 (0.55)	2.9 (0.61)	365	- 5.18	<.0001	- 0.54
Situational interest value-related	2.85 (0.67)	2.62 (0.59)	364	- 3.5	0.002	- 0.37

(experiments or simulation) of both studies (Tables 4 and 5). We checked these possible differences in the pre-test questionnaires to exclude an effect on differences of the dependent variables in the later randomly divided groups. The analysis of the prior knowledge questionnaire showed that students achieved an average of 1.3 points (*SD*=1.05) out of possible 4 points in the first study. In the second study students achieved an average of 3.38 points (*SD*=2.46) out of 18 possible points for the three open questions in the pre-test.

The *t*-tests of the dependent variables of both studies confirmed that on the one hand, learning with the simulation compared to learning with experiments resulted in higher students’ content knowledge achievement ($d_{Study I} = -0.19$; $d_{Study II} = -1.05$) and a higher

Fig. 4 Differences in content knowledge achievement in % between the methods for both studies (Study I: Intervention at school $N_{Exp} = 221, N_{Sim} = 222$; Study II: Out-of-school intervention $N_{Exp} = 198, N_{Sim} = 169$); * $p < .05$, *** $p < .001$

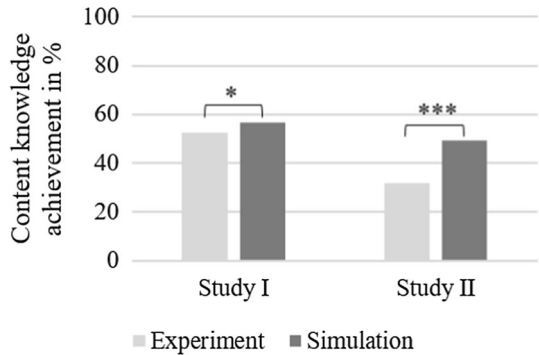


Fig. 5 Differences in mean cognitive load between the methods of both studies (Study I: Intervention at school $N_{Exp} = 221, N_{Sim} = 222$; Study II: out-of-school intervention $N_{Exp} = 198, N_{Sim} = 169$); * $p < .05$

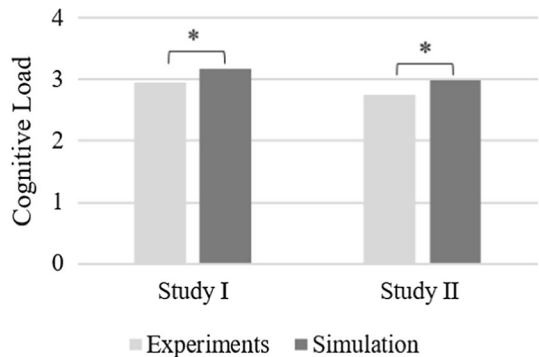


Fig. 6 Differences in mean situational interest (SI) between the methods of Study I ($N_{Exp}=221$, $N_{Sim}=222$), subdivided into different components; ** $p < .01$, *** $p < .001$

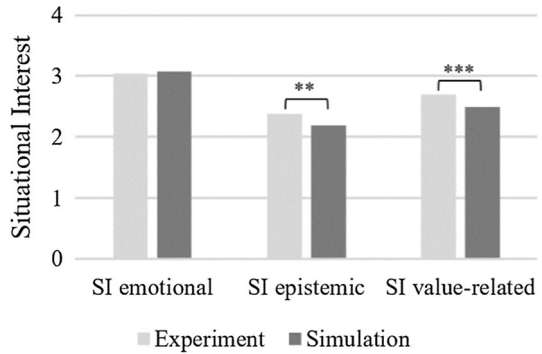
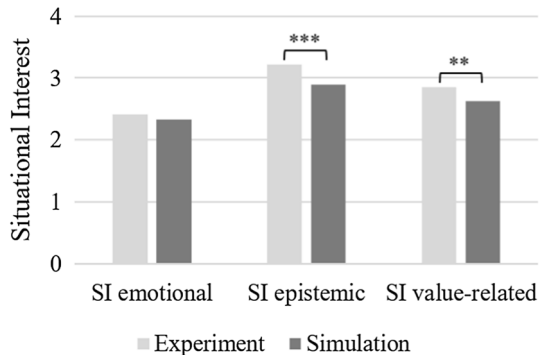


Fig. 7 Differences in mean situational interest (SI) between the methods of Study II ($N_{Exp}=198$, $N_{Sim}=169$), subdivided into different components; ** $p < .01$, *** $p < .001$



level of cognitive load ($d_{Study I}=0.19$; $d_{Study II}=0.21$). On the other hand, students who worked with the experiments showed a higher level of the epistemic ($d_{Study I}=-0.33$; $d_{Study II}=-0.54$) and value-related ($d_{Study I}=-0.35$; $d_{Study II}=-0.37$) component of the situational interest by than by working with the simulation. We could not observe a difference in the emotional component of situational interest between the two methods for both studies. All means, standard deviations and statistic values are shown in Tables 6 and 7 and presented in Figs. 4, 5, 6 and 7. We could not find any significant differences in subgroups (age, gender, school form, or grade).

Discussion

The goal of this research was to investigate the differences between learning with computer simulations and experiments concerning (a) the contribution of content knowledge, (b) the extent of affected cognitive load, and (c) the extent of situational interest to find out which method is better suited to communicate current and complex topics. Therefore, we conducted two studies using a simple experimental research design with type of learning location as between-subjects factor. The following sections show detailed explanations of the assumptions for each research question and resulting implications.

Content knowledge

The low achievements in the pre-test indicate that the prior knowledge was very low in both studies, as expected. This is probably because the topics did not belong to the standard German curriculum. The expectedly non-significant differences of prior knowledge as well as marks in science subjects between the two randomly assigned treatment groups suggest that the results for the differences in content knowledge achievement cannot be attributed to differences in prior knowledge or school success in science subjects.

As expected (Evangelou & Kotsis, 2019; Madden et al., 2020), students had no disadvantage learning with the simulation compared to the experiments; in fact, they learned more despite higher cognitive load. Besides the different modality itself (simulation vs. experiments), there are several indications on why the simulation in this case facilitated learning the interaction of multiple variables: We suggest, that the focus of the students' tasks in the two methods could have an impact on the content learned (Puntambekar et al., 2020). Since the students did not have to deal with physical activities during working on the simulation, but had more time to investigate processes and contexts, it stands to reason that they were also able to learn this content better. Additionally, since everyone had their own device, students could deal with the single processes individually to a greater or lesser extent. Thus, they were able to determine at their own learning pace from when they started to deal with the interactions. They also had more time to do so, since physical activities were absent. In the experiments, time distribution was quite strictly predetermined by the experimental procedure. The fact that the students had their own device but still worked in groups may allow another advantage for learning with simulations compared to experiments: Students had the opportunity to first manipulate each change or perform the tasks themselves and then discuss things as a group. More division of labor was required in conducting the experiments, and therefore not every student performed all the tasks themselves. This high level of interactivity and the possibility of self-determined action within the simulation are, as already shown, of great importance for learning challenging content (Linn et al., 2010). However, since learning which is solely oriented towards self-determined learning of scientific principles and concepts often leads to less effective learning success (Jong & van Joolingen, 1998; Mayer, 2004), the students in our studies worked in groups or with a partner. This might have helped to bypass the low level of knowledge acquisition often found in exploratory learning studies using computer simulations (Jong, 2010).

The topics dealt with in our studies both conveyed not only rudimentary content but also complex processes and structures especially in the second study. This means that the interventions conveyed actual and difficult to communicate topics with social relevance (socio-scientific issues). Several aspects require simultaneous consideration, to understand the effects and impacts of ocean acidification, warming, eutrophication, or lower salinity on an ecosystem and society. The second study showed an even higher degree of complexity than the first, as all processes were considered together, whereas in the first study "only" one process was considered. The great effect of the learning advantage of simulations compared to the experiments in the second study ($d=1.05$) indicates that interactive computer simulations are particularly suitable for teaching complex processes and systems, as already assumed by other authors (Smetana & Bell, 2012). Scheuring and Roth (2017) and Lichti and Roth (2018), respectively, were able to show that computer simulations primarily promote complex thinking processes. This could be due to the fact that multiple variables can be manipulated simultaneously during simulation. Simulations therefore seem

to be particularly suitable for understanding the interaction of multiple variables. Thus, if educators with the goal of imparting (complex) knowledge—rather than teaching scientific methods—are looking for a suitable learning method, simulations could be beneficial here.

However, the appropriate instructional support in the simulations plays a significant role here. Results from research literature show that especially for students with little prior knowledge—such as those in our studies—instructional support offers effective assistance for successful knowledge acquisition (Eckhardt et al., 2018; Jong & van Joolingen, 1998; Mayer, 2004; Rutten et al., 2012). We claim that these assumptions about the importance of instructional support should be considered since increasing digitalization is currently creating more and more digital and interactive learning environments. Future work in this area could include providing additional support for teachers and students working with simulations. For example, guidance on the use of appropriate instructional support, on supporting students in interacting with simulations through dialog systems, or on artificial agents would be useful (e.g. (Wallace et al., 2009; Zapata-Rivera et al., 2014).

Cognitive load

The expected higher cognitive load when working with computer simulations can be explained at the content, task, and working form level. First, the students could change several variables at the same time (Study I: CO₂ concentration + scenarios; Study II: warming + acidification + eutrophication + salinity) and had to observe their effects on several aspects (Study I: equilibrium reaction + calcification process of the coral; Study II: bladderwrack + amphipods + epiphytes + their interactions). This non-linear structure of model-based simulations resulted in higher complexity of the learning method. As already stated several times this is attributed to a high level of cognitive load (de Jong, 2010; Stull & Hegarty, 2016). Second, the digital work on the tablet (typing, pushing, observing, reading, etc.) possibly led to higher cognitive demands than the execution of real experiments (cutting, filling, measuring, recording, etc.). Especially for students with low media competence, the work on the simulation could have led to cognitive load. Third, the different working methods may have led to a perceived higher load during the simulations. In the experiments, the division of labor allowed weaker students in particular to take a back seat. This was not the case with the simulations since everyone had to do the same work. Having one's own device for working on the simulation requires that each student participate equally and makes it difficult to "submerge" oneself in the group.

Situational interest

Since no difference was found between the two treatment groups in interest in biology and chemistry, this may indicate that the observed differences in situational interest between the groups are not due to differences in individual interest. The non-existing differences in interest in working at the different learning locations indicate that in our case it was not crucial where the students worked but rather with which methods. Since the outreach activities were carried out by the same persons at both learning sites, the dependence of the teacher could also be of decisive importance (Hattie, 2012). As shown in Sect. 2.3., situational interest does not depend on the individual preference for a certain object but rather on the interest of the object in a concrete situation. The subdivision of the interest into emotional, value-related and epistemic components allows a more detailed analysis of the students' perceived interest.

The learners associated working with both methods with positive feelings and emotions (emotional component). The higher situational interest on epistemic and value-related level while working on experiments compared to computer simulations contradicts the hypothesis of our third research question. The literature, which emphasizes the interest-promoting and motivating character of computer simulations, largely originates from a pre-digital era in science education (Jain & Getis, 2003; Rutten et al., 2012). Today's students grow up with digital media (so-called digital natives) and are not easily impressed by a "novelty" effect, the interactivity, the possibility of self-determined interaction within the simulation, or the digital medium itself. This will probably have an impact on the value-related component in particular. Nowadays, working with a digital medium itself no longer has any special value for the learners because they are used to it. This leads to the question of whether the existing assumptions are still valid or need to be reassessed considering the digital transformation of society. Consequently, the often-shown interest-enhancing effect of computer-based (learning) environments might be outdated — while hands-on learning experiences might have a greater motivational and interest-enhancing impact than expected (especially with regard to the value-related and epistemic components). One way of enhancing situational interest through digital media could be the use of gaming effects or immersive aspects like augmented/virtual reality (Makransky et al., 2020; McClarty et al., 2012; Parong & Mayer, 2018) and should be considered in the development of future digital learning formats.

Besides the medium itself, the activity during conducting the methods may also have an influence on the development of situational interest: Practical work while conducting experiments is usually more diverse and possibly more exciting than operating digital controllers, sliders and toolboxes. In everyday school life teachers often lack time, equipment and opportunities to let students conduct experiments independently (Hanekamp, 2014). Therefore, practical work might be special and peak their situational interest. This conforms to the research of Lin et al. (2013) who indicated a positive influence of activities that are novel or conspicuous on students' situational interest. These attributes could have led in particular to the students' desire to learn more about the topics they dealt with through working with the methods (epistemic component). Furthermore, the division of labor in the experiments might have been more enjoyable for the students, as the interdependence in the different steps of the work led to a higher interactivity within the group and supported social learning more than in the simulations.

Limitations

The following aspects limit the above described implications of this paper. First of all, the results and implications presented here are of course only valid for the interventions described in this paper. Especially the specific context of the activities (complex, current scientific topics) and characteristics of the methods (instructional support, collaboration forms, etc.) influenced the results. The transferability to other outreach activities should therefore be investigated in further research. The role of the educator has been shown to have a major impact on the outcomes of educational activities (Hattie, 2012). We tried to keep this factor constant by using the same supervisors in both studies. However, an influence of the educator on the outcomes cannot be excluded. The two studies differed not only in the learning location but also in the duration of the interventions, which is a theoretical limitation. The difference in duration was due to the more complex content

of the second intervention compared to the first. However, since the results show a high degree of similarity, we conclude that this difference did not have a significant impact on the outcomes. The low reliability we found for the instrument for cognitive load (Table 3) raises some doubts about the value of our results. However, we have used this measurement because it has been shown that this subjective evaluation scale is the most sensitive measure available to differentiate the cognitive load of different teaching methods (Sweller, 2011). Furthermore, Naismith and Cavalcanti (2015) stated in a review that the validity of self-report instruments does not differ from other instruments such as secondary task methods or instruments with physiological index. To improve the rigor of measuring cognitive load in future the usage of a more sophisticated measurement like dual-task measurement (Brünken et al., 2004) or psychometric scales that distinguish between categories of cognitive load (Leppink et al., 2013) might be preferable. A certain limitation of the results regarding situational interest represents the low reliability of the scales in Study I (Table 3). This is especially noticeable for the epistemic and value-related situational interest scales, which show a significant difference between the methods. That leads to questioning whether these items are accurately measuring the associated constructs. However, since in Study II the Cronbach's alpha values were sufficiently high (all α -values > 0.7) and the results very similar, we conclude that the results are nevertheless reliable.

Conclusion

This paper provides insights into some students' motivational and cognitive effects while working on experiments and computer simulations. We could show repeatedly the advantages of interactive computer simulations with regard to the communication of complex contents as well as the importance of hands-on learning activities to promote students' situational interest (especially with regard to the value-related and epistemic components). We assume that the results are independent of the learning location, because the investigations in both studies showed the same results. Furthermore, the high number of test persons ($N_{\text{total}} = 810$ students) makes this paper particularly meaningful. The assumptions offer educators indications for their choice of an appropriate method for communicating complex socio-scientific issues for school and out-of-school education. We pointed out that promoting interest is just as important for science education as teaching content knowledge. Since both methods may complement and supplement each other, a combined approach of both methods seems to be the right way to benefit most from the positive effects we could observe (knowledge gain + motivation). Subsequent to our findings, Puntambekar et al. (2020) indicated that learning with one method can complement and complete the weaknesses of the other (different focus of the content learned), and also point out to combine both strategically. In their study, they found that students' discussions were influenced by the distinguishing characteristics of the methods: Students conducting hands-on physics labs talked about how to set up equipment, make accurate measurements, and use formulas to calculate output quantities. Conversely, students performing virtual labs engaged more with variable inputs, predictions, and interpretation of phenomena. Additionally, Smetana and Bell (2012) as well as Lichti and Roth (2018) also pointed out the combination of laboratory experiments and model-based simulations as a potentially powerful learning tool. In addition, de Jong et al. (2013) emphasize that a combination of virtual and non-virtual

Table 8 Used questions for pre-test knowledge acquisition

Study I (multiple-choice questions)	Study II (open questions)
<p>(1) What happens when CO₂ is dissolved in sea-water?</p> <p>(a) Mainly hydrogen carbonate ions (HCO₃⁻) are formed</p> <p>(b) Mainly carbon acid (H₂CO₃) is formed</p> <p>(c) The pH-value rises</p> <p>(d) The content of calcium carbonate ions increases</p> <p>(2) What is the current CO₂ concentration in the atmosphere? (indicated in ppm = parts per million)</p> <p>(a) 275 ppm</p> <p>(b) 1005 ppm</p> <p>(c) 885 ppm</p> <p>(d) 408 ppm</p> <p>(3) Increasingly, scientists are observing that corals (organisms made of lime: calcium carbonate, CaCO₃) are unable to build their skeletons sufficiently quickly to form coral reefs. Which statement in this regard is true?</p> <p>(a) Lime is a chemical compound of calcium, carbon and oxygen. The increased input of CO₂ consumes so much oxygen that corals are therefore unable to build up their calcium carbonate skeleton</p> <p>(b) The lower pH value in the sea ensures a lower physiological activity of corals, which is why they cannot build up their skeleton sufficiently quickly</p> <p>(c) The warmed atmosphere due to an increased CO₂ concentration (greenhouse effect) ensures a higher average temperature of the oceans. However, corals are cold-water organisms and can therefore no longer build up their calcium carbonate skeleton sufficiently quickly</p> <p>(d) The buffering effect of seawater causes a decrease in carbonate ions (CO₃²⁻), which are needed for the formation of lime</p>	<p>(1) Ecosystems change under human influence. List all the influences you can think of from which the Baltic Sea is suffering/will suffer</p> <p>(2) What is meant by "eutrophication of a water body"? Describe the causes, the process and possible effects</p> <p>(3) What is meant by "acidification of a water body"? Describe the causes, the process and possible effects</p>

learning elements seems to enable a deeper and more nuanced understanding, especially when teaching complex issues. We conclude by stating that motivational and cognitive effects of such a combination should be investigated in future research.

Appendix

See Tables 8, 9 and 10.

Table 9 Example items of the multiple-choice questionnaire for each content area of Study I

Content area	Example items	No. of items
pH-value	<p>Since the beginning of industrialization, the pH value of the ocean's surface water has decreased by an average of 0.1 units. Which statements about the pH value are true? Several answers are correct</p> <p>(a) The pH value provides information about the acidic or basic character of an aquatic solution</p> <p>(b) A doubling of the pH value, e.g. from pH=2 to pH=4, means that the basic or acidic character of the solution has doubled</p> <p>(c) The pH value is a logarithmic scale</p> <p>(d) The pH scale is divided into: < 7 = basic, 7 = neutral, > 7 = acidic</p> <p>(e) The more oxonium ions (H_3O^+) there are in the solution, the higher the pH value</p>	3
Increase of global CO_2 concentration	<p>What is the current CO_2 concentration in the atmosphere? (indicated in ppm = parts per million)</p> <p>(a) 275 ppm</p> <p>(b) 1005 ppm</p> <p>(c) 885 ppm</p> <p>(d) 408 ppm</p>	2
Equilibrium reaction of CO_2 in seawater	<p>What happens when CO_2 is dissolved in seawater?</p> <p>(a) Mainly hydrogen carbonate ions (HCO_3^-) are formed</p> <p>(b) Mainly carbon acid (H_2CO_3) is formed</p> <p>(c) The pH-value rises</p> <p>(d) The content of calcium carbonate ions increases</p>	3
Effect on calcifying organisms	<p>Increasingly, scientists are observing that corals (organisms made of lime: calcium carbonate, CaCO_3) are unable to build their skeletons sufficiently quickly to form coral reefs. Which statement in this regard is true?</p> <p>(a) Lime is a chemical compound of calcium, carbon and oxygen. The increased input of CO_2 consumes so much oxygen that corals are therefore unable to build up their calcium carbonate skeleton</p> <p>(b) The lower pH value in the sea ensures a lower physiological activity of corals, which is why they cannot build up their skeleton sufficiently quickly</p> <p>(c) The warmed atmosphere due to an increased CO_2 concentration (greenhouse effect) ensures a higher average temperature of the oceans. However, corals are cold-water organisms and can therefore no longer build up their calcium carbonate skeleton sufficiently quickly</p> <p>(d) The buffering effect of seawater causes a decrease in carbonate ions (CO_3^{2-}), which are needed for the formation of lime</p>	1

Table 10 Example items of the multiple-choice questionnaire for each content area of Study II

Content area	Example items	No. of items
Organisms and interactions	Which of the following is a characteristic of bladderwrack? (a) The grass-like alga grows mainly on sandy substrates (b) The bladderwrack is a seaweed genus of red algae (c) Gas bubbles give the kelp buoyancy in the water (d) The bladderwrack grows on other plants to get a better supply of light	4
Processes of the changes	Which statement regarding the effects of eutrophication on the Baltic Sea is not true? (a) Increased decomposition of organic material results in oxygen minimum zones in the surface water (b) The increased nutrient load in the water results in increased biomass production (c) Especially fast growing plants benefit particularly from eutrophication and cause "algae blooms" (d) The oxygen minimum zones on the seafloor resulting from eutrophication represent a life-threatening situation for all benthic organisms	4
Effects of single change	Due to the lower salinity in the Baltic Sea... (a) There will be more amphipods (b) There will be less epiphytes (c) There will be more bladderwrack (d) There will be less amphipods	4
Effects of combined changes	What causes particularly strong growth of epiphytes? (a) Eutrophication + acidification (b) Warming + lower salinity (c) Acidification + lower salinity (d) Eutrophication + warming	3

Funding Open Access funding enabled and organized by Projekt DEAL.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Anmarkrud, Ø., Andresen, A., & Bråten, I. (2019). Cognitive load and working memory in multimedia learning: Conceptual and measurement issues. *Educational Psychologist, 54*(2), 61–83.
- Bennett, D. J., & Jennings, R. C. (Eds.). (2011). *Successful science communication: Telling it like it is*. Cambridge University Press.
- Bennett, R. E., Persky, H., Weiss, A., & Jenkins, F. (2010). Measuring problem solving with technology: A demonstration study for NAEP. *The Journal of Technology, Learning and Assessment, 8*(8), 4–89.
- Berger, M. (2018). *Neue Medien im experimentellen Physikunterricht der Sekundarstufe I: Eine empirisch-explorative Studie zur Untersuchung der Auswirkungen von virtuell durchgeführten physikalischen Experimenten auf die Motivation der Lernenden im Sekundarstufenbereich I*. Dissertation, Pädagogische Hochschule Heidelberg, Heidelberg.

- BIOACID (Ed.) (2012). *Das andere CO2-Problem: Ozeanversauerung: Acht Experimente für Schüler und Lehrer*. Helmholtz-Zentrum für Ozeanforschung Kiel (GEOMAR).
- Braund, M., & Reiss, M. (2006). Towards a more authentic science curriculum: The contribution of out-of-school learning. *International Journal of Science Education*, 28(12), 1373–1388.
- Brünken, R., Plass, J. L., & Leutner, D. (2004). Assessment of cognitive load in multimedia learning with dual-task methodology: Auditory load and modality effects. *Instructional Science*, 32(1–2), 115–132.
- Chini, J. J., Madsen, A., Gire, E., Rebello, N. S., & Puntambekar, S. (2012). Exploration of factors that affect the comparative effectiveness of physical and virtual manipulatives in an undergraduate laboratory. *Physical Review Special Topics Physics Education Research*, 8(1), 24.
- Darrah, M., Humbert, R., Finstein, J., Simon, M., & Hopkins, J. (2014). Are virtual labs as effective as hands-on labs for undergraduate physics?: A comparative study at two major universities. *Journal of Science Education and Technology*, 23(6), 803–814.
- Develaki, M. (2017). Using computer simulations for promoting model-based reasoning. *Science & Education*, 26(7–9), 1001–1027.
- de Jong, T. (2010). Cognitive load theory, educational research, and instructional design: Some food for thought. *Instructional Science*, 38(2), 105–134.
- 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.
- Di Fuccia, D., Witteck, T., Markic, S., & Eilks, I. (2012). Trends in practical work in German science education. *EURASIA Journal of Mathematics Science and Technology Education*, 8(1), 48.
- Eckhardt, M., Urhahne, D., & Harms, U. (2018). Instructional support for intuitive knowledge acquisition when learning with an ecological computer simulation. *Education Sciences*, 8(3), 94.
- Ekwueme, C. O., Ekon, E. E., & Ezenwa-Nebife, D. C. (2015). The impact of hands-on-approach on student academic performance in basic science and mathematics. *Higher Education Studies*, 5(6), 47.
- Engeln, K. (2004). *Schülerlabors: authentische, aktivierende Lernumgebungen als Möglichkeit, Interesse an Naturwissenschaften und Technik zu wecken*. Dissertation, Christian-Albrechts-Universität zu Kiel, Kiel.
- Evangelou, F., & Kotsis, K. (2019). Real vs virtual physics experiments: Comparison of learning outcomes among fifth grade primary school students: A case on the concept of frictional force. *International Journal of Science Education*, 41(3), 330–348.
- Finkelstein, N. D., Adams, W. K., Keller, C. J., Kohl, P. B., Perkins, K. K., Podolefsky, N. S., et al. (2005). When learning about the real world is better done virtually: A study of substituting computer simulations for laboratory equipment. *Physical Review Special Topics - Physics Education Research*, 1(1), 340.
- Frey, A., Taskinen, P., Schütte, K., Prenzel, M., Artelt, C., Baumert, J., et al. (2009). *PISA-2006-Skalenhandbuch: Dokumentation der Erhebungsinstrumente*. Münster: Waxmann.
- Glowinski, I., & Bayrhuber, H. (2011). Student labs on a university campus as a type of out-of-school learning environment: Assessing the potential to promote students' interest in science. *International Journal of Environmental and Science Education*, 6(4), 371–392.
- Hanekamp, G. (2014). Zahlen und Fakten: Allensbach-Studie 2013 der Deutsche Telekom Stiftung. In J. Maxton-Küchenmeister & J. Meßinger-Koppelt (Eds.), *Naturwissenschaften. Digitale Medien im naturwissenschaftlichen Unterricht* (pp. 21–28). Joachim-Herz-Stiftung Verlag.
- Hattie, J. (2012). *Visible learning for teachers: Maximizing impact on learning*. Routledge.
- Hofstein, A., & Lunetta, V. N. (2004). The laboratory in science education: Foundations for the twenty-first century. *Science Education*, 88(1), 28–54.
- Höft, L., Bernholt, S., Blankenburg, J. S., & Winberg, M. (2019). Knowing more about things you care less about: Cross-sectional analysis of the opposing trend and interplay between conceptual understanding and interest in secondary school chemistry. *Journal of Research in Science Teaching*, 56(2), 184–210.
- Itzek-Greulich, H., Flunger, B., Vollmer, C., Nagengast, B., Rehm, M., & Trautwein, U. (2015). Effects of a science center outreach lab on school students' achievement: Are student lab visits needed when they teach what students can learn at school? *Learning and Instruction*, 38, 43–52. <https://doi.org/10.1016/j.learninstruc.2015.03.003>
- Jain, C., & Getis, A. (2003). The effectiveness of internet-based instruction: An experiment in physical geography. *Journal of Geography in Higher Education*, 27(2), 153–167.
- Jong, T., de Linn, M. C., & Zacharia, Z. C. (2013). Physical and virtual laboratories in science and engineering education. *Science (New York, NY)*, 340(6130), 305–308.
- Kalyuga, S. (2007). Enhancing instructional efficiency of interactive E-learning environments: A cognitive load perspective. *Educational Psychology Review*, 19(3), 387–399.

- Kiroğlu, K., Türk, C., & Erdoğan, İ. (2019). Which one is more effective in teaching the phases of the moon and eclipses: Hands-on or computer simulation? *Research in Science Education*, 23(11), 1095.
- Köck, H. (2018). Fachliche Konsistenz und Spezifik in allen Dimensionen des Geographieunterrichts. In A. Rempfler (Ed.), *Unterrichtsqualität Wirksamer Geographieunterricht 5* (pp. 110–121). Schneider Verlag Hohengehren GmbH.
- Krapp, A. (2007). An educational-psychological conceptualisation of interest. *International Journal for Educational and Vocational Guidance*, 7(1), 5–21.
- Krapp, A., Hidi, S., & Renninger, A. (1992). Interest, learning and development. In A. Renninger, S. Hidi, & A. Krapp (Eds.), *The role of interest in learning and development* (pp. 3–25). Hillsdale.
- Lamb, R., Antonenko, P., Etopio, E., & Seccia, A. (2018). Comparison of virtual reality and hands on activities in science education via functional near infrared spectroscopy. *Computers & Education*, 124, 14–26.
- Leppink, J., Paas, F., van Gog, T., van der Vleuten, C. P. M., & van Merriënboer, J. J. G. (2013). Effects of pairs of problems and examples on task performance and different types of cognitive load. *Learning and Instruction*, 30, 32–42.
- Lichti, M., & Roth, J. (2018). How to foster functional thinking in learning environments using computer-based simulations or real materials. *Journal for STEM Education Research*, 1(1–2), 148–172.
- Lin, H.-S., Hong, Z.-R., & Chen, Y.-C. (2013). Exploring the development of college students' situational interest in learning science. *International Journal of Science Education*, 35(13), 2152–2173.
- Lin, Y.-C., Liu, T.-C., & Sweller, J. (2015). Improving the frame design of computer simulations for learning: Determining the primacy of the isolated elements or the transient information effects. *Computers & Education*, 88, 280–291.
- Linn, M. C., Chang, H.-Y., Chiu, J., Zhang, H., & McElhaney, K. (2010). Can desirable difficulties overcome deceptive clarity in scientific visualizations. *Successful Remembering and Successful Forgetting*, 24, 239–262.
- Madden, J., Pandita, S., Schuldt, J. P., Kim, B., Won, S., & Holmes, N. G. (2020). Ready student one: Exploring the predictors of student learning in virtual reality. *PLoS ONE*, 15(3), 0229788.
- Makransky, G., Petersen, G. B., & Klingenberg, S. (2020). Can an immersive virtual reality simulation increase students' interest and career aspirations in science? *British Journal of Educational Technology*, 51(6), 2079–2097.
- Marshall, J. A., & Young, E. S. (2006). Preservice teachers' theory development in physical and simulated environments. *Journal of Research in Science Teaching*, 43(9), 907–937.
- Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? The case for guided methods of instruction. *The American Psychologist*, 59(1), 14–19.
- Mayer, R. E., Hegarty, M., Mayer, S., & Campbell, J. (2005). When static media promote active learning: Annotated illustrations versus narrated animations in multimedia instruction. *Journal of Experimental Psychology: Applied*, 11(4), 256–265.
- McClarty, K. L., Orr, A., Frey, P. M., Dolan, R. P., Vassileva, V., & McVay, A. (2012). A literature review of gaming in education. *Gaming in Education*, 24, 1–35.
- Mutlu-Bayraktar, D., Cosgun, V., & Altan, T. (2019). Cognitive load in multimedia learning environments: A systematic review. *Computers & Education*, 141, 103618.
- Naismith, L. M., & Cavalcanti, R. B. (2015). Validity of Cognitive Load Measures in Simulation-Based Training: A Systematic Review. *Academic Medicine*, 90(11), S24–35.
- National Research Council (NRC) (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. National Academies Press.
- National Science Teachers Association (NSTA). (2013). *Next generation science standards: For states, by states*, from <https://www.nextgenscience.org/get-to-know>.
- Paas, F. G., van Merriënboer, J. J., & Adam, J. J. (1994). Measurement of cognitive load in instructional research. *Perceptual and Motor Skills*, 79(1 Pt 2), 419–430.
- Paas, F., Renkl, A., & Sweller, J. (2003a). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist*, 38(1), 1–4.
- Paas, F., Tuovinen, J. E., Tabbers, H., & van Gerven, P. W. M. (2003b). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist*, 38(1), 63–71.
- Palmer, D. H. (2009). Student interest generated during an inquiry skills lesson. *Journal of Research in Science Teaching*, 46(2), 147–165.
- Park, M. (2019). Effects of simulation-based formative assessments on students' conceptions in physics. *EURASIA Journal of Mathematics, Science and Technology Education*, 15(7), 14.
- Parong, J., & Mayer, R. E. (2018). Learning science in immersive virtual reality. *Journal of Educational Psychology*, 110(6), 785–797.

- Paul, I., & John, B. (2020). Effectiveness of computer simulation in enhancing higher order thinking skill among secondary school students. *UGC Care Journal*, 31(4), 343–356.
- Puntambekar, S., Gnesdilow, D., Dornfeld Tissenbaum, C., Narayanan, N. H., & Rebello, N. S. (2020). Supporting middle school students' science talk: A comparison of physical and virtual labs. *Journal of Research in Science Teaching*, 58(3), 392–419.
- Quellmalz, E. S., DeBarger, A. H., Haertel, G., Schank, P., Buckley, B. C., Gobert, J., et al. (2008). *Exploring the role of Technology-Based Simulations in Science Assessment: The Calipers Project: Assessing Science Learning: Perspectives From Research and Practice*.
- Renninger, A. (2000). Individual interest and its implications for understanding intrinsic motivation. In *Intrinsic and Extrinsic Motivation* (pp. 373–404). Elsevier.
- Rutten, N., van Joolingen, W. R., & van der Veen, J. T. (2012). The learning effects of computer simulations in science education. *Computers & Education*, 58(1), 136–153.
- Scharfenberg, F. J., Bogner, F. X., & Klautke, S. (2007). Learning in a gene technology laboratory with educational focus: Results of a teaching unit with authentic experiments. *Biochemistry and Molecular Biology Education*, 35(1), 28–39.
- Scheuring, M., & Roth, J. (2017). Real materials or simulations? Searching for a way to foster functional thinking. *CERME*, 10, 2677–2679.
- Siahaan, P., Suryani, A., Kaniawati, I., Suhendi, E., & Samsudin, A. (2017). Improving students' science process skills through simple computer simulations on linear motion conceptions. *Journal of Physics: Conference Series*, 812, 12017.
- Smetana, L. K., & Bell, R. L. (2012). Computer simulations to support science instruction and learning: A critical review of the literature. *International Journal of Science Education*, 34(9), 1337–1370.
- Stull, A. T., & Hegarty, M. (2016). Model manipulation and learning: Fostering representational competence with virtual and concrete models. *Journal of Educational Psychology*, 108(4), 509.
- Sweller, J. (2011). Cognitive Load Theory. In *Psychology of learning and motivation* (55th edn., pp. 37–76). Academic Press.
- Trundle, K. C., & Bell, R. L. (2010). The use of a computer simulation to promote conceptual change: A quasi-experimental study. *Computers & Education*, 54(4), 1078–1088.
- Tunnicliffe, S. D. (2017). Overcoming 'Earth Science Blindness' earth science in action in natural history dioramas. In M. F. P. D. C. M. Costa & J. B. Vázquez Dorrió (Eds.), *Hands-on science: Growing with science*. Universidade de Vigo.
- van Merriënboer, J. J. G., & Sweller, J. (2005). Cognitive load theory and complex learning: recent developments and future directions. *Educational Psychology Review*, 17(2), 147–177.
- Wallace, P., Graesser, A., Millis, K., Halpern, D., Cai, Z., Britt, M. A., Magliano, J., & Wiemer, K. (1945). Operation ARIES: A computerized game for teaching scientific inquiry. *Frontiers in Artificial Intelligence and Applications*, 200, 602–604.
- Winberg, T. M., & Berg, C. A. R. (2007). Students' cognitive focus during a chemistry laboratory exercise: Effects of a computer-simulated prelab. *Journal of Research in Science Teaching*, 44(8), 1108–1133.
- Zapata-Rivera, D., Jackson, T., Liu, L., Bertling, M., Vezzu, M., & Katz, I. R. (2014, June). Assessing science inquiry skills using dialogues. In International Conference on Intelligent Tutoring Systems (pp. 625–626). Springer, Cham.
- Zendler, A., & Greiner, H. (2020). The effect of two instructional methods on learning outcome in chemistry education: The experiment method and computer simulation. *Education for Chemical Engineers*, 30, 9–19.