Journal of Innovation in Psychology, Education and Didactics
ISSN: 2247-4579, E-ISSN: 2392-7127
http://www.jiped.ub.ro/
Covered in: EBSCO, CEEOL, ProQuest, ERIH PLUS, DOAJ, Scipio, International Innovative Journal Impact Factor, CiteFactor, EuroPub database, Open Academic Journals Index, ResearchBib, Universal Impact Factor
2026, Volume 30, Issue 1, pages: 125-160, doi:10.29081/JIPED.2026.30.1.09

How are Simulations Used in Science Education? – A Status Quo on Research

Mahdi EL TEGANI [1]*, Claas WEGNER [2]

Received: 31 March 2026/ Accepted: 19 June 2026/ Published: 18 July 2026

Abstract

Simulations are becoming increasingly common as a method of teaching scientific phenomena. Research into the impact of using simulations has also been ongoing for over 20 years. Regarding their impact, various studies have found positive effects on interest, procedural and declarative knowledge, scientific thinking and behavior, causal relationships, and spatial thinking. However, the presentation of the results does not always make it clear in which contexts the respective findings were obtained. This paper aims to provide deeper insight into research on the use of simulations in science education and to present study designs more explicitly. This article should help to make research results easier to interpret and to identify gaps in research methods.

Keywords: digitalization; education; review; Science; simulation; STEM

How to cite*:* El Tegani, M., & Wegner, C. (2026). How are Simulations Used in Science Education? – A Status Quo on Research. *Journal of Innovation in Psychology, Education and Didactics*, *30*(1), 125-160. doi:10.29081/JIPED.2026.30.1.09

[1] M.Ed., Bielefeld University, Germany, E-mail: mahdi.el_tegani@uni-bielefeld.de
[2] Prof. PhD, Bielefeld University, Germany, E-mail: claas.wegner@uni-bielefeld.de
* Corresponding author

1. Introduction

Computer simulations can be understood as programs that replicate dynamic (natural) processes and provide insights that can be applied in the real world (Zauner & Schrempf, 2009). Given their similarities to real experiments (Jebeile, 2017), they can also be viewed as virtual experiments in a scientific context (Bungartz et al., 2013). However, simulations are not limited to science and economics; they are also particularly well-suited for educational contexts. Here, simulations offer a valuable complement to experiments and traditional analog teaching for various reasons (safety concerns, time and space constraints, limited materials or equipment, and visualization options) (Stinken-Rösner, 2020). Simulations have been used in science education for more than 20 years (El Tegani & Wegner, 2025).

Given their similarities to experiments (Jebeile et al., 2017), simulations are particularly relevant for modern science education. They can serve as a tool for knowledge acquisition through inquiry-based learning (Huber et al., 2013), in which students independently investigate research questions using the scientific method, formulate hypotheses, test them through simulations, and draw logical conclusions (Kollar et al., 2011). Students can thus observe the phenomenon of global warming and the death of coral reefs around the world. From this, they develop research questions such as "What happens when the oceans become more acidic?", formulate hypotheses, and can investigate these using a simulation relevant to their topic (Krüger et al. 2022). Thanks to the simulation, students can even pause in the middle of the digital experiment and reflect on the learning process so far (Parush et al., 2002). The process of scientific inquiry is specifically supported by the use of simulations, as simulations allow students to influence variables and observe immediate changes as consequences. Studies have shown that scientific skills, in particular, can be positively influenced. For example, Lin et al. (2012) found a positive effect on scientific competencies. There are further arguments in favor of using simulations in science teaching. By visualizing invisible phenomena, they help make scientific concepts easier to understand (Bell et al., 2013; Ryoo & Linn, 2012; Yoon et al., 2018). They can also help create mental models (Landriscina, 2009). The creation of mental models depends heavily on cognitive load. In 1994, Paas & van Merriënboer developed Cognitive Load Theory (CLT) based on the concept of cognitive load. Here, cognitive load refers to the mental strain an individual experiences while performing a task. The theory posits that working memory has only limited capacity and that tasks an individual performs consume a portion of that capacity. Consequently, the number of tasks an individual can handle simultaneously is limited. A high cognitive load can have negative consequences for learning and lead to the learner being overwhelmed by the task (Paas & van Merriënboer, 1994). This can ultimately be reflected in the learner's performance.

Cognitive load can be reduced by decreasing task complexity, through repetition and the development of routines, or by setting clear goals (Sweller, 1988). Choi et al. (2014) expand the theory to include the influencing factor "physical learning environment," such that learners, tasks, and the environment influence one another in bidirectional and tridirectional ways. It follows that the design and conditions of using simulations as a learning environment and task medium can influence cognitive load and thus learning success. The influence of simulations on cognitive load can be argued on several levels. For instance, simulations can simplify complex processes, visualize invisible phenomena, or incorporate verbal and visual learning aids (Liu & Chuang, 2011). At the same time, simulations are not yet a common learning medium and vary greatly in their design and interactivity depending on the subject and age group, which can increase cognitive load. Cognitive load is one of the central variables examined in the studies presented here.

Compared to real experiments, simulations expand the scope of learning by allowing the speed of processes to be reduced, accelerated, or even stopped, thus allowing time for reflection on the learning process in the middle of the investigation rather than only afterwards (Parush et

al., 2002). By adding elements (e.g., labels) or highlighting certain aspects, attention can be directed to areas relevant to the learning object, but this can also lead to misinterpretation and the wrong aspects being given special significance (López & Pintó, 2017).

The integration of simulations can therefore pose a challenge as well. Simulations are not necessarily reality-based (Jebeile, 2017) and can therefore promote misconceptions. In addition, their didactic effects may be limited, as although they make scientific thinking and working methods more comprehensible (Cayvaz et al., 2020), they cannot replace practical experience in handling equipment and tools (Karlsson et al., 2013). However, these are particularly indispensable for scientific work. In response to the problem of a lack of physical practical experience, new theoretical frameworks such as "Bifocal Modeling" have emerged, which utilize physical laboratory experiments and computer simulations in parallel to generate epistemological knowledge through discrepancies (Blikstein et al., 2016). The use of simulations requires technical prerequisites such as suitable end devices, software, and, if necessary, licenses. In addition, teaching staff must know how to use simulations meaningfully (Stinken-Rösner, 2020). The ongoing digitization of schools is making it easier to use simulations in science lessons. The coronavirus pandemic and the associated shift to digital distance learning in Germany, for example, have meant that by 2023, more than three-quarters of all schools had at least one class set of digital devices (VBE, 2023). As a result, simulations have once again gained importance as a (more or less effective) substitute for laboratory experiences in so-called virtual laboratories (Schäfers et al., 2020). However, to exploit this potential for meaningful simulation use, appropriate resources, structures, and educational opportunities for teachers are required (Kabaum & Anders, 2020). Over the past few decades, simulations have become an increasingly important focus of educational research and represent a relevant teaching tool for both children and young people (Samavedam, 2020) and for higher education (Chernikova et al., 2024). The topics covered in science lessons are as diverse as the teaching models and simulations that can be used in these areas of education. Although the natural sciences share a common foundation in terms of the methods used to acquire knowledge, the subject-specific skills required vary greatly. For example, the ability to analyze and interpret unadjusted measurement data containing confounding factors is essential in physics. This therefore requires special consideration in the design of simulations for physics instruction (Bumbacher et al., 2018). The visualization and quantification of forces and fields is also of particular interest in physics and chemistry. This example demonstrates that each subject has different requirements for simulation properties, particularly within the research landscape on simulations in science education. Previous research on simulations in science education varies widely in its methods, topics, and research focus (El Tegani & Wegner, 2025). Broadly speaking, the variables examined in the recent literature can be grouped into two main theoretical areas: "Knowledge, Performance, and Competencies" and "Affective Characteristics." Regarding Knowledge, Performance, and Competencies, many studies focus on the effect on the technical communication of topics (Jaakkola et al., 2011; Magana et al., 2019; Pucholt, 2021; Yoon et al., 2018).

Research in this area distinguishes between the acquisition of declarative factual knowledge and deeper procedural skills, such as mathematical modeling or setting up chemical equations (Bilek et al., 2018; Shegog et al., 2012). Regarding affective characteristics, several studies investigate how the use of simulations impacts students' psychological and motivational states. These studies focus on intrinsic and extrinsic motivation, interest, scientific identity, and self-efficacy (Dickes et al., 2019; Garneli & Chorianopoulos, 2018; Makransky et al., 2020). Ng & Chu (2021) demonstrated that higher intrinsic motivation and peer support correlate significantly with better performance in simulation tasks, while Makransky et al. (2020) highlighted how immersive simulations can specifically elevate scientific ambition and self-efficacy, particularly among diverse demographic groups. Even within the subjects, the simulation topics vary greatly. They range from crystal structures in chemistry (Urhahne et al., 2009) to the ratio of acids and bases (Tolentino et al., 2009) to the formulation of reaction

equations (Bílek et al., 2018). This demonstrates the diversity of research focuses in this field. However, the research methods used to investigate this field also vary greatly. This highlights the need to categorize the diverse research focuses in this field. Although each subject has its own specific requirements and methods, they all share the scientific method by testing hypotheses in a similar way. Therefore, simulations can be useful for all of them. The diversity of different subjects limits the amount of knowledge each subject can gather by focusing solely on its own research. Therefore, a recent overview of a holistic approach demonstrates significant value for all science subjects and offers new perspectives on how to use simulations in science education. Bifocal modeling could therefore also be successfully used in chemistry or physics to enrich education. Consequently, mapping the current state of how simulation methods are actually integrated into the teaching of these exact sciences in schools is both a necessity and a way to move beyond isolated findings, to assess the real pedagogical effects of their use, and to equip educators with evidence-based teaching strategies.

2. Methodology

This study was conducted as a systematic literature review to examine how simulations are used in science education and their reported impact on students learning process . The review followed the principles displayed in the former Article of El Tegani & Wegner (2025).

A comprehensive search was carried out in major academic databases, including FIS Fachportal Pädagogik & Web of Science. The search combined keywords such as "simulation*", "bildung*", "Naturwissen*", "MINT*", "physik*", "Chemi*", "Informatik*" and their English equivalents in the categories, using Boolean operators (AND, OR). Only peer-reviewed articles published in English between 2000 and 2023 were considered.

The inclusion criteria comprised empirical studies investigating the use of simulations in science education at primary, secondary, or higher education levels. Studies were required to report educational outcomes related to student learning, engagement, conceptual understanding, scientific skills, or attitudes toward science. Review articles, conference abstracts, editorials, dissertations, studies unrelated to science education, and publications without full-text access were excluded.

The selection process consisted of four stages: identification of records through database searches, removal of duplicate publications, screening of titles and abstracts, and full-text assessment for eligibility. Relevant information extracted from each study included publication year, country, educational level, scientific discipline, type of simulation used, research design, sample characteristics, and key findings.

The included studies were synthesized using a qualitative thematic analysis. Findings were grouped according to recurring themes, including the educational purposes of simulations, their effectiveness in improving learning outcomes, factors influencing successful implementation, and reported limitations. The synthesis aimed to identify common trends, differences across educational contexts, and gaps in the existing literature.

3. Results

Table 1 provides a systematic overview of the study designs and includes key indicators such as sample size, variables, methods, measurement instruments used, location, level, age of participants, subject, and type of school, and briefly presents the relevant content and results. Main characteristics of each study were collected in a summarizing table and then visualized in statistical histograms, which show the count of studies divided by different characteristics. The analysis results are presented clearly in graphical form below (Figure 1 – Figure 7).

Excluding dropouts, most studies had a total sample size of 25-75 participants, while a few studies had 250 or more participants. This is not a surprise due to the fact that most studies

investigated learning groups in classes which contain mostly 25 to 30 students. Figure 1 shows the number of studies divided by sample size.

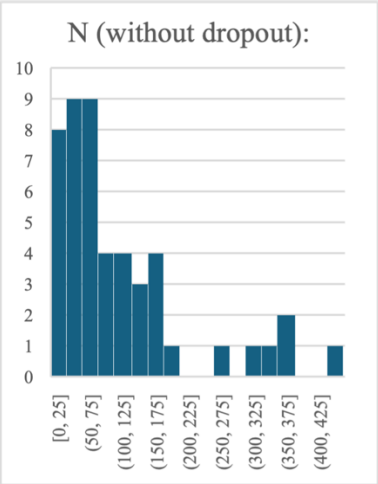


Figure 1. Deviation of studies (y-axis) ordered by number of participants per study (x-axis)

It is also striking that the studies focused on grades 7 to 11. Primary school (grades 1 to 4), in particular, was hardly studied at all (Figure 2). This lack may stem from the fact that the use of digital media is usually systematically implemented only from secondary school onwards (Wastiau et al., 2013). Similarly, science topics become increasingly complex and abstract as pupils progress through the school years, making the use of visualized and didactically simplified representations of phenomena more attractive. The distribution of investigated class levels in the included studies is displayed in Figure 2.

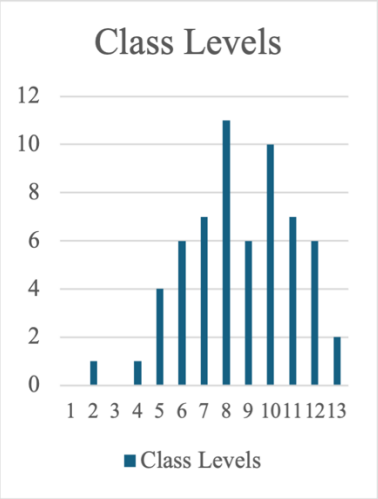


Figure 2. Number of studies (y-axis) investigating different class levels (x-axis)

Regarding the timing of the test, most studies used a pre-post design (36). However, a few studies also used pre-midway-post designs (2; Makransky et al., 2020; Tapola et al., 2013) and pre-post follow-up designs (4; Han & Black, 2011; Pucholt, 2021). The imbalance of the test timing approaches can be seen in Figure 3.

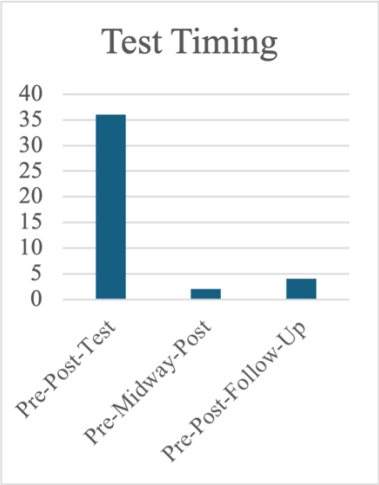


Figure 3. Number of studies (y-axis) investigating different points of time(x-axis)

Most studies used treatment as the independent variable. The treatments differed significantly from one another. The simulation was often researched without a control group (e.g., Magana et al., 2019; Ng & Chu, 2021; Reilly et al., 2021; Rosenberg & Lawson, 2019). Other studies compared the use of simulations with traditional laboratory environments or experiments (Krüger et al., 2022; H. Liu & Su, 2011; Pucholt, 2021; Zendler & Greiner, 2020), traditional teaching (Eckhardt et al., 2013; Gelbart et al., 2009; Sabagh, 2011; Thisgaard & Makransky, 2017), or other media such as overhead projections, as an example of static images (Price, 2013), videos (Chang, 2017; Makransky et al., 2020; Mešić et al., 2015; Ndihokubwayo et al., 2020). The way in which simulations can be used has also been investigated by several studies. For example, the sequence with other methods (Chen et al., 2016; Li et al., 2022; Shegog et al., 2012), as well as the supporting measures for learning with simulations (Chang, 2017; Eckhardt et al., 2018; Jaakkola et al., 2011; H.-C. Liu & Chuang, 2011; Yoon et al., 2018). In some cases, the multimodal representation of the simulation (Han & Black, 2011) and its thematic content (Han & Black, 2011; López & Pintó, 2017; Tapola et al., 2013; Yoon et al., 2018) were also varied. In addition to the treatment, some studies also included other independent variables. These were primarily age (2) (Makransky et al., 2020), gender (3) (Makransky et al., 2020; Tapola et al., 2013), and prior knowledge (3) (Barab et al., 2007; Magana et al., 2019; Tapola et al., 2013). The deviation of independent variables investigated by the studies is shown in Figure 4.

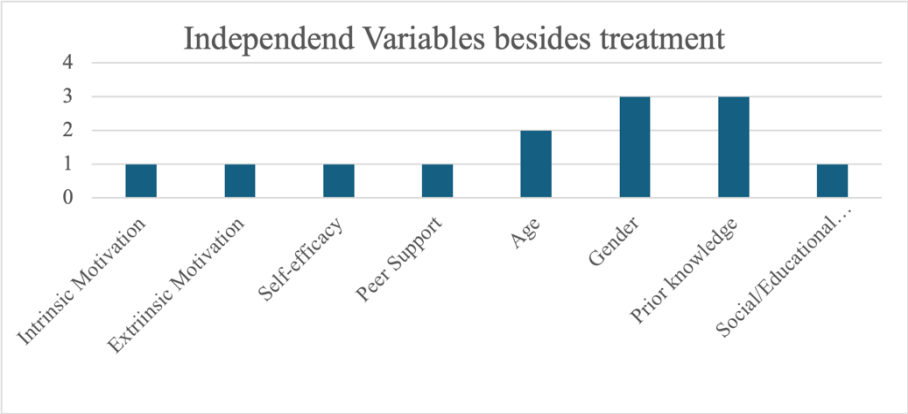


Figure 4. Number of studies (y-axis) that investigated the influence of independent variables (x-axis) besides the use of simulations

A total of 17 studies conducted one-shot treatments, while 29 studies conducted interventions at multiple time points (Figure 5). The testing period ranged from one lesson to a three-month treatment.

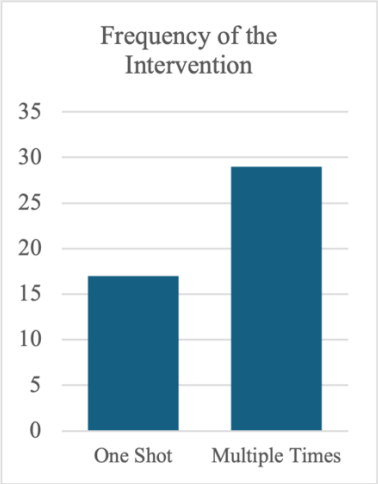


Figure 5. Number of studies (y-axis) that used simulations once or multiple times (x-axis)

Most studies used quantitative research methods, such as performance tests or questionnaires (Figure 6).

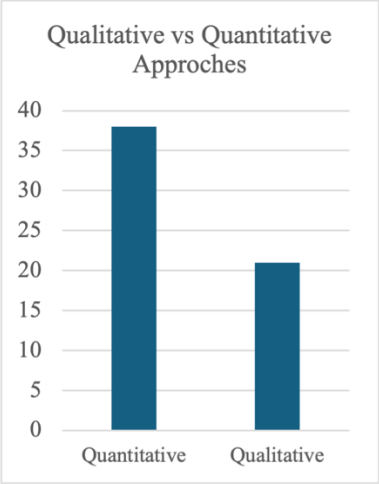


Figure 6. Number of studies (y-axis) that reported quantitative or qualitative approaches

Those studies that also used qualitative methods often used interviews (12), video recordings (7), and open questions (9) (Figure 7).

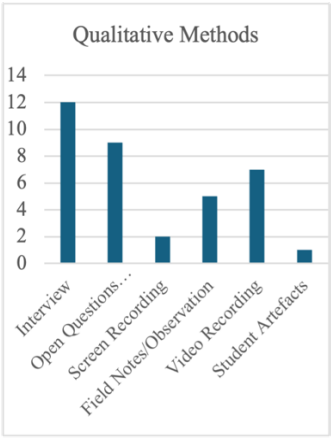
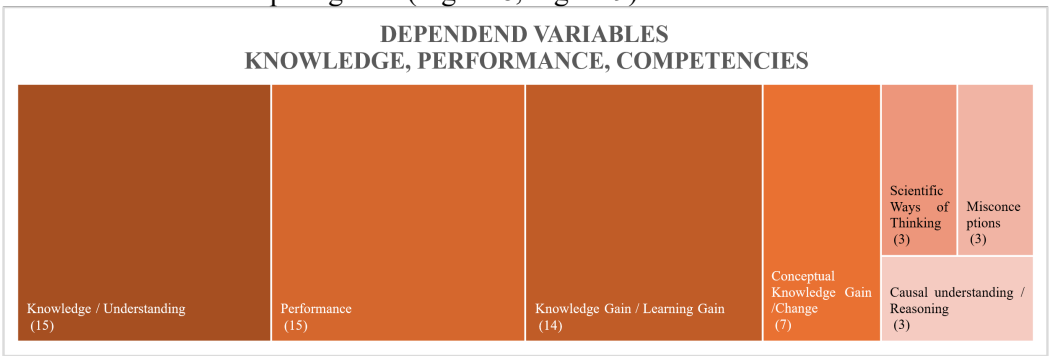


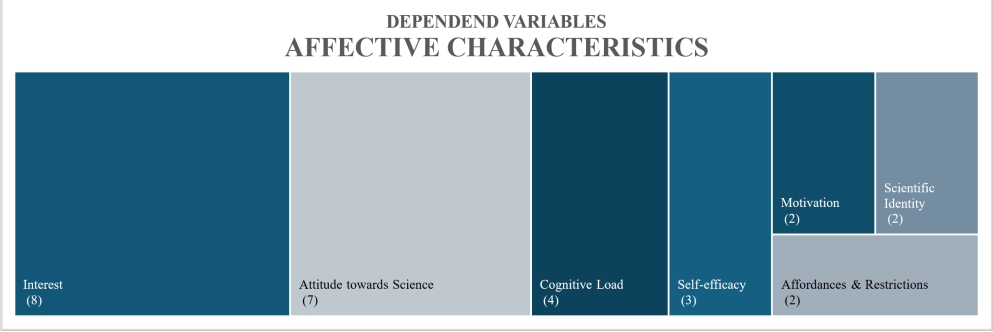
Figure 7. Number of studies (y-axis) regarding the reported qualitative methods used

Two main objective categories can be identified as dependent variables. One objective can be summarized as "Knowledge, Performance and Competencies" (Figure 8), which includes dependent variables that measure absolute knowledge or knowledge gain, overall performance across a variety of tasks or scientific skills, and competencies such as reasoning and scientific ways of thinking. The other objective was to examine "Affective characteristics" (Figure 8), such as interest, attitude towards science, and cognitive load. For an easier overview, the analysis results are shown in treemap diagrams (Figure 8, Figure 9).



Note: Total number of studies investigated

Figure 8. Treemap diagram of studies investigated on dependent variables regarding the categories "Knowledge, Performance or Competencies" (with numbers in brackets)



Note: Total number of studies investigated

Figure 9. Tree map diagram of studies investigation on dependent variables regarding "Affective characteristics" (with numbers in brackets)

4. Discussion

This article provides a comprehensive overview of research on the use and impact of simulations in science education. The analyzed papers indicate that simulations can be used in a variety of different settings in science education.

A major strength of the studies conducted is the great diversity of methodological approaches and research foci. Although most studies used quantitative methods, qualitative and mixed-method approaches were also effectively utilized to capture a holistic overview of simulation-based learning. Furthermore, the literature successfully evaluates the instructional potential of simulations across a wide spectrum of variables, showing positive effects on both cognitive gains—such as knowledge and performance—and affective characteristics, which should be investigated even more closely in the future.

In addition to these strengths, several methodological limitations must be noted. The results show that the pre-post test approach was used most frequently. Consequently, only very limited—if any—conclusions can be drawn regarding the long-term impact of simulations in science education. Only one study employed follow-up surveys; however, no differences were found compared to the control group (Pucholt, 2021). Studies are therefore needed to investigate the long-term effects of using simulations in science education. The majority of the studies included fewer than 100 participants. These numbers could be explained by the fact that the studies were often conducted in class groups. These are naturally rarely larger than 30 students. Since simulations are strongly content-related, their use is also thematically anchored in the curriculum. That anchoring, therefore, limits their use to certain grade levels. Cross-school research could help obtain a larger sample. Finally, there is an imbalance in the targeted demographics. Most studies focused on secondary school or upper secondary school grades. That focus could be due to lower use of technical equipment in lower grades (Wastiau et al., 2013) or to lower topic complexity. In addition, primary school classes often lack dedicated science lessons. The reason the 13th grade is particularly underrepresented could be that not all education systems in the studies examined provide 13 years of schooling. In addition, the final years of school are particularly exam-intensive in preparation for graduation, which can lead to a lower willingness to participate in research. Given the increasing use of technologies in education and the accompanying ongoing discussions about the meaningful use of these technologies, studies are needed to close the research gap, particularly in the lower grade levels. However, the selection of simulations could be complicated by students' still-limited reading proficiency. Similar patterns can also be observed in the distribution of the dependent variables.

A critical issue identified across the literature is the limitations regarding the reporting of results. In many cases, it was not possible to extract comprehensive information on the study design. Some studies specified only the participants' age, while others provided information on grade levels. Additionally, the specific test instruments were not detailed in all studies. Furthermore, it was often not possible to determine the exact proportion of total teaching time spent using simulations.

Because simulations can vary significantly in terms of content, usability, interface, and complexity, these omissions in reporting make it exceedingly difficult to draw absolute causal comparisons between different educational interventions. Due to the small number of participants in the individual studies and the wide range of research approaches, subjects, settings, and variables examined, it is not possible to make universally valid statements about the use of simulations, meaning no absolute causalities can be compared. When comparing research results, this limiting factor must be taken into account to avoid drawing false conclusions.

5. Conclusion

These findings yield several important implications for educational practice. Repeated use of simulations could help reduce cognitive load, as the working environment is no longer

perceived as entirely new, which could lead to improved learning (Choi et al., 2014). However, educators must exercise caution: interest can arise when knowledge about a subject is lacking and perceived as new (Kramer, 2025). This novelty effect can be lost with repeated use, leading to a decline in interest. The use of simulations can foster knowledge acquisition, interest, causal understanding, self-efficacy, and scientific thinking. Simulations are a promising addition to real experiments in new learning approaches such as "Bifocal Modeling" and can be used in a variety of settings, ranging from replacing experiments that are too dangerous to providing in-depth tasks for completion at home to fostering scientific competencies in projects. Although this paper provides a comprehensive overview of research on simulations in science education, several limitations should be considered when interpreting the results. The results presented here have been significantly condensed; for more comprehensive information, it is recommended to consult the original studies. Additionally, only studies that 1) were available in German or English, 2) were accessible to the authors, and 3) met the additional criteria outlined by El Tegani and Wegner (2025) were evaluated. Future research on this topic could also incorporate perspectives from teachers and educational institutions to pursue a holistic research approach. The educational shift from STEM to STEAM holds additional potential for creativity and learning (Boy, 2013). Including artistic subjects in an analysis of simulations could therefore yield new insights. Some age groups are also underrepresented in the available studies. The impact of simulations in primary school teaching is particularly relevant given the increasing digitalization. Nevertheless, this study represents an important contribution, as it summarizes current findings, clearly presents research approaches, and highlights existing gaps.

References

1. Aşkar, P., Yavuz, H., & Köksal, M. (1992). Students' perceptions of computer assisted instruction environment and their attitudes towards computer assisted learning. *Educational Research*, *34*(2), 133–139. https://doi.org/10.1080/0013188920340204
2. Bakeman, R., & Gottman, J. M. (1997). *Observing Interaction: An Introduction to Sequential Analysis* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511527685
3. Bandura, A. (1986). Social foundations of thought and action. *Englewood Cliffs, NJ*, *1986*(23–28), 94–104.
4. Barab, S. A., Sadler, T. D., Heiselt, C., Hickey, D., & Zuiker, S. (2007). Relating Narrative, Inquiry, and Inscriptions: Supporting Consequential Play. *Journal of Science Education and Technology*, *16*(1), 59–82. https://doi.org/10.1007/s10956-006-9033-3
5. Beichner, R. J. (1994). Testing student interpretation of kinematics graphs. *American Journal of Physics*, *62*(8), 750–762. https://doi.org/10.1119/1.17449
6. Bell, R. L., Maeng, J. L., & Binns, I. C. (2013). Learning in context: Technology integration in a teacher preparation program informed by situated learning theory. *Journal of Research in Science Teaching*, *50*(3), 348–379. https://doi.org/10.1002/tea.21075
7. Belland, B. R. (2014). Scaffolding: Definition, Current Debates, and Future Directions. In J. M. Spector, M. D. Merrill, J. Elen, & M. J. Bishop (Eds.), *Handbook of Research on Educational Communications and Technology* (pp. 505–518). Springer New York. https://doi.org/10.1007/978-1-4614-3185-5_39
8. Belland, B. R., Kim, C., & Hannafin, M. J. (2013). A Framework for Designing Scaffolds That Improve Motivation and Cognition. *Educational Psychologist*, *48*(4), 243–270. https://doi.org/10.1080/00461520.2013.838920
9. Bílek, M., Nodzyńska, M., Kopek-Putała, W., & Zimak-Piekarczyk, P. (2018). Balancing chemical equations using sandwich making computer simulation games as a supporting teaching method. *Problems of Education in the 21st Century*, *76*(6), 779–799. https://doi.org/10.33225/pec/18.76.779

10. Blikstein, P., Fuhrmann, T., & Salehi, S. (2016). Using the Bifocal Modeling Framework to Resolve "Discrepant Events" Between Physical Experiments and Virtual Models in Biology. *Journal of Science Education and Technology*, *25*(4), 513–526. https://doi.org/10.1007/s10956-016-9623-7
11. Boy, G. A. (2013). From STEM to STEAM: Toward a human-centred education, creativity & learning thinking. *Proceedings of the 31st European Conference on Cognitive Ergonomics*, 1–7. https://doi.org/10.1145/2501907.2501934
12. Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa
13. Britner, S. L., & Pajares, F. (2001). Self-Efficacy Beliefs, Motivation, Race, And Gender In Middle School Science. *Journal of Women and Minorities in Science and Engineering*, *7*(4), 15. https://doi.org/10.1615/JWomenMinorScienEng.v7.i4.10
14. Bumbacher, E., Salehi, S., Wieman, C., & Blikstein, P. (2018). Tools for Science Inquiry Learning: Tool Affordances, Experimentation Strategies, and Conceptual Understanding. *Journal of Science Education and Technology*, *27*(3), 215–235. https://doi.org/10.1007/s10956-017-9719-8
15. Caglar, F., Shekhar, S., Gokhale, A., Basu, S., Rafi, T., Kinnebrew, J., & Biswas, G. (2015). Cloud-hosted simulation-as-a-service for high school STEM education. *Simulation Modelling Practice and Theory*, *58*, 255–273. https://doi.org/10.1016/j.simpat.2015.06.006
16. Cayvaz, A., Akcay, H., & Kapici, H. O. (2020). Comparison of Simulation-Based and Textbook-Based Instructions on Middle School Students' Achievement, Inquiry Skills and Attitudes. *International Journal of Education in Mathematics, Science and Technology*, *8*(1), 34. https://doi.org/10.46328/ijemst.v8i1.758
17. Chang, H.-Y. (2017). How to augment the learning impact of computer simulations? The designs and effects of interactivity and scaffolding. *Interactive Learning Environments*, *25*(8), 1083–1097. https://doi.org/10.1080/10494820.2016.1250222
18. Chen, C., Schneps, M. H., & Sonnert, G. (2016). Order Matters: Sequencing Scale-Realistic Versus Simplified Models to Improve Science Learning. *Journal of Science Education and Technology*, *25*(5), 806–823. https://doi.org/10.1007/s10956-016-9642-4
19. Chen, F., Zhou, J., Wang, Y., Yu, K., Arshad, S. Z., Khawaji, A., & Conway, D. (2016). *Robust Multimodal Cognitive Load Measurement*. Springer International Publishing. https://doi.org/10.1007/978-3-319-31700-7
20. Chernikova, O., Holzberger, D., Heitzmann, N., Stadler, M., Seidel, T., & Fischer, F. (2024). Where salience goes beyond authenticity: A meta-analysis on simulation-based learning in higher education. *Zeitschrift Für Pädagogische Psychologie*, *38*(1–2), 15–25. https://doi.org/10.1024/1010-0652/a000357
21. Choi, H.-H., Van Merriënboer, J. J. G., & Paas, F. (2014). Effects of the Physical Environment on Cognitive Load and Learning: Towards a New Model of Cognitive Load. *Educational Psychology Review*, *26*(2), 225–244. https://doi.org/10.1007/s10648-014-9262-6
22. Correia, A.-P., Koehler, N., Thompson, A., & Phye, G. (2019). The application of PhET simulation to teach gas behavior on the submicroscopic level: Secondary school students' perceptions. *Research in Science & Technological Education*, *37*(2), 193–217. https://doi.org/10.1080/02635143.2018.1487834
23. Creswell, J. W. (2012). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (4. ed., international ed). Pearson.
24. Dayal, R., Faries, P. L., Lin, S. C., Bernheim, J., Hollenbeck, S., DeRubertis, B., Trocciola, S., Rhee, J., McKinsey, J., Morrissey, N. J., & Kent, K. C. (2004). Computer simulation as a component of catheter-based training. *Journal of Vascular Surgery*, *40*(6), 1112–1117. https://doi.org/10.1016/j.jvs.2004.09.028

25. Deci, E. L., Eghrari, H., Patrick, B. C., & Leone, D. R. (1994). Facilitating Internalization: The Self-Determination Theory Perspective. *Journal of Personality*, *62*(1), 119–142. https://doi.org/10.1111/j.1467-6494.1994.tb00797.x
26. Denzin, N. K. (1989). *The research act: A theoretical introduction to sociological methods* (3rd ed). Prentice Hall; WorldCat.
27. Dickes, A. C., Kamarainen, A., Metcalf, S. J., Gün-Yildiz, S., Brennan, K., Grotzer, T., & Dede, C. (2019). Scaffolding ecosystems science practice by blending immersive environments and computational modeling. *British Journal of Educational Technology*, *50*(5), 2181–2202. https://doi.org/10.1111/bjet.12806
28. Dodick, J., & Orion, N. (2003). Measuring student understanding of geological time. *Science Education*, *87*(5), 708–731. https://doi.org/10.1002/sce.1057
29. Eckhardt, M., Urhahne, D., Conrad, O., & Harms, U. (2013). How effective is instructional support for learning with computer simulations? *Instructional Science*, *41*(1), 105–124. https://doi.org/10.1007/s11251-012-9220-y
30. 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. https://doi.org/10.3390/educsci8030094
31. Ekstrom, R. B., French, J. W., & Harman, H. H. (1979). Cognitive factors: Their identification and replication. *Multivariate Behavioral Research Monographs*, *79*(2), 84–84.
32. El Tegani, M., & Wegner, C. (2025). Simulations in science education. A systematic literature review. *Journal of Innovation in Psychology, Education and Didactics*, *29*(2).
33. Engeln, K. (2004). *Schülerlabors. Authentische, aktivierende Lernumgebungen als Möglichkeit, Interesse an Naturwissenschaften und Technik zu wecken.* Logos.
34. Enger, S. K., & Yager, R. E. (1998). *The Iowa Assessment Handbook.* Iowa Univ., Iowa City. Science Education Center.
35. Enger, S. K., & Yager, R. E. (2001). *Assessing student understanding in science: A standards-based K-12 handbook*. Corwin Press.
36. Erşahan, O. (2016). *Yapılandırmacı yaklaşımı temel alan etkileşimli video öğretim yönteminin 7. Sınıf öğrencilerinin iş ve enerji konusu ile ilgili bilişsel ve duyuşsal öğrenmelerine etkisi* [Gazi-Universität]. https://tez.yok.gov.tr/UlusalTezMerkezi/tezDetay.jsp?id=aJ1hQfYG88k5bjLLiDkN5g&no=-U8c4poCIZ-JVMyIixUinw
37. Eskrootchi, R., & Oskrochi, G. R. (2010). A Study of the Efficacy of Project-based Learning Integrated with Computer-based Simulation—STELLA. *Journal of Educational Technology & Society*, *13*, 236–245.
38. Frey, A., Taskinen, P., Schütte, K., Prenzel, M., Artelt, C., Baumert, J., Blum, W., Hammann, M., Klieme, E., Pekrun, R., Carstensen, C. H., Asseburg, R., & Drechsel, B. (Eds.). (2009). *PISA 2006 Skalenhandbuch. Dokumentation der Erhebungsinstrumente.* Waxmann.
39. Garneli, V., & Chorianopoulos, K. (2018). Programming video games and simulations in science education: Exploring computational thinking through code analysis. *Interactive Learning Environments*, *26*(3), 386–401. https://doi.org/10.1080/10494820.2017.1337036
40. Gelbart, H., Brill, G., & Yarden, A. (2009). The Impact of a Web-Based Research Simulation in Bioinformatics on Students' Understanding of Genetics. *Research in Science Education*, *39*(5), 725–751. https://doi.org/10.1007/s11165-008-9101-1
41. Gelbart, H., & Yarden, A. (2006). Learning genetics through an authentic research simulation in bioinformatics. *Journal of Biological Education*, *40*(3), 107–112. https://doi.org/10.1080/00219266.2006.9656026
42. Gibson, J. J. (1979). *The ecological approach to visual perception*. Houghton Mifflin.
43. Han, I., & Black, J. B. (2011). Incorporating haptic feedback in simulation for learning physics. *Computers & Education*, *57*(4), 2281–2290. https://doi.org/10.1016/j.compedu.2011.06.012

44. Heller, K. A., & Perleth, C. (2000). *Kognitiver Fähigkeitstest für 4.-12. Klassen, Revision (KFT 4-12+ R)*. Beltz Testgesellschaft.
45. Henderson, L., Klemes, J., & Eshet, Y. (2000). Just Playing a Game? Educational Simulation Software and Cognitive Outcomes. *Journal of Educational Computing Research*, *22*(1), 105–129. https://doi.org/10.2190/EPJT-AHYQ-1LAJ-U8WK
46. Hestenes, D., & Wells, M. (1992). A mechanics baseline test. *The Physics Teacher*, *30*(3), 159–166. https://doi.org/10.1119/1.2343498
47. Honey, M. A., & Hilton, M. (Eds.). (2011). *Learning Science through Computer Games and Simulations*. National Academies Press. 500 Fifth Street NW, Washington, DC 20001. Tel: 888-624-8373 and Tel: 202-334-2000 and Fax: 202-334-2793 and e-mail: Customer_Service@nap.edu and Web site: http://www.nap.edu. http://www.nap.edu/catalog.php?record_id=13078
48. Huber, L., Hellmer, J., & Schneider, F. (Eds.). (2013). *Forschendes Lernen im Studium. Aktuelle Konzepte und Erfahrungen. 2. Aufl.* Univ.-Verl. Webler.
49. Ito, M., Gutiérrez, K., Livingstone, S., Penuel, B., Rhodes, J., Salen, K., Schor, J., Sefton-Green, J., & Watkins, S. C. (2013). *Connected Learning*. BookBaby.
50. Jaakkola, T., Nurmi, S., & Veermans, K. (2011). A comparison of students' conceptual understanding of electric circuits in simulation only and simulation-laboratory contexts. *Journal of Research in Science Teaching*, *48*(1), 71–93. https://doi.org/10.1002/tea.20386
51. Jebeile, J. (2017). Computer Simulation, Experiment, and Novelty. *International Studies in the Philosophy of Science*, *31*(4), 379–395. https://doi.org/10.1080/02698595.2019.1565205
52. Karlsson, G., Ivarsson, J., & Lindström, B. (2013). Agreed discoveries: Students' negotiations in a virtual laboratory experiment. *Instructional Science*, *41*(3), 455–480. https://doi.org/10.1007/s11251-012-9238-1
53. Kärrqvist, C. (1985). The development of concepts by means of dialogues centered on experiments. *Aspects of Understanding Electricity*, 215–226.
54. Kollar, I., Wecker, C., Langer, S., & Fischer, F. (2011). Webbasiertes Forschendes Lernen im naturwissenschaftlichen Unterricht. *Psychologie in Erziehung und Unterricht*, *58*(4). https://doi.org/10.2378/peu2011.art18d
55. Kramer, N. (2025). *Sport-Bio-logisch! Konzeption und Evaluation fächerübergreifenden Unterrichts in der Fächerkombination Biologie und Sport* [Application/pdf]. 24445075 bytes. https://doi.org/10.4119/UNIBI/3004852
56. Krüger, J. T., Höffler, T. N., Wahl, M., Knickmeier, K., & Parchmann, I. (2022). Two comparative studies of computer simulations and experiments as learning tools in school and out-of-school education. *Instructional Science*, *50*(2), 169–197. https://doi.org/10.1007/s11251-021-09566-1
57. Landriscina, F. (2009). Simulation and learning: The role of mental models. *Journal of E-Learning and Knowledge Society*, *Vol 5*, No 2 (2009): Journal of eLearning and Knowledge Society. https://doi.org/10.20368/1971-8829/316
58. Lent, R. W., Brown, S. D., & Hackett, G. (1994). Toward a Unifying Social Cognitive Theory of Career and Academic Interest, Choice, and Performance. *Journal of Vocational Behavior*, *45*(1), 79–122. https://doi.org/10.1006/jvbe.1994.1027
59. Lent, R. W., Lopez, A. M., Lopez, F. G., & Sheu, H.-B. (2008). Social cognitive career theory and the prediction of interests and choice goals in the computing disciplines. *Journal of Vocational Behavior*, *73*(1), 52–62. https://doi.org/10.1016/j.jvb.2008.01.002
60. Li, M., Donnelly-Hermosillo, D. F., & Click, J. (2022). Comparing Simulation Sequencing in a Chemistry Online-Supported Project-Based Learning Unit. *Journal of Science Education and Technology*, *31*(1), 27–51. https://doi.org/10.1007/s10956-021-09929-w
61. Lin, L.-F., Hsu, Y.-S., & Yeh, Y.-F. (2012). The Role of Computer Simulation in an Inquiry-Based Learning Environment: Reconstructing Geological Events as Geologists. *Journal of*

Science Education and Technology, *21*(3), 370–383. https://doi.org/10.1007/s10956-011-9330-3
62. Linn, M. C., & Eylon, B.-S. (2011). *Science Learning and Instruction* (0 ed.). Routledge. https://doi.org/10.4324/9780203806524
63. Liu, H., & Su, I. (2011). Learning residential electrical wiring through computer simulation: The impact of computer-based learning environments on student achievement and cognitive load. *British Journal of Educational Technology*, *42*(4), 598–607. https://doi.org/10.1111/j.1467-8535.2009.01047.x
64. Liu, H.-C., & Chuang, H.-H. (2011). Investigation of the impact of two verbal instruction formats and prior knowledge on student learning in a simulation-based learning environment. *Interactive Learning Environments*, *19*(4), 433–446. https://doi.org/10.1080/10494820903356940
65. López, V., & Pintó, R. (2017). Identifying secondary-school students' difficulties when reading visual representations displayed in physics simulations. *International Journal of Science Education*, *39*(10), 1353–1380. https://doi.org/10.1080/09500693.2017.1332441
66. Magana, A. J., Elluri, S., Dasgupta, C., Seah, Y. Y., Madamanchi, A., & Boutin, M. (2019). The Role of Simulation-Enabled Design Learning Experiences on Middle School Students' Self-generated Inherence Heuristics. *Journal of Science Education and Technology*, *28*(4), 382–398. https://doi.org/10.1007/s10956-019-09775-x
67. 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. https://doi.org/10.1111/bjet.12954
68. McDermott, L. C., & Shaffer, P. S. (1992). Research as a guide for curriculum development: An example from introductory electricity. Part I: Investigation of student understanding. *American Journal of Physics*, *60*(11), 994–1003. https://doi.org/10.1119/1.17003
69. Mešić, V., Dervić, D., Gazibegović-Busuladžić, A., Salibašić, D., & Erceg, N. (2015). Comparing the Impact of Dynamic and Static Media on Students' Learning of One-Dimensional Kinematics. *EURASIA Journal of Mathematics, Science and Technology Education*, *11*(5). https://doi.org/10.12973/eurasia.2015.1385a
70. Miles, M. B., Huberman, A. M., & Saldaña, J. (2013). *Qualitative data analysis: A methods sourcebook* (Edition 3). Sage.
71. Ndihokubwayo, K., Uwamahoro, J., & Ndayambaje, I. (2020). Effectiveness of PhET Simulations and YouTube Videos to Improve the Learning of Optics in Rwandan Secondary Schools. *African Journal of Research in Mathematics, Science and Technology Education*, *24*(2), 253–265. https://doi.org/10.1080/18117295.2020.1818042
72. Ng, D. T. K., & Chu, S. K. W. (2021). Motivating Students to Learn STEM via Engaging Flight Simulation Activities. *Journal of Science Education and Technology*, *30*(5), 608–629. https://doi.org/10.1007/s10956-021-09907-2
73. Niemivirta, M. (2002). Motivation And Performance In Context: The Influence Of Goal Orientations And Instructional Setting On Situational Appraisals And Task Performance. *Psychologia*, *45*(4), 250–270. https://doi.org/10.2117/psysoc.2002.250
74. Olakanmi, E. E. (2015). The Effects Of A Web-Based Computer Simulation On Students' Conceptual Understanding Of Rate Of Reaction And Attitude Towards Chemistry. *Journal of Baltic Science Education*, *14*(5), 627–640. https://doi.org/10.33225/jbse/15.14.627
75. Osborne, R. (1983). Towards modifying children's ideas about electric current. *Research in Science & Technological Education*, *1*(1), 73–82.
76. Paas, F. G. W. C., & Van Merriënboer, J. J. G. (1993). The Efficiency of Instructional Conditions: An Approach to Combine Mental Effort and Performance Measures. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *35*(4), 737–743. https://doi.org/10.1177/001872089303500412

77. Paas, F. G. W. C., & Van Merriënboer, J. J. G. (1994). Instructional control of cognitive load in the training of complex cognitive tasks. *Educational Psychology Review*, *6*(4), 351–371. https://doi.org/10.1007/BF02213420
78. Paas, F. G. W. C., Van Merriënboer, J. J. G., & Adam, J. J. (1994). Measurement of Cognitive Load in Instructional Research. *Perceptual and Motor Skills*, *79*(1), 419–430. https://doi.org/10.2466/pms.1994.79.1.419
79. Paas, F., Renkl, A., & Sweller, J. (Eds.). (2003a). *Cognitive Load Theory* (0 ed.). Routledge. https://doi.org/10.4324/9780203764770
80. Paas, F., Renkl, A., & Sweller, J. (2003b). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist*, *38*(1), 1–4.
81. Pintrich, P., Smith, D., Duncan, T., & Mckeachie, W. (1991). A Manual for the Use of the Motivated Strategies for Learning Questionnaire (MSLQ). *Ann Arbor. Michigan*, *48109*, 1259.
82. Price, N. T. (2013). Teaching Strategies for Using Projected Images to Develop Conceptual Understanding: Exploring Discussion Practices in Computer Simulation and Static Image-Based Lessons. *Open Access Dissertations*, *762*. https://doi.org/10.7275/WC4G-5J50
83. Pucholt, Z. (2021). Effectiveness of simulations versus traditional approach in teaching physics. *European Journal of Physics*, *42*(1), 015703. https://doi.org/10.1088/1361-6404/abb4ba
84. Pugh, K. J., Linnenbrink-Garcia, L., Koskey, K. L. K., Stewart, V. C., & Manzey, C. (2010). Motivation, learning, and transformative experience: A study of deep engagement in science. *Science Education*, *94*(1), 1–28. https://doi.org/10.1002/sce.20344
85. Reilly, J. M., McGivney, E., Dede, C., & Grotzer, T. (2021). Assessing Science Identity Exploration in Immersive Virtual Environments: A Mixed Methods Approach. *The Journal of Experimental Education*, *89*(3), 468–489. https://doi.org/10.1080/00220973.2020.1712313
86. Rosenberg, J. M., & Lawson, M. A. (2019). An Investigation of Students' Use of a Computational Science Simulation in an Online High School Physics Class. *Education Sciences*, *9*(1), 49. https://doi.org/10.3390/educsci9010049
87. Rubin, H., & Rubin, I. (2005). *Qualitative Interviewing (2nd ed.): The Art of Hearing Data*. SAGE Publications, Inc. https://doi.org/10.4135/9781452226651
88. Sabagh, H. A. el-Aziz el-. (2011). *The impact of a web-based virtual lab on the development of students's conceptual understanding and science process skills* [Universität Dresden]. http://d-nb.info/1013549376/34
89. Samavedam, S. (2020). Simulation Benefits Both the Teacher and the Taught. *Indian Journal of Critical Care Medicine*, *24*(6), 373–374. https://doi.org/10.5005/jp-journals-10071-23472
90. Sawada, D., Piburn, M. D., Judson, E., Turley, J., Falconer, K., Benford, R., & Bloom, I. (2002). Measuring Reform Practices in Science and Mathematics Classrooms: The Reformed Teaching Observation Protocol. *School Science and Mathematics*, *102*(6), 245–253. https://doi.org/10.1111/j.1949-8594.2002.tb17883.x
91. Scardamalia, M. (2002). Collective cognitive responsibility for the advancement of knowledge. *Liberal Education in a Knowledge Society*, *97*, 67–98.
92. Schäfers, M. S., Schmiedebach, M., & Wegner, C. (2020). Virtuelle Labore im Biologieunterricht: Auswirkungen von Labster auf die Selbsteinschätzung von Schülerinnen und Schülern. *MedienPädagogik: Zeitschrift für Theorie und Praxis der Medienbildung*, 140–167. https://doi.org/10.21240/mpaed/00/2020.08.31.X
93. Shegog, R., Lazarus, M. M., Murray, N. G., Diamond, P. M., Sessions, N., & Zsigmond, E. (2012). Virtual Transgenics: Using a Molecular Biology Simulation to Impact Student Academic Achievement and Attitudes. *Research in Science Education*, *42*(5), 875–890. https://doi.org/10.1007/s11165-011-9216-7
94. Shegog, R., Markham, C., Peskin, M., Dancel, M., Coton, C., & Tortolero, S. (2007). "It's Your Game": An Innovative Multimedia Virtual World to Prevent HIV/STI and Pregnancy in Middle School Youth. In *MEDINFO 2007* (pp. 983–987). IOS Press.

https://ebooks.iospress.nl/volumearticle/11123?_gl=1*14vynqk*_up*MQ..*_ga*MzA2MDQ5MDA0LjE3MTYzODU0ODE.*_ga_6N3Q0141SM*MTcxNjM4NTQ4MC4xLjEuMTcxNjM4NTU2NC4wLjAuMA..

95. Shen, J. (2011). *A transformative modeling perspective on the use of manipulative materials for conceptual learning in science education*. https://www.researchgate.net/profile/Ji-Shen-3/publication/284199866_A_Transformative_Modeling_Perspective_on_the_Use_of_Manipulative_Materials_for_Conceptual_Learning_in_Science_Education/links/564fcde208aeafc2aab4060c/A-Transformative-Modeling-Perspective-on-the-Use-of-Manipulative-Materials-for-Conceptual-Learning-in-Science-Education.pdf
96. Shipstone, D. (1984). A study of children's understanding of electricity in simple DC circuits. *European Journal of Science Education*, *6*(2), 185–198. https://doi.org/10.1080/0140528840060208
97. Shkedi, A. (2005). *Multiple Case Narrative*. John Benjamins Publishing Company. http://digital.casalini.it/9789027293893
98. Slater, T. F., Ryan, J. M., & Samson, S. L. (1997). Impact and dynamics of portfolio assessment and traditional assessment in a college physics course. *Journal of Research in Science Teaching*, *34*(3), 255–271. https://doi.org/10.1002/(SICI)1098-2736(199703)34:3%253C255::AID-TEA4%253E3.0.CO;2-R
99. Stinken-Rösner, L. (2020). Simulations in Science Education – Status Quo. *Progress in Science Education (PriSE)*, 26-34 Pages. https://doi.org/10.25321/PRISE.2020.996
100. Sweller, J. (1988). Cognitive Load During Problem Solving: Effects on Learning. Cognitive Science, 12(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4
101. Tapola, A., Veermans, M., & Niemivirta, M. (2013). Predictors and outcomes of situational interest during a science learning task. *Instructional Science*, *41*(6), 1047–1064. https://doi.org/10.1007/s11251-013-9273-6
102. Thisgaard, M., & Makransky, G. (2017). Virtual Learning Simulations in High School: Effects on Cognitive and Non-cognitive Outcomes and Implications on the Development of STEM Academic and Career Choice. *Frontiers in Psychology*, *8*, 805. https://doi.org/10.3389/fpsyg.2017.00805
103. Thornton, R. K., & Sokoloff, D. R. (1998). Assessing student learning of Newton's laws: The Force and Motion Conceptual Evaluation and the Evaluation of Active Learning Laboratory and Lecture Curricula. *American Journal of Physics*, *66*(4), 338–352. https://doi.org/10.1119/1.18863
104. Urhahne, D. (2002). *Motivation und Verstehen: Studien zum computergestützten Lernen in den Naturwissenschaften*. Zugl.: Kiel, Univ., Diss., 2001.
105. Urhahne, D., Nick, S., & Schanze, S. (2009). The Effect of Three-Dimensional Simulations on the Understanding of Chemical Structures and Their Properties. *Research in Science Education*, *39*(4), 495–513. https://doi.org/10.1007/s11165-008-9091-z
106. VBE. (2023). *Licht und Schatten im digitalen Klassenzimmer* [Pressebericht]. Verband Bildung und Erziehung. https://www.vbe.de/presse/pressedienste/pressedienste-2023/licht-und-schatten-im-digitalen-klassenzimmer
107. Wastiau, P., Blamire, R., Kearney, C., Quittre, V., Van De Gaer, E., & Monseur, C. (2013). The Use of ICT in Education: A survey of schools in Europe. *European Journal of Education*, *48*(1), 11–27. https://doi.org/10.1111/ejed.12020
108. Wen, C.-T., Chang, C.-J., Chang, M.-H., Fan Chiang, S.-H., Liu, C.-C., Hwang, F.-K., & Tsai, C.-C. (2018). The learning analytics of model-based learning facilitated by a problem-solving simulation game. *Instructional Science*, *46*(6), 847–867. https://doi.org/10.1007/s11251-018-9461-5

109. Widiyatmoko, A. (2018). The Effectiveness of Simulation in Science Learning on Conceptual Understanding: A Literature Review. *Journal of International Development and Cooperation*, *24*(1 & 2), 35–43.
110. Yoon, S. A., Anderson, E., Park, M., Elinich, K., & Lin, J. (2018). How augmented reality, textual, and collaborative scaffolds work synergistically to improve learning in a science museum. *Research in Science & Technological Education*, *36*(3), 261–281. https://doi.org/10.1080/02635143.2017.1386645
111. Yoon, S. A., & Wang, J. (2014). Making the Invisible Visible in Science Museums Through Augmented Reality Devices. *TechTrends*, *58*(1), 49–55. https://doi.org/10.1007/s11528-013-0720-7
112. Zauner, M., & Schrempf, A. (2009). *Informatik in der Medizintechnik*. Springer Vienna. https://doi.org/10.1007/978-3-211-89189-6
113. Zendler, A., & Greiner, H. (2020). The effect of two instructional methods on learning outcomes in chemistry education: The experiment method and computer simulation. *Education for Chemical Engineers*, *30*, 9–19. https://doi.org/10.1016/j.ece.2019.09.001
114. Zwingenberger, A. (2009). *Wirksamkeit multimedialer Lernmaterialien: Kritische Bestandsaufnahme und Metaanalyse empirischer Evaluationsstudien*. Waxmann.

Appendix

Authors, year	Study design	Key findings / (selected) results
Krüger et al. (2022) Study I	**N:** 443 (group 2 n = 222, group 2 n = 221) from 19 secondary schools **uV:** Treatment (simulation, hands-on real experiments) Group 1: Simulation Group 2: Experiment **aV:** Expertise, cognitive load, interest **Method**: Pre- & post-test **Instruments used:** Personal data (age, gender, school grades in STEM subjects), individual interest in biology & chemistry (according to Frey et al., 2009, α = .917 - .919), prior knowledge (own questionnaire), cognitive load (according to Paas et al., 2003a, 2003b, α = .595), situational interest (according to Engeln, 2004, emotional α = .825; epistemic α = .621; value-related α = .555) **Duration:** 90 min **Country:** Germany **Grade level:** 10 - 13 **Specialized context:** Ocean acidification **Type of school:** secondary school	The study focused on the influence of simulation-based and real-experiment-based learning at school. The pupils were divided into two groups. While one group only learnt with simulations during the intervention, the other group worked exclusively with real-life experiments. In the case of the students, higher values were achieved by the simulation in both *subject knowledge* (d = 0.19) and *cognitive load* (d = 0.19) than was the case in the group with the experiments. At the same time, the level of *epistemic* (d = 0.33) and *value-related* (d = 0.35) components of *situational interest* was lower in the simulation group than in the experimental group. No difference was found between the groups with regard *to emotional interest*. Age, *gender*, *school type* and *grade level* also had no significant influence on the results.
Krüger et al. (2022 Study II	**N1:** 367 (group 1 n = 169, group 2 n = 198) **uV:** Treatment (simulation / real experiment) Group 1: Simulation Group 2: Real experiments **aV:** Expertise, *cognitive load*, interest **Method**: Pre- & post-test **Instruments used:** Personal date (age, gender, school grades in STEM subjects), individual interest in biology & chemistry (according to Frey et al., 2009, α = .90 - .91), prior knowledge (own questionnaire), *cognitive load* (according to Paas et al., 2003a, 2003b, α = .586), situational interest (according to Engeln, 2004, emotional α = .801; epistemic α = .748; value-related α = .763) **Duration:** 180 min **Country:** Germany **Grade level:** 10 - 13 **Specialist context:** Warming, salinization, acidification of the Baltic Sea, interactions and challenges at the ecological and social level **Type of school:** secondary schools	This second study also focused on the influence of simulation-based and real-life experiment-based learning. This time, however, the study was conducted in a learning lab. The students were divided into two groups. While one group only worked with simulations during the intervention, the other group worked exclusively with real-life experiments. The same test instruments were used as in Study I. Here, too, it was shown that the students achieved significantly higher values in both *specialized knowledge* (d = 1.05) and *cognitive load* (d = 0.21) with the simulation than the group with the experiments. In this study, the level of *epistemic* (d = 0.54) and *value-related* (d = 0.35) components of *situational interest* was also lower in the simulation group than in the experimental group. No difference was found between the groups concerning *emotional interest* either. Age, *gender*, *school type,* and *grade level* also had no significant influence on the results.
Li et al. (2022)	**N:** 72 (group 1 n = 37, group 2 n = 35) **uV:** Sequence of simulation and project-based work (design, construction, testing) Group 1 (planning): Simulation; project-based work Group 2 (reflection): project-based work; simulation **aV:** learning gain, affordances (facilitations), and restrictions **Method**: Research question 1: quasi-experimental two-group pre-post test Research question 2: qualitative design, questionnaire, interviews, video recordings	In the following study, 72 students were analyzed to determine the extent to which the sequence of learning content and media affects *learning gains*. To this end, the order of the learning content was adapted and its impact analyzed. To do this, the students worked with the WISE tool, which contains interactive visualizations, collaboration tools and assessment systems for various science curricula. At the same time, WISE logged the student data during the interaction with the curriculum. It was shown that the combination of simulation and project-based work led to significant

	Instruments used: Web-Base Inquiry Science Environment-WISE (according to Linn & Eylon 2011), pre-post test (α = .782) Pre-post test (α_{IR}= 88.9% to 100%; Cohen's κ = .702 - 1.000) Embedded items (α_{IR}= 86.1% to 98.6%; Cohen's κ = .618 to .974) Qualitative analysis: Thematic analysis (according to Braun & Clarke, 2006) Topic identification through affordance frameworks (according to Gibson, 1979) **Duration:** seven units over two weeks **Grade level:** 10 **Subject context:** chemistry, physics, concentration, desalination **School type:** High School	*improvements in performance* in both groups (planning: d = 1.12, reflection: d = 0.97), but no significant differences between the order (1st simulation & 2nd project work / 1st project work & 2nd simulation) could be demonstrated (F(1,70) = 0.962, p = .330). The qualitative analysis showed that the simulation facilitated the transfer of theory into practice and, at the same time, enabled meaningful learning. In addition, pen and paper tasks were perceived as more of a hindrance. However, some students preferred pen and paper and perceived the technology as less accessible and more complicated.
Ng & Chu (2021)	**N:** 345 (female n = 167; male n = 178) from 8 primary schools **uV:** treatment (flight simulation), intrinsic motivation, extrinsic motivation, self-efficacy and peer support **aV:** (Flight) performance **Method**: Questionnaire survey pre- and posttest **Instruments used:** ASMQ (intrinsic motivation [α = .93], extrinsic motivation [α = .84], self-efficacy [α = .89] and peer support [α = .91]) **Duration:** 1.5 hours in three phases **Country:** Hong Kong **Grade:** 4 (n = 117), 5 (n = 106) and 6 (n = 122) **Subject context:** physics, technology, flight simulation **School type:** Primary school,	In this study by Ng & Chu, the researchers investigated the correlation between motivation, grade level, gender and children's flight performance during simulation-based learning. The flight simulation was divided into three phases (control, take-off, landing). They found that the greatest motivator for the pupils was peer support (M = 3.44). In addition, positive correlations were found between *intrinsic motivation* (r = 0.39, p < .001), *extrinsic motivation* (r = 0.30, p < .001), *peer support* (r = 0.36, p < .001) and *self-efficacy* (r = 0.26, p < .001) in relation to flight performance. In addition, the *motivational constructs* were found to be significantly associated with student *performance* (F = 72.05, p < .001) and *gender* (F = 21.48, p < .001), but no significant interaction effects between gender and performance were found. These results indicate that higher *motivation* is associated with better *flight performance*. In contrast, there were no significant differences between the different Grade levels.
Pucholt (2021) Sub-study I	**N:** 29 (group 1 n = 14, group 2 n = 15) **uV:** Treatment CG: classic experimental teaching and blackboard EG: simulation-based lessons, projector and blackboard Group 1: CG Group 2: EG **aV:** Memory, misconceptions, percentage progress **Method**: Pre-post follow-up test, questionnaire with closed questions and math problems **Instruments used:** own questionnaire, Force Concept Inventory-FCI test **Duration:** 2 x 45 min **Country:** Czech Republic **Age:** 15 - 16 **Subject context:** Physics, Newtonian motion **School type:** Grammar school **Place of learning:** School	The study by Pucholt (2021), divided into three sub-studies, investigated whether the increase in performance - *measured by percentage progress* - of pupils in relation to certain physics topics through purely simulation-based lessons differs significantly from lessons that are purely experimental in structure. The sub-study designs differed in their respective structure and subject content. The extent to which existing misconceptions in the respective groups could be eliminated beforehand was also investigated. In sub-study I, the students first completed a pre-test before being divided into two groups. While group 1 initially received analogue instruction, group two was taught using simulation. After the first unit, the two groups swapped rooms so that group 1 now received simulation-based teaching and group 2 analogue teaching. After the second unit, both groups completed a post-test. The students performed significantly better in the blocks in which they were assigned to the

		EG- treatment than those from the other group (Block 1: p = .0361, U = 305, Block 2: p = .0385, U = 304).
Pucholt (2021) Sub-study II	**N:** 29 (group 1 n = 14, group 2 n = 15) **uV:** Treatment CG: classic experimental teaching and blackboard EG: simulation-based lessons as well as projector and blackboard Group 1: CG Group 2: EG **aV:** Memory, misconceptions, percentage progress (knowledge test) **Method**: Pre-post follow-up test, questionnaire with closed questions and math problems **Instruments used:** own questionnaire, Force Concept Inventory-FCI-Test **Duration:** 2 x 45 min **Country:** Czech Republic **Age:** 15 - 16 **Subject context:** Physics, Newtonian motion **School type:** Grammar school **Place of learning:** School	In this sub-study, the students completed a pre-test before being divided into two groups. While one group received traditional analogue instruction, the other group received simulation-based instruction. After the lessons, the students completed a post-test, which was followed by another follow-up test two months later to test *their memorization skills*. While the students in group 2 performed significantly better in the first post-test (p = .025, U = 150), no significant differences could be identified in the follow-up test (p = .693, U = 94).
Pucholt (2021) Sub-study III	**N:** 27 (group 1 n = 14, group 2 n = 13) **uV:** Treatment CG: classic experimental teaching and blackboard EG: simulation-based teaching and projector and blackboard Group 1: CG Group 2: EG **aV:** misconceptions, percentage progress **Method**: Pre-post follow-up test, questionnaire with closed questions and math problems **Instruments used:** Own questionnaire, Force Concept Inventory-FCI test **Duration:** 2 x 45 min **Country:** Czech Republic **Age:** 15 - 16 **Subject context:** Physics, motion in the gravitational field **School type:** Grammar school	In this sub-study, the pupils completed a pre-test before being divided into two groups. They then underwent two teaching units. While one group received traditional analogue instruction, the other group received simulation-based instruction. Group 2 achieved significantly higher percentage progress than group 1 (p = .046, U = 126.5).
Reilly et al (2021)	**N:** 126 (minus dropout n = 103) **uV:** Treatment (simulation-based teaching) **aV:** scientific identity, interest, ecosystem science content, understanding of causality, correlation versus causality, use of experimental methods, epistemology of science **Method**: Mixed method, pre-post survey, interviews **Instruments used:** Pre-post test based on the Likert scale: self-efficacy (according to Chen et al., 2014, α = .89), self-perception and scientific identity (according to Pugh et al., 2010, α = .94) and interest (according to Deci et al., 1994, α = .90) Qualitative: iterative, hybrid, emic and etic thematic coding method (after Creswell, 2012; Rubin and Rubin, 2005) **Duration:** 13 x 50 min **Country:** USA **Grade level:** 7 **Age:** 12 - 13 **Subject context:** Ecology, Ecosciences	Three schools, each with two seventh-grade classes, were analyzed as part of the study. The students, aged 12 and 13, were divided into groups of two to three per group, resulting in a total of 63 groups. The study aimed to investigate 1) whether the specially developed EcoXPT curriculum promotes the exploration and change of scientific identity, 2) how the students describe their own scientific identity and what it means to them to be a science person and 3) how the scientific identity, the shift in interests and the change in self-efficacy are reflected in the students' descriptions and whether the qualitatively and quantitatively collected data contradict each other. The EcoXPT tool is the further development of one of the tools ecoMuve, in which users can investigate an ecosystem in the virtual world with guidance. exoXPT complements this environment with experimental tools based on real-world techniques used by ecosystem scientists.

	School type: Secondary school **Place of learning:** School	While the intervention significantly increased *self-efficacy* (6.1%, $t(92) = 3.44$, $p < .001$, $d = 0.33$), the sub-constructs *self-perception* and *interest* showed no significance. Interest even decreased slightly. The qualitative surveys indicate that the students internalized ecological science content through the unit. In addition, the curriculum helped to "demystify" the working methods and thinking of scientists. The students also identified more strongly with science.
Cayvaz et al (2020)	**N:** 188 (Group 1 n = 98, Group 2 n = 90) **uV:** Treatment Group 1: Simulation Group 2: Book **aV:** Knowledge, scientific ways of thinking and working, attitude towards natural sciences **Method**: Pre-post test **Instruments used:** Multiple-choice Achievement Test (according to Erşahan, 2016, α = .71), Inquiry Skills Test (according to Enger & Yager, 1998, α = .72) Science Attitude Scale (according to Enger & Yager, 2001, α = .70) **Duration:** 3 weeks (4 hours per week) **Country:** Turkey **Grade level:** 7 Age: 13 - 14 **Subject context:** Labour and energy **School type:** Middle School	This study investigated the effect of the type of instruction (simulation vs. book) on *knowledge, scientific thinking and working methods* and *attitudes towards science*. The results of this study show that the students in the simulation group significantly improved their skills in terms of *scientific thinking and working methods* ($t(97) = 3.812$, $p < .001$), while the control group, which worked with textbooks, did not. In terms of *acquired knowledge*, the subjects in the simulation group increased their performance significantly more than those in the control group ($t(186) = 2.148$, $p < .001$). The *attitude towards natural sciences* did not change significantly in either group. There was also no significance between the groups.
Makransky et al (2020) Study I	**N:** 99 (female n = 47, male n = 52) **uV:** Treatment (simulation), demographic characteristics (age, gender) **aV:** Interest, self-efficacy, scientific ambition **Method**: Pre-post test **Duration:** 1 week **Instruments used:** Interest (according to Thisgaard & Makransky, 2017, α = .77, 0.78), self-efficacy (according to Pintrich et al., 1991, α = .85, 0.91), scientific ambition (according to Thisgaard & Makransky, 2017, α = .92, .93) **Country:** Denmark **Grade level:** 7th and 8th grade **Age:** 13 - 16 **Subject context:** Biology, safety briefing laboratory **School type:** Middle school	In this study, 49 seventh graders and 50 eighth-graders, 52 boys and 47 girls, were used to investigate the extent to which an immersive virtual simulation affects *interest, self-efficacy* and *scientific ambitions*. In particular, gender-specific differences were to be analyzed. The study found that *interest* in laboratory work and safety was significantly increased by the intervention ($F(1,97) = 6.354$, $p = .013$, $\eta^2 = 0.061$). However, gender-specific differences were not significant. *Self-efficacy* was also significantly increased ($F(1,97) = 36.321$, $p < .001$, $\eta^2 = 0.272$). Here, too, there was no significant difference between the genders. Although a higher *scientific ambition* could be measured, this was not significant. While, as mentioned, there was no significant interaction in relation to *gender* for *interest* and *self-efficacy*, this was significant for *scientific ambition* ($F(1,97) = 4.015$, $p = .048$, $\eta^2 = 0.040$). While the intervention had no significant influence on the male participants in the study, the intervention significantly increased the *scientific ambition* of the female students ($t(46) = 2.349$, $p = .023$, $d = 0.30$).
Makransky et al (2020) Study II	**N:** 131 (male n = 47, female n = 84; simulation n = 64, video n = 67) **uV:** Demographic characteristics (age, gender), treatment (IVR-CSI simulation, video) Group 1: IVR-SCI simulation Group 2: Video **aV:** interest, self-efficacy, scientific ambition, expectation of results (social, physical, self) **Method**: Pre-post test	The follow-up study by Makransky et al., (2020) involved 77 second-year and 54 third-year high school students. A total of 47 male and 84 female test subjects took part. This time, in addition to an intervention group, a control group that watched a video was added. While group 1 went through an immersive virtual simulation, group 2 watched a film in which the simulation was played out perfectly.

	Instruments used: Interest (after Thisgaard & Makransky, 2017, α = .65, 0.74), self-efficacy (after Pintrich et al., 1991, α = .80, 0.85), academic ambition (after Thisgaard & Makransky, 2017, α = .82, 0.83), outcome expectancy (after Thisgaard & Makransky, 20017; Bandura, 1986; Lent et al., 1994, social (α = .63), physical (α = .71), self (α = .77)) **Duration:** 12 minutes 3 times a week **Country:** Denmark **Grade level:** 3rd and 4th high school **Age:** 17 - 20 **Subject context:** Biology, DNA analysis, PCR experiment **School type:** High school	No significant interactions were found between *gender* and the dependent variables. With regard to *interest*, significant increases were found in the simulation group compared to the pre-test ($t(63) = 4.644$, $p < .001$, $d = 0.35$). This was not the case for the video group. Significant increases in *self-efficacy* were found in both groups (group 1: $t(63) = 3.819$, $p < .001$, $d = 0.30$, group 2: $t(66) = 3.637$, $p = .001$, $d = 0.20$). No significant differences were found between the two treatments in this respect. Both groups showed significant differences in the pre-post comparison for *physical outcome expectations* (group 1: $t(63) = 2.815$, $p = .001$, $d = 0.25$, group 2: $t(66) = 2.217$, $p = .030$, $d = 0.15$), but no significant differences were found between the two treatments. With regard to *social outcome expectancy*, the two groups differed significantly in favor of the simulation group ($F(1,129) = 4.312$, $p = .040$, $\eta^2 = 0.032$). The differences in *self-expectation of results* were not significant at $p = .053$. Similarly, there was no significant difference in *scientific ambition* between the test subjects in both groups ($F(1,129) = 1.458$, $p = .229$, $\eta^2 = 0.011$)
Ndihokubwayo et. al. (2020)	**N:** 136 (group 1 n = 45, group 2 n = 46, group 3 n = 45) **uV:** Treatment (simulation, video, classic) Group 1: PhET simulation Group 2: YouTube videos Group 3: Conventional teaching (analogue) **aV:** Power **Method**: Pre-post test **Instruments used:** Geometric Optics Conceptual Understanding Test (GOCUT) **Duration:** 3 months **Country:** Rwanda **Grade level:** Senior 4 (10th grade) **Subject context:** Mirrors and lenses, physics **School type:** Senior Secondary School	In this study, 136 tenth-grade students from six different schools were examined. The study investigated whether the medium or teaching method had an influence on the students' performance. The study found that the students achieved significantly better results compared to conventional teaching methods (books, blackboard and chalk) when they used a simulation ($p < .01$, $d = 0.73$) or YouTube videos ($p < .001$; $d = 0.62$) to process the content.
Zendler & Greiner (2020)	**N:** 45 (group 1 n = 23, group 2 n = 21) **uV:** Treatment (instruction method), time (before the unit, after the unit), learning groups (class) Group 1: instructed by computer simulation Group 2: instructed by experiment **aV:** Learning success (prior knowledge, observation knowledge, application knowledge) **Method**: SPF-2x2*2, pre-post test **Instruments used:** n.a. **Duration:** 90 min **Country:** Germany **Grade level:** 8 **Subject context:** Chemistry, metals **School type:** Grammar school	In this study, the two methods of real-life experiment and computer simulation were compared. For this purpose, two eighth-grade classes were divided in half and half of the students were taught using experiments, while the other half worked exclusively with computer simulations. The study found no significant differences in learning success between the two methods - experiment and computer simulation - in terms of prior knowledge ($F(1,41) = 0.73$, $p < .40$), *observational knowledge* ($F(1,41) = 0.74$, $p < .40$) or *application knowledge* ($F(1,41) = 1.22$, $p < .28$).
Dickes et al (2019)	**N:** 43(female n = 19, male n = 24) **uV:** Treatment (work in immersive environment simulation) **aV:** Change in causal explanations, self-efficacy in science and programming **Method**: Mixed method, interviews, video recordings, screen recordings, student artefacts, field notes	In this pilot study, an immersive simulation was used to investigate the extent to which models designed by students change through the realistic simulation of an ecosystem. To analyze this, the recordings were examined for transformative moments, which the researchers defined as moments in which the students transform material in such a way that the

	Instruments used: Analysis of transformative moments (according to Shen 2011) **Duration:** 15 days of 50 minutes each **Country:** Great Britain **Age:** 8 - 11 **Subject context:** Biology, forest and pond ecosystem **School type:** n.a.	material conveys different information after the transformation than before the transformation. Over the course of the units, the test subjects increasingly developed their models "for a reason" and paid more attention to the fit with mechanisms of the reference object than to the functionality of the models when evaluating them. As time went on, the students developed a more precise idea of causal relationships in the simulated ecosystem.
Magana et al (2019)	**N:** 318 (Group 1 n = 131, Group 2 n = 187) **uV:** Treatment (simulation), cluster Group 1: naive developing heurists Group 2: half-knowledgeable, fixed heurists **aV:** conceptual understanding, factual knowledge, conceptual/contextual knowledge and representative knowledge **Method**: Mixed method, pre-post test, multiple choice, open questions & answers **Instruments used:** own multiple-choice questionnaire (α_{IR} = .917), n-TARP clustering **Duration:** 2 weeks of 45 minutes per day **Country:** USA **Grade level:** 6 - 8 **Subject context:** physics, energy consumption, energy saving, planning of buildings **School type:** Middle school	The aim of this study was to examine the extent to which the use of simulation-based learning-by-design learning experiences had an influence on the heuristics generated by students. It was found that the intervention resulted in a significant increase in performance in the pre-post test comparison (X^2= 14.692, four degrees of freedom, p = .005385). In addition, an analysis according to clusters showed that two significant groupings could be distinguished. The allocation to these was determined on based on the results of the pretest. The two groups differed significantly (X^2= 62.51, three degrees of freedom, p < .001). While group 1, who were labelled "naive, developing heurists", improved significantly in both the multiple-choice test (X^2= 6.1134, two degrees of freedom, p = .047) and in answering open questions on the interpretation of graphics (X^2= 17.005 , two degrees of freedom, p < .001), group 2, which was labelled "semi-knowledgeable, fixed heurists", did not show any significant performance gains and even recorded a significant *decline in performance* in the category of open-ended conceptual/contextual questions (X^2= 12.087, two degrees of freedom, p = .0024). The results indicate that simulation-based learning-by-design learning experiences can be helpful in generating inherent heuristics, especially for students with little to no prior knowledge and understanding of concepts.
Rosenberg & Lawson (2019)	**N:** 13 **uV:** Treatment (simulation) **aV:** Conceptual understanding **Method**: qualitative explorative data analysis, open questionnaire **Instruments used:** own open questions Qualitative: Explorative data analysis, structure and in- vivo coding framework, content analysis **Duration:** 3 hours **Country:** USA **Grade level:** Upper school **Subject context:** physics, thermodynamics, diffusion **School type:** High school	This study investigated how students in online science courses use simulations to develop their conceptual understanding of a phenomenon and to serve as a context for analyzing a complex, quantitative data source. In addition, the students should critically analyze the strengths and weaknesses of the simulation used, also with regard to the variability of data sets. The exploratory analyses indicated that the simulation can bring about an improvement in formal academic language and scientific understanding. The students formulated their mechanistic explanations from previously descriptive to increasingly causal explanations. In addition, students cited the ease of understanding complex phenomena as one of the main advantages of simulations.
Bilek et al. (2018)	**N:** 91 **uV:** Treatment (simulation), social background and educational background **aV:** interest, performance	This study investigated how the use of computer simulations affects students' ability to set up chemical equations. The aim was to transfer principles from a known system (sandwiches) to

	Method: Pre-post test, multiple-choice test **Instruments used:** own questionnaire **Duration:** 50 min (5 min pre-test, 30 min intervention, 15 post-test) **Country:** Czech Republic **Age:** 8 - 12 **Subject context:** Chemistry, reaction equations **School type:** Primary school	an unknown one (chemical reactions) and thus simplify learning and also teach the setting up of chemical equations. The researchers wondered whether this approach would allow students to set up correct chemical equations before they started their chemistry lessons. The results showed that the students did not perform significantly differently depending on their social and educational background (sandwich p = .685, molecules p = .941). Although the students performed identical mathematical and logical actions in the two simulations, their performance in the simulation for setting up chemical equations of molecules was significantly worse than in the simulation for "making sandwiches" (t= -2.836, df= 180, p = .005). At the same time, the hypothesis that the students' performance would have an effect on *interest* could neither be confirmed for the sandwich simulation (X^2=3.310, dF = 2, p = .191) nor for the simulation of molecules (X^2=0.376, dF = 2, p = .829). Ultimately, the researchers concluded from the high proportion of correct answers in the post-test (sandwich 71%, molecules 57%) that the simulation is suitable for teaching students how to set up chemical reaction equations without them having had chemistry lessons beforehand.
Correia et al (2018)	**N:** 114 (minus dropout n = 109) **uV:** Treatment (simulation) **aV:** Attitude towards computer simulation, conceptual change **Method**: Pre-post test, questionnaire, mixed method **Instruments used:** own questionnaire Qualitative: Thematic content analysis, triangulation (according to Denzin, 1989) **Duration:** 2x 45 min **Country:** USA **Grade level:** 9 - 12 **Subject context:** Chemistry, behaviour of gas at the atomic level **School type:** High School	This study investigated how the perception and understanding of the behavior of gas on a sub-microscopic level is achieved mainly through the use of simulations. The vast majority (approx. 85%) of the students studied recommended the program to others and were positive about their learning experience. They described the program as an effective and efficient method. In addition, 50% more test subjects were able to answer the questions in the test correctly compared to before the learning unit.
Eckhardt et al. (2018)	**N:** 117 (Group 1 n = 19, Group 2 n = 15, Group 3 n = 20; Group 4 n = 17, Group 5 n = 23, Group 6 n = 23) **uV:** Treatment (didactic support for self-regulation (support for reflection), solution (not given / given / generated)) Group 1(NS/NR): no solution, no reflection support Group 2(GiS/NR): given solution, no reflection support Group 3(GeS/NR): generated solution through solution support, no reflection support Group 4(NS/R): no solution, reflection support Group 5(GiS/R): given solution, reflection support Group 6(GeS/R): generated solution with solution support, reflection support **aV:** intuitive increase in knowledge, self-assessment **Method**: Pre-post test, 3x2 study design, multiple-choice test, open questions	This study investigated how different conditions affect learning with simulations. The study focused on how different framework conditions and aids when working with a simulation affect the *increase in knowledge* and the *perceived cognitive load*. For this purpose, two variables were adjusted for each group. One group (NS/NR) received no solutions and no support in reflecting on the tasks. This group was declared the control group. The second group was given possible solutions and had to choose between them, but received no support in reflecting on the tasks. The third group (GeS/NR) received help in solving the tasks, for example by being given the beginnings of sentences but having to complete them themselves so that they created the solutions themselves. This group was not given any help with reflection. The fourth group (NS/R) did not

	Instruments used: own questions **Duration:** 70 min **Country:** Germany **Grade level:** 8 **Subject context:** biology, ecology, predator-prey relationship **School type:** Secondary school	receive any support in solving the tasks, but did receive support in reflecting on them afterwards. The fifth group (GiS/R) received both predefined solutions and support with reflection. The sixth group (GeS/R) received support in creating their own solutions and help with reflection. It was found that the students recorded a significant increase in knowledge in the pre-post comparison across all six conditions ($F(1,111) = 15.95$, $p < .001$, $\eta^2 = 0.13$). When comparing the conditions, only the groups with *a given solution and reflective support* (GiS/R $F(1,39) = 4.99$, $p < 0.05$, $\eta^2 = 0.11$) and *a given solution and no reflective support* (GeS/NR; $F(1,37) = 5.58$, $p < .05$, $\eta^2 = 0.13$) showed significant knowledge gains. Other conditions showed no significant knowledge acquisition. The results suggest that the simulation contributed significantly to intuitive knowledge acquisition.
Garneli & Chorianopoulos (2018)	**N:** 34 (video game = 17, simulation = 17) **uV:** Treatment, medium (simulation / video game) Group 1: Simulation Group 2: Video game **aV:** CT skills, number of primitives, motivation **Method**: Pre-post test, mixed-method study, interview (semi-structured interview guide) **Instruments used:** Dr Scratch (CT capability) Qualitative: Inductive content analysis **Duration:** 5 sessions of 2 hours each **Country:** Greece **Grade level:** 9 **Subject context:** Computer science, programming, Computational Thinking **School type:** Middle School	This study focused on the influence of video game and simulation design on students' *computational thinking* (CT) skills and *motivation*, which was interpreted from the *number of primitives used* (Nr_of_P) and subsequent interviews. Primitives are the smallest coherent units of a programming language that can contain information (arithmetic operations, data types, ...). The simulation group showed a significant increase in *CT ability* ($z = -2.310$, $p = .021$, $r = -0.31$) in the pre-post comparison, while the change in Nr_of_P was not significant. In the video game group, the changes were significant for both CT skills ($z = -2.298$, $p = .022$, $r = -0.31$) and Nr_of_P ($z = -3.198$, $p = .001$, $r = -0.43$). No significant difference was found in the comparison of the two test groups with regard to CT skills, but the video game group performed significantly better than the simulation group in the comparison of the Nr_of_P ($U = 215.500$, $z = 2.448$, $p = .013$). The qualitative analysis of the interviews showed that 58% of the students in the simulation group rated their programming skills as improved.
Wen et al (2018)	**N:** 25 **uV:** Treatment (simulation) **aV:** Modelling progress, modelling strategies **Method**: supervised and unsupervised lag sequential analysis (LSA) **Instruments used:** LSA (according to Bakeman and Gottman, 1997) Qualitative: content analysis $\alpha_{IR} = .73$ **Duration:** 90 min **Country:** Taiwan **Grade level:** 10 **Subject context:** Physics, kinematics **School type:** Secondary school	In this study, students' modelling progress and strategies were examined using a problem-solving simulation game. It was found that the simulation game has the potential to help students reflect on a scientific problem and thus construct a feasible and correct model to solve the problem. Certain patterns of student activity seemed to lead to concrete modelling and problem-solving results. Students who used the model-centered or reference material-based modelling pattern were more likely to be able to solve the problem. Of the 15 students who designed their models in the "momentum" or "velocity change" categories, which require a basic understanding of kinematics, 14 were able to solve the problem during the modelling test phase. It was also found that relying on intuitive

		knowledge in reference materials only produces weak and unstable models that cannot be applied to other contexts.
Yoon et al (2018)	**N:** 374 (minus dropout n = 365, group 1 n = 64, group 2 n = 63, group 3 n = 67, group 4 n = 162) **uV:** Treatment (scaffold method), content Group 1 (control): interactive visualisation object + description Group 2: Scaffold method from group 1 + augmented reality simulation Group 3: Scaffold method from group 2 + reference questions for interaction with the object Group 4: Scaffold method from group 3 +worksheet with questions and sentence starters, answers from other students and tips for reaching a consensus **aV:** Understanding of the phenomenon, perception of the method **Method**: Pre-post test, qualitative survey, questionnaire, interviews (n = 117), video observation **Instruments used:** own questionnaire (α_{IR}> 0.8) Qualitative: Coding of scaffolding affordances (according to Miles et al., 2013), coding of AR scaffolds (according to Yoon & Wang, 2014) Coding according to collaborative scaffolding (according to Belland, 2014; Belland et al., 2013; Ito et al., 2013; Scardamalia 2002) **Duration:** 1 museum visit **Country:** USA **Grade level:** 5 - 8 **Subject context:** Physics, magnetism, Bernoulli effect **School type:** Museum	The study investigated synergistic effects of different scaffold methods on students' understanding of the phenomenon and their perception of the scaffold methods. The results from the magnetism part showed that group 4 only achieved significantly better results compared to group 1 ($F(3,128) = 2.729$, $p = .047$). In the "Bernoulli effect" section, group 4 performed significantly better than group 1, group 2 and group 3 ($F(3,170) = 14.50$, $p = .001$). In the interview, 5 categories were identified in which the students thought when they answered the question of how the AR simulation helped them. 38% stated that the AR simulation made phenomena visible. 22% stated that the AR simulation contained details. 22% felt that the dynamics of the AR simulation made learning easier. 9% found the AR simulation helpful because it made it easier for them to stay focused. A further 9% mentioned that the interactivity of the AR simulation was helpful.
Chang (2017) Study I	**N:** 62 (group 1 n = 32, group 2 n = 30) **uV:** Treatment (simulation, observation) Group 1(Simulation): Simulation Group 2(Observation): Video of simulation **aV:** learning time, result after the test, number of attempts, learning efficiency **Method**: Pre-post test, multiple-choice questionnaire, open questions, **Instruments used:** own questionnaire and coding (κ =0.91) **Duration:** approx. 30 minutes **Country:** Taiwan **Grade level:** 8 **Age:** 14 - 15 **Subject context:** Physics, buoyancy **School type:** Secondary school	This study investigated whether interactive engagement with a simulation or pure observation using a video resulted in differences in *learning time*, *post-test score* and *number of attempts*. In the pre-post comparison, both treatment groups showed significant learning gains (simulation: $Z = 4.62$, $p < .001$; observation: $Z = 3.87$, $p < .001$). No significant differences in *learning time*, *post-test score*, *number of trials* and *learning efficiency* were found in the comparison of the two treatment groups. The *learning efficiency* was defined as the quotient of the result after the test and the *number of attempts*.
Chan (2017) Study II	**N:** 66 (group 1 n = 32, group 2 n = 30) **uV:** Treatment group (key questions, structured questions) Group 1: guiding questions Group 2: structured questions **aV:** learning time, result after the test, number of attempts, learning efficiency **Method**: Pre-post test, multiple-choice questionnaire, open questions, **Instruments used:** own questionnaire and coding (κ = 0.91) **Duration:** approx. 30 minutes	In the second study by Chan (2017), the aim was to investigate whether scaffolding in the questions has an influence on *learning time*, the result after the test, the *number of attempts* and *learning efficiency*. This time, all students worked with the simulation. The students had to learn to understand the principle of buoyancy. While group 1 only received guiding questions that focused on the final learning objective, group 2's questions were more step-by-step and guided them through the simulation in a structured manner.

	Country: Taiwan **Grade level:** 8 **Subject context:** Physics, buoyancy **School type:** Secondary school	In the pre-post comparison, both treatment groups showed significant learning gains (guiding questions: $Z = 4.62$, $p < .001$; structured questions: $Z = 4.94$, $p < .001$, $d = 0.90$). The two groups showed no significant differences with regard to the *result after the test* and the *learning time*. In terms of the *number of attempts*, the students in group 2 scored approximately twice as high as group 1. The difference was significant ($U = 56.50$, $z = -4.01$, $p < .001$, $d = 0.90$). The difference between the two groups was also significant with regard to *learning efficiency*. Here, group 1 showed a higher learning *efficiency* than group 2 ($U = 140$, $z = -2.39$, $p=0.02$, $d = 0.91$).
López & Pintó (2017)	**N:** 18 (group 1 n = 9, group 2 n = 9) **uV:** Simulation content Group 1: Friction Group 2: Electromagnetic induction **aV:** Difficulties in reading and interpreting **Method**: Semi-structured interview **Instruments used:** Content analysis and classification **Duration:** 20 minutes **Country:** Spain **Age:** 14 - 16 **Subject context:** physics, frictional heat, electromagnetic induction **School type:** Secondary school	In this study, the researchers focused on the question of what difficulties can occur when reading visual representations. They had 18 pupils work with simulations and then interviewed them. They then categorized the answers into five categories: Structure of the representation (CS); Difficulty in giving appropriate relevance to each visual element in the composition (RE); Difficulty in giving appropriate semantic meaning to each visual element in the composition (SEM); Difficulty in dealing with multiple representations (MR); Difficulty in perceiving and interpreting dynamic information (DY). The compositional structure of the representation can be misunderstood in simulations, which then affects the students' understanding of the entire composition. Misreading not only contributes to confusion, but can also lead to visual elements of the simulation being given the wrong meaning, so that different information is conveyed than intended. Emphasising elements influences reading and can thus direct students' perception and attention. Strong colours are one example of this. However, this can also have counterproductive consequences, so that some visual elements "overshadow" other visual elements through their presentation. The presentation of scientific objects that require a certain amount of prior knowledge can lead to difficulties in assigning the corresponding meaning.
Thisgaard & Makransky (2017)	**N:** 128 (group 1 n = 66, group 2 n = 62) **uV:** Sequence Group 1: Simulation, classical teaching Group 2: classical teaching, simulation **aV:** increase in knowledge, self-efficacy, intrinsic motivation, interest, expectation of results **Method**: Pre-midway post-test, crossover design, multiple-choice questionnaire **Instruments used:** Questionnaire on motivated learning strategies (Pintrich et al., 1991), intrinsic motivation inventory (Deci et al., 1994), method for assessing interest in computer science disciplines (Lent et al., 2008, 2011), O∗ Net online database for occupations (Honey and Hilton, 2011),	This study investigated the effect of simulations and traditional lessons on *knowledge growth, self-efficacy, intrinsic motivation, interest* and *outcome expectations.* To this end, the students first attended lessons under one condition (simulation or traditional lessons) and then swapped after the lunch break so that the group that had previously worked with simulations now received traditional lessons. Group 1 showed a significantly higher *increase* in knowledge from the pre-test to the midway test than Group 2 ($(M = 1.84)$, $t(126) = -3.89$, $p < .001$, $d = 0.69$). The addition of classical instruction in group 1 did not result in a

	Bandura's definition of outcome expectations (Bandura, 1986; Lent et al., 1994), self-efficacy (according to Pintrich et al., 1991). **Duration:** 2 times 1.5 hours **Country:** Denmark **Grade level:** Upper school **Subject context:** Biology/ Biotechnology **School type:** Grammar school	significant *increase in knowledge*. When simulation was added to group 2, a significant *increase in* knowledge was found (M = 1.40 to M = 13.13, t(61) = -5.59, p < .001, d = 0.59). In addition, group 2, in which the simulation started at level 2, had a significantly higher change (M = 1.73) than group 1, which started the lesson at level 2 (M = 0.35, t(126) = 3.80, p < .001, d = 0.67). It was also found that the change in *knowledge* between immediately before and immediately after the simulation (M = 2.66) was significantly higher than the change in *knowledge* from immediately before to immediately after the lesson ((M = 1.07), t(127) = -4.31, p < .001, d = 1.32). While *self-efficacy* did not increase significantly in the pre-midway test comparison in group 1, it increased significantly in group B (M = 4.06 to M = 4.20; t(61) = -2.74; p = .008; d = 0.10). In addition, a further significant increase in *self-efficacy* was found when the simulation was added to group 2 ((M = 0.01), t(126) = 3.19; p = .002; d = 0.56). No significant differences were found between the two treatment groups with regard to *motivation*, *interest* and *expected results.* However, there were significant differences between the two groups in terms of *self-efficacy*. In a direct comparison immediately before and after the simulation, the students showed a significant increase in interest (d = 0.14).
Blikstein et al. (2016)	**N:** 13 (female n = 4, male n = 9) **uV:** Treatment (bifocal modelling) **aV:** Modelling behavior of the students **Method**: Video recording, screen capture software, interviews, field notes **Instruments used:** Content and context analysis **Duration:** 3 days (15 hours) **Country:** USA **Grade level:** n.a. **Subject context:** Biology **School type:** High School	Blikstein et al. (2016) investigated a new scientific research framework, so-called "bifocal modelling", in which students were able to simultaneously test the simulated computer models they had created with the help of experiments. The researchers focused on situations in which an explicit comparison was made between the experiment and the virtual model. These moments were then analyzed in terms of content and context and iterative comparison moments were identified. When students designed simulated models, the bifocal framework was shown to be a possibility for new interaction between the micro and macro levels of a phenomenon. The simulation and its mathematical algorithms were used and retained by the students as an anchor in the process of understanding the reproductive behavior of bacteria at the micro level. This enabled a new kind of epistemic game through which the students were able to accelerate and deepen their insights. The discrepancies between the experiment and the simulated model encouraged the students to research beyond the scope of the experiment and to further investigate the phenomenon on different levels. The newly acquired knowledge was then incorporated into the simulated model and enabled the modelling of a more complex and realistic image of reality. The bifocal approach proved to be conducive to working through discrepancies between the model and the

		physical experiment. The close coupling of virtual modelling and physical experiment proved to be both conceptually and epistemologically generative and not detrimental to the learning process.
Chen et al. (2016)	**N:** 152 (subsamples not mentioned) **uV:** Treatment (sequence of the simulation), scale fidelity of the simulation Group 1: True-to-scale model, orrery model Group 2: Orrery model, true-to-scale model **aV:** Performance in scale-dependent and scale-independent questions **Method**: Pre-post test **Instruments used:** Astronomy and Space Science Concept Inventory (ASSCI inventory, content knowledge) **Duration:** 3 days **Country:** USA **Age:** 15 **Subject context:** Astronomy, solar system **School type:** High School	This study involved the use of two different simulated visual models, one representing the solar system in a more true-to-scale and complex way (True-to-Scale), while the other model represented the solar system in a simplified way (Orrey model). When using the true-to-scale model, students recorded significant increases in correctly answering *scale-dependent questions* ($p < .001$). This was not the case when using the Orrery model. With regard to the *scale-independent questions*, the students made significant gains when using both models (true-to-scale: $p < .01$, EF = 0.40 SD; orrery: $p < .001$, EF = 0.31 SD). The true-to-scale model led to significantly better results when it was used as the first rather than the second model ($p < .01$). In addition, students made significantly greater progress when they first worked with the true-to-scale model and then with the Orrey model than vice versa ($X^2 = 8.12$, $p < .01$).
Caglar et al. (2015)	**N:** 64 (minus dropout n = 37) **uV:** web-based simulation **aV:** Learning gain in computational thinking, scientific **Method**: Pre-post test, questionnaire **Instruments used:** own questionnaire **Duration:** 5 days **Country:** USA **Grade level:** n.a. **Specialized context:** Computer science, control and regulation of traffic **School type:** High School	This study investigated the impact of using a web-based cloud-based simulation on *learning gains in computational thinking and scientific understanding*. The use of the web-based simulation showed no learning gains in *scientific reasoning* and *computational thinking* in the overall group except for one question that tested *understanding of the conditional construct* ($p = .01$; $d = 0.42$). A subsequent group subdivision in which only the students who had recorded *CT learning gains* were recorded showed that the group with positive *CT learning gains* (n = 17) showed significant improvements ($p < .001$; $d = 0.68$).
Mešić et al. (2015)	**N:** 63 (minus dropout n = 52, group 1 n = 17, group 2 n = 16, group 3 n = 19, female n = 32, male n = 31) **uV:** Treatment (simulation / pictures of the simulation / classic diagrams from the textbook) Group 1 (S): Simulation Group 2 (SSF): Images of the simulation Group 3 (CSD): Classic diagrams from the textbook **aV:** attitude, understanding **Method**: pre-post test, quasi-experimental design, questionnaire **Instruments used:** own questionnaire (based on Force Concept Inventory (Hestenes et al., 1992), Mechanics Baseline Test (Hestenes & Wells, 1992), Test of Understanding Graphs in Kinematics (Beichner, 1994) and Force and Motion Conceptual Evaluation (Thornton & Sokoloff, 1998), $\alpha = .55$; 0.65) **Duration:** 10 hours over 5 weeks (including 6 hours of intervention)	In this study, three different teaching methods (simulations, image sections of the simulations and classic diagrams) were compared with each other to find out how the handling affects the students' *understanding* and *attitude*. Teaching with simulations is significantly more effective than conventional teaching with static diagrams ($t(48) = 2.25$, $p < .05$, $r = 0.31$). There were no statistically significant differences when comparing the images of the simulation and the simulation itself. Girls benefited significantly more from the intervention than boys in both the S and SSF groups (S: $t(15) = 2.26$, $p < .05$, $r = 0.5$ and SSF: $t(14) = 3.53$, $p < .05$, $r = 0.69$). There were no gender-specific differences in traditional teaching. There were significant differences between the groups with regard to *attitudes* towards learning and teaching one-dimensional kinematics ($F(2,46) = 4.30$, $p < .05$, $\eta^2 = 0.16$). Students from groups S and SSF showed significantly higher differences

	Country: Bosnia and Herzegovina **Grade level:** Middle school **Age:** 11 - 16 **Subject context:** Physics, one-dimensional kinematics **School type:** Grammar school	compared to group CSD (group S: ($t(46) = 2.26$, $p < .05$, $r = 0.32$) and group SSF: $t(46) = 2.75$, $p < .05$, $r = 0.38$)). The S and SSF groups did not differ significantly from each other. 92% of students in the S group agreed (23%) or strongly agreed (65%) with the statement "I think that physlets/sequences of simulation frameworks should also be used in other physics lessons". In contrast, only 53% of the SSF group agreed with this statement. 13% of the SSF group did not agree with the statement at all.
Olakanmi (2015)	**N:** 66 (group 1 n = 34, group 2 n = 32), no information on gender distribution **uV:** Gender, Treatment Group 1(experimental group): web-based simulation Group 2(control group): traditional analogue teaching **aV:** Performance, attitude to chemistry **Method**: Pre-post test, observation and interviews **Instruments used:** "Rate of Reactions Knowledge Test" (RRKT), "Chemistry Attitude Scale" (CAS), "Class Activity on Rate of Reaction (ARR)" Qualitative: "think aloud" protocol **Duration:** 3 hours per week over 3 weeks **Country:** Nigeria **Grade level:** 10 **Subject context:** Chemistry, reaction rate **School type:** Secondary School	Two teaching settings were compared in this study. The experimental group learned with the help of web-based simulations, while the control group attended the analogue lessons provided by the school. Both groups were taught by the same person. The influence of the two teaching settings on the students' *performance* and *attitude towards chemistry* was analyzed. Compared to the control group, the experimental group showed significantly higher *changes in performance* ($t(64) = 4.952$, $p < .05$) with regard to chemical reaction speed. The academic performance of the experimental group before and after the test differed significantly in favor of the results after the test ($t(33) = -9.622$, $p < .05$). Web-based simulations have a positive influence on *attitudes* towards chemistry. The differences to the control group were significantly better ($t(64) = 3.066$, $p < .05$). No significant differences in *performance* were found with regard to *gender*. The simulations had a positive effect on the teachers' attention and awareness of being able to identify misconceptions more quickly. The simulation also helped them to transfer what they had learnt to real-life examples.
Eckhardt et al. (2013)	**N:** 117 (Group 1 n = 20, Group 2 n = 18, Group 3 n = 21; Group 4 n = 18, Group 5 n = 24, Group 6 n = 23) **uV:** Treatment (didactic support for self-regulation (support for reflection), solution (not given / given / generated)) Group 1(NS/NR): no solution, no reflection support Group 2(GiS/NR): given solution, no reflection support Group 3(GeS/NR): generated solution through solution support, no reflection support Group 4(NS/R): no solution, reflection support Group 5(GiS/R): given solution, reflection support Group 6(GeS/R): generated solution with solution support, reflection support **aV:** Increase in knowledge, perceived cognitive load **Method**: 3x2 study design, pre-post-test per day, pen-and-paper questionnaire, multiple-choice test **Instruments used:** own knowledge test (Pre: $\alpha = .68$; Post: $\alpha = .76$), perceived cognitive load (Urhahne, 2002, $\alpha = .86$) **Duration:** 2 days of 90 minutes each **Country:** Germany **Grade level:** 8	As part of this study, researchers looked at how different framework conditions and aids when working with a simulation affect *knowledge growth* and *perceived cognitive load*. With the exception of the generated solution and reflection support condition, all groups showed a significant increase in knowledge. In terms of further *knowledge gain*, it was found that the condition with given solution and without reflection support (GiS/NR), ($F(1,36) = 4.63$, $p < .05$, $\eta^2=0.11$), with given solution and reflection support (GiS/R), ($F(1, 42) = 5.68$, $p < .05$, $\eta^2=0.12$), and with generated solution and without reflection support (GeS/ NR), ($F(1,39) = 9.06$, $p < .01$, $\eta^2=0.19$), performed significantly better than the control group. No significant differences were found between the control group and the condition without solution and reflection support (NS/R) or the condition with generated solution and reflection support (GeS/R). With regard to the *perceived cognitive load*, the analysis of variance showed significant main effects for the didactic intervention ($F(2,118) = 4.85$, $p < .01$, $\eta^2=0.08$) and the support for self-regulation ($F(1,118) = 18.39$, p

	Subject context: Biology, ecosystem, water **School type:** n.a.	< .001, η^2= 0.14). The lowest values of *perceived cognitive load* were achieved by learners in the NS/NR, NS/R and GeS/NR conditions.
Price (2013)	**N:** 224 (group 1 n = 107, group 2 n = 117) **uV:** Treatment (overhead simulation), teacher:in (Mr R, Mr S, Mr T) Group 1 (SIM): Simulation Group 2 (OV): Overhead **aV:** percentage improvement **Method**: Pre-post test, mixed method, multiple choice and modified multiple choice (short answers), open questions (long answers) Video recordings (18 hours) **Instruments used:** own questionnaire Qualitative: constant comparative method, interpretation (according to Clement, 2000), microanalysis of image-based discussion strategies, lesson comparison study of overhead and simulation lessons taught by two different teachers, lesson comparison study of six pairs of overhead simulation lessons taught by three different teachers **Duration:** 4 weeks **Country:** USA **Grade level:** 8 **Subject context:** Natural sciences, gases **School type:** Middle School	In her dissertation, Price (2013) investigated the differences in learning in simulation-based and overhead-based discursive lessons. In addition, the influence of the teacher on learning success was investigated. Both students in the simulation group (SIM) and the overhead group (OV) showed significant improvements in the pre-post test comparison (SIM: short: $p < .001$; long: $p < .001$; OV: short: $p < .001$; long: $p = .005$). For Mr S, the students in both test groups showed significant improvements in the long questions (SIM: $p< .001$; OV: $p < .001$) but none in the short answers (SIM: $p = .206$; OV: $p = .705$). For Mr R, students in both experimental groups showed significant improvements in the long answers (SIM $p < .001$; OV $p = .023$), but only the simulation group also showed significant improvements in the short answers (SIM: $p = .002$; OV: $p = .090$). Mr T's students showed significant improvements in all responses in both groups ($p < .001$). Students who received simulation-based instruction showed significantly greater improvements than students who received overhead-based instruction (short: $p = .02$; long: $p < .01$).
Tapola et al. (2013)	**N:** 57 (minus dropout n = 52, group 1 n = 26, group 2 n = 26) **uV:** Conditions in the simulation (fixed concreteness, decreasing concreteness), gender, prior knowledge Group 1: fixed concreteness of the simulation Group 2: decreasing concreteness of the simulation **aV:** performance, interest **Method**: Pre-midway post-test, paper-and-pencil questionnaire **Instruments used:** Performance goal orientation (according to Niemivirta, 2002), own questionnaire (mastery-intrinsic ($\alpha = .87$), mastery-extrinsic ($\alpha = .87$)), performance orientation ($\alpha = .77$), performance avoidance ($\alpha = .77$) and work avoidance ($\alpha = .67$), prior knowledge ($\alpha = .77$) **Duration:** 2 hours **Country:** Finland **Grade level:** 5 and 6 **Subject context:** Mathematics, electrical circuits **School type:** Primary school	In this study, researchers looked at the effect of the degree of concreteness in the presentation of a simulation on students' *interest* and *performance*. Interest in maths was higher in boys than in girls ($t(50) = -2.07$, $p < .05$). There were significant interaction effects between simulation condition and *situational interest* ($F(2,84) = 6.50$, $p = .002$, η^2= 0.13). Although the initial level of both groups was initially the same, the situational interest decreased with decreasing concreteness during the course of the simulation and increased in the concrete simulation. The *situational interest rating* remained higher for boys than for girls over the entire period ($F(1,42) = 5.10$, $p = .029$, η^2= 0.11). Neither *interest in maths* nor *situational interest* were predictors of students' *performance*. However, *prior knowledge* was a significant predictor of *performance* (β= 0.47, $p < .01$). Although no correlation was found between *prior knowledge of electricity and interest in maths*, it was associated with student *performance*. The researchers then presented a model according to which mastery-intrinsic (β=0.30, $p < .05$) and work-avoidance orientation (β=-0.26, $p < .10$) were the strongest predictors of *interest in maths*. In addition, neither *interest in maths* nor *situational interest* in the tasks had a significant influence on the students' *performance*. However, the prior knowledge of the students (β=0.47, $p < .01$) and the task setting(β=0.36, $p < .05$) had a significant

		influence on *performance*.
Lin et al. (2012) Study I	**N:** 55 (male n = 27, female n = 31) **uV:** Treatment (simulation+teaching) **aV:** Research competence (planning, analysis, modelling), knowledge **Method**: Pre-post test, questionnaire **Instruments used:** Pre-post-inquiry skill test (planning: α = .87, analysis: α = .83, modelling: α = .83) **Duration:** 3 weeks of 2 hours each **Country:** Taiwan **Grade level:** 9 **Subject context:** Geology, fossils **School type:** Middle School	This first sub-study investigated how the use of the *FossilSim* simulation affects students' *research skills* and *knowledge*. The simulation was designed to allow students to conduct their research in a linear but recursive three-stage investigation process. Each investigation cycle contained guiding questions to help apply and develop the respective research skills. Significant positive changes were found in the students' *research skills* ($t = 3.696, p < .001, d = 0.47$). However, they did not show any significant improvement across all three research skills. While *planning* ($t = 6.663, p < .001, d = 0.87$) and *analysis* ($t = 3.807, p < .001, d = 0.50$) improved significantly, no significance was found in *modelling*.
Lin et al. (2012) Study II	**N:** 11 **uV:** Simulation **aV:** Change in the idea of geological time **Method**: semi-structured interviews, questionnaire **Instruments used:** Coding (Dodick & Orion 2003, $\alpha_{IR} = .83$) **Duration:** 35 min Intervention and (Interview **Country:** Taiwan **Grade level:** 9 **Subject context:** Geology, fossils **School type:** Middle School	The second sub-study looked at how the simulation helped the students to correctly categorise geological events. For this, students were asked to explain how they manipulated the FossilSim to recall and categorise the specific event. Only 24% of the students showed that they fully understood the laws and evidence for their answers. Therefore, the ability to recognise evidence is not enough to draw the correct conclusion. Nevertheless, 94% of students were able to show that they can draw logical conclusions. Occasionally, it was also shown that *FossilSIM* promoted logical thinking in relation to the sequence of geological events by providing macroscopic and microscopic evidence that was used by the students. However, according to the authors, there was a lack of opportunity to be guided through the simulation and to draw these conclusions themselves with assistance.
Shegog et al (2012)	**N:** 44 (group 1 n = 23, group 2 n = 21) **uV:** Treatment Group 1(T):1. Simulation + 2. simulation Group 2(C):1st homework+2nd simulation **aV:** procedural knowledge, declarative knowledge, attitudes towards the use of computers, manageability of the program **Method**: Pre-2-post test, questionnaire, multiple-choice test **Instruments used** Procedural knowledge: Error rate from the lab book, time per task (according to Dayal et al.2004) Declarative knowledge: 15-item multiple choice knowledge questionnaire (α = .90) Attitude towards computers for learning: 10-item survey with a 3-point response scale (according to Askar et al., 1992, α = .81) Attitudes towards careers in the natural sciences: modified 30-item questionnaire (after Britner & Pajares, 2001) Program usability: 36-item survey with 5-point Likert scales (according to Shegog et al., 2007) **Duration:** 1 day **Country:** USA	In this study, the effect of simulations on *procedural knowledge*, *declarative knowledge* and various *attitudes* was investigated. In terms of *procedural knowledge*, the students showed significantly fewer errors after the second session than in the first session ($p < .01$). The treatment group showed significantly fewer errors and time spent using the simulation for the first time than the control group ($M = 1.48$ vs. $M = 2.24$ errors, $p < .005$, $M = 16.65$ vs. $M = 13.29$ in minutes, $p < .001$). Students in the treatment group also showed significant improvements in *procedural knowledge* in posttest 1 ($p < .05$), while the control group did not. The control group showed a significant increase in *declarative knowledge* after using the simulation between posttest 1 and posttest 2 ($p < .005$). The treatment group showed significantly greater improvements in *declarative knowledge* than the control group ($p < .001$). *Attitudes towards the use of computers* improved significantly ($p < .001$). This increase in scores was recorded after each use of the simulation. There were no significant

	Grade level: 9 to 12 **Age:** 15 - 18 **Subject context:** Biology, molecular biology, DNA **School type:** High school	differences between the two groups in terms of attitude change towards computers. No significant changes were found in *attitudes towards science and careers in science.* The handling of the program was rated positively in both groups. (T: 89.4%, C: 84%).
Han & Black (2011)	**N:** 220 (minus dropout n = 175, group 1 n = 59, group 2 n = 54, group 3 n = 62) **uV:** type of simulation/multimodal representation (force, kinesthetic, not haptic), content (window handle, salad spinner, cogwheels) Group 1(FK): Strength and kinesthetics Group 2(K): strength Group 3(NH): non-haptic representation **aV:** Ability to create and customise a multimodal representation, knowledge of mechanisms of simple machines **Method**: Pre-post follow-up test, multiple choice, open questions **Instruments used:** recall test, inference test, transfer test, pretest**.** **Duration:** 2 times 45 min over 2 weeks **Country:** USA **Grade level:** 5 **Subject context:** physics, simple machines, cogwheels, (window handle **School type:** Elemetary School	In this study, the effectiveness of a haptically supported simulation on the development of a multimodal representation and functioning of simple machines by primary school pupils was investigated. Three different multimodal representations were tested. One group received feedback in the simulation through force and kinesthetics (FK), while the second group only received kinesthetic feedback (K). The third group had no haptic feedback (NH). In addition, three different contents (window handle, salad spinner and cogwheels) were offered. With regard to the *creation of multimodal representations*, the groups showed significant differences ($F(2,172) = 3.591$, $p = .030$, $\eta^2 = 0.038$) The students in Group 1 performed significantly better than those in Group 3 ($p = .044$), but not significantly better than the students in Group 2 ($p = 1.00$). In terms *of understanding further instructions,* the students who were in group 1 performed best in all three areas. However, the results were only significantly better than those of the second-best group in the content of cogwheels (FK, K: $p = .090$). No significant differences were found with regard *to knowledge retention*. With regard to the *transfer of knowledge to new tasks*, there were significant main effects for the simulation ($F(2,115) = 3.622$, $p= .030$,$\eta^2 = .059$). Here, also, group 1 proved to be the best group.
Jaakkola et al. (2011)	**N:** 50 (group 1 n = 12, group 2 n = 14, group 3 n = 12, group 4 n = 12) **uV:** environment (simulation (S), simulation and laboratory (C)), instruction (explicit (E), implicit (I)) Group 1(SI): Simulation + implicit instructions Group 2(SE): Simulation + explicit instructions Group 3(CI): Laboratory + implicit Group 4(CE): Laboratory + explicit instructions **aV:** Subjective knowledge, time, learning efficiency **Method**: Pre-post test, mixed method, recall interviews, qualitative content analysis **Instruments used:** prior knowledge ($\alpha = .667$), specialized knowledge ($\alpha = .822$), Qualitative: Categorization of responses (according to Kärrqvist, 1985, McDermott & Shaffer, 1992; Osborne, 1983; Shipstone, 1984, $\alpha_{IR} = .91$) **Duration:** 1 day (pre-test: one week before, post-test: 1 day after) **Country:** Finland **Grade level:** 5 and 6 **Subject context:** Physics, electrical circuits **School type:** Primary school	This study investigated different modalities of learning with simulations on *subjective knowledge*. While SI showed no significant changes in *subjective knowledge* ($p = .0314$, $d = 0.22$), Group 2 ($p = 0.013$, $d = 0.78$), Group 3 ($p = .001$, $d = 1.51$) and Group 4 ($p < .001$, $d = 1.24$) showed significant improvements. The changes from the CI and CE conditions were greater than one standard deviation, indicating a substantial *increase in knowledge* regarding the function of electrical circuits. Group 3, in which both the simulation and the laboratory experiment were used, showed significantly better results than the SI group in *subjective knowledge* (CI vs. SI $p < .001$, $g=1.62$(95% CI: 0.70 - 2.52)) and *learning efficiency* (CI vs. SI $p < .001$, $g=1.43$ (95% CI: 0.54 - 2.30). With regard to the instructions, it was found that the SE group performed significantly better than the SI group in *subjective knowledge* ($F(1,45) = 4.83$, $p = .03$, $g=0.84$ (95% CI: 0.05 - 1.61)), but not in learning efficiency. No significant differences were found between groups 4 and 3. The qualitative analyses showed that 67% of Group 4 were able to reproduce the Ohm model

		correctly, followed by 58% of Group 3. Less than half of the students in Group 2 (43%) were able to do so.
Liu & Chuang (2011)	**N:** 262 (no information on sub-samples) **uV:** instructions (verbal+ on-screen (RM)/ on-screen (SM)), prior knowledge (PK) Group 1 (experimental group): Verbal and visual instruction Group 2 (control group): Visual instruction **aV:** *Cognitive load* (CL), performance (P), cognitive efficiency (calculated from CL and P) **Method:** 2x2 between-groups factorial design, questionnaire **Instruments used:** prior knowledge (13 items: 5 true/false, 8 multiple choice, α = .769), learning success (27 items: 12 multiple-choice, 15 cloze, α = .794), *cognitive load* (according to Paas & van Merriënboer, 1993), **Duration:** Prior knowledge (one week before the intervention), intervention (90 min), *cognitive load* (immediately after the intervention), performance (one week after the intervention), performance (one week after the intervention). **Country:** Taiwan **Grade level:** 7 **Age:** 13 - 14 **Subject context:** Chemistry, oxidation and reduction **School type:** Middle school	This study investigated the extent to which the students' prior knowledge and the type of instructions affect the *cognitive load* and *performance* in dealing with simulations. The experimental group received auditory instructions in addition to on-screen instructions, while the control group only received on-screen instructions. The students were then tested on their prior knowledge (PK) one week before the intervention and the 27 best and worst per cent of their group were divided into two subgroups (high-PK, low-PK) based on their test results. A significant interaction was found between prior knowledge and *performance* ($r = 0.34$, $p < .01$), but no significant interaction was found between instruction types and prior knowledge. Although no significant differences were found between the low-PK and high-PK groups in terms of *performance*, the type of instruction had a significant effect on *performance* ($F = 4.08$, $p = .046$). There were also no significant differences between the high-PK and low-PK groups in terms of *cognitive load*. However, the RM and SM groups differed significantly ($F = 0.58$, $p = .034$) in favor of RM. With regard to *cognitive efficiency*, no significant differences were found between the groups.
Liu & Su (2011)	**N:** 169 (group 1 n = 86, group 2 n = 83) **uV:** Treatment (simulation+virtual lab, classical teaching and real lab) Group 1 (experimental group): Simulation+virtual laboratory Group 2 (control group): traditional teaching and real laboratory **aV:** Residential Wiring Achievement (performance in the wiring of (fictitious) private households), *Cognitive Load* **Method:** quasi-experimental design, pre-post test **Instruments used:** Electricity Concepts (α = .89), Residential wiring Achievement (30 multiple-choice items, α = .88), *Cognitive Load* (according to Paas et al.1994, 7-point scale, α = .90) **Duration:** 90 min per week over 6 weeks **Country:** Taiwan **Grade levels:** 11 **Age:**15 - 16 **Subject context:** physics, technology, electrical circuits **School type:** High school	This study investigated the extent to which students' performance in learning how to wire (fictitious) private households as an example of wiring electrical circuits and their self-perceived *cognitive load* depends on the learning environment they use. In the first two of a total of three learning units, students in the experimental group used a feedback-providing simulation in which they learnt the basics of electrical circuits. During the same period, the control group received regular lessons without simulation, in which illustrations from the simulation were also used. In the third learning unit, the experimental group used a virtual laboratory to solve tasks, while the control group used a real laboratory to complete the same tasks. Immediately after the intervention, both groups were tested in two tests on *Residential Wiring Achievement* and *Cognitive Load*. There were no significant differences between the two groups with regard to the pre-test. The results showed that the experimental group performed significantly better in the *Residential Wiring Achievement Test* than the control group ($t = 9.9$, $p < .001$). More specifically, the experimental group performed significantly better than the control group in all three subcategories of the test: *knowledge* ($t = 2.76$, $p = .006$), *understanding* ($t = 4.09$, $p < .001$) and *application* ($t = 12.38$, $p < .001$). With regard to

		cognitive load, the experimental group showed significantly higher values than the control group (t = 2.34, p = .02).
Sabagh (2011)	**N:** 70 (minus dropout n = 64, group 1 n = 31, group 2 n = 33, female n = 35, male n = 29) **uV:** Treatment (simulation / classical teaching), gender Group 1 (experimental group): Simulation (Web-Based Virtual Lab) Group 2(control group): traditional teaching **aV:** Conceptual understanding, scientific methodological competences **Method:** Pre-post test, questionnaire, multiple-choice test **Instruments used:** own questionnaire: conceptual understanding test (13 items; α = .82), scientific methodological competences (14 items; α = .85) **Duration:** Four weeks **Country:** Egypt **Grade level**: 4 **Subject context:** physics, biology, chemistry, density, microscopy, states of matter **School type:** Primary school	In Sabagh's dissertation (2011), the scientist focused on the influence of web-based virtual labs on *conceptual understanding* and *scientific methodological skills.* The 64 students were first randomly divided into two groups. The experimental group was then taught for four weeks using a web-based virtual laboratory, while the control group received traditional lessons (textbooks). In the experimental group, the intervention led a significant increase in the results in the pre-post comparison for both *conceptual understanding* (t(30) = 6.5, p< .001) and *scientific methodological skills* (t(30) = 4.9, p< .001). Furthermore, the experimental group performed significantly better than the control group in the skill *observation* (t = 2.224, p = .030, d = 0.556). However, this was only significant for the girls (p = .024) but not for the boys (p = .359). The same was true for the *predictive* skill (all: t = 2.3, p = .026, d = 0.572; female: t = 3.8, p = .001; male: t = 0.3, p = .741). The experimental group also performed significantly better in skill *interference* (p = .002, d = 0.799). This time, significant differences were found for both genders (female: p = .041; male: p = .031). The significances were the same for the *classification* construct (all: t = 3.9, p< .001, d = 0.975; female: t = 3.2, p = .003; male: t = 2.4, p = .024). No significant differences were found between the experimental and control groups in the *measurement* construct. In Skill *Communication*, the experimental group also performed significantly better than the control group. This time, however, the girls showed no significance, while the boys did the same (all: t = 2.6, p = .012, d = 0.648; female: t = 1.2, p = .236; male: t = 2.4, p = .022). No significant differences between the genders could be identified in any of the skills.
Eskrootchi & Oskrochi (2010)	**N:** 72 (group 1 n = 19, group 2 n = 33, group n = 20, female n = 40, male n = 32) **uV:** gender, grade level, treatment (project-based, experiment, simulation), computer background Group 1 (PB): project-based Group 2 (PBES): project-based + experiment + simulation Group 3 (PBS): project-based + simulation **aV:** Content knowledge, understanding of the topic, attitude of the students towards the project **Method:** Questionnaire survey **Instruments used:** Questionnaire (8 true/false questions, 7 open-ended questions, 23 multiple-choice questions) **Duration:** n.a. **Country:** USA **Grades:** 6th to 8th grade **Subject context:** Geography, Ecology **School type:** Middle school	In the study by Eskrootchi & Oskrochi (2010), sixth to eighth grade students were examined in three different treatments. They analyzed how the different combinations of project-based teaching, experiment and simulation affected the *students' content knowledge, understanding* and *attitude towards the project*. One group (PB) received traditional project-based lessons, the second group (PBES) was taught using an experimental model and a simulated model. The third group (PBS) was taught using a simulated model only. The results showed no significant differences between the three groups in terms of *content knowledge*. The PBES group performed significantly better than the other two groups in terms of *understanding the topic* (PBES, PBS: p = .002; PBES, PB: p< .001). The PBS group and the PB group did not differ significantly in this

		construct (p = .692). No significant interactions were found between the *boys* in the PBS and PB groups, but significant interactions were found in the *girls,* indicating that the treatments had a greater effect on the girls. No evidence was found that grade level had a significant effect on students' knowledge gains. 85% of the students had a positive attitude towards the simulation, which helped them to better understand the topic.
Gelbart et al. (2009)	**N:** 95 (group 1 n = 56, group 2 n = 39) **uV:** Treatment (simulation, control) Group 1(experimental group): Simulation Group 2(control group): Regular lessons with textbook **aV:** understanding, ability to explain **Method:** Mixed method, pre-post test, quantitative (n = 95): Questionnaire, qualitative (n = 4): Interviews, observation **Instruments used:** Comprehension (adapted from Gelbart & Yarden 2006; 12 true/false questions, α = .797) Qualitative: Categorization and reorganization of episodes (according to Shkedi, 2005) α_{IR}= 1.0 **Duration:** n.a. **Country:** Israel **Grade levels**: 12 **Subject context:** Biology, **School type:** Grammar school	In this study, pupils aged between 17 and 18 were divided into two groups and taught differently. While the experimental group learnt with the help of a laboratory simulation divided into five lessons, the control group learnt in parallel in regular lessons with the textbook. Both groups were tested both before and one month after the last lesson, with the second questionnaire containing the same topic but not the same tasks. Quantitative results: The results show that the experimental group showed significant increases in the true/false questions (t(55) = 5.61, p < .001) and in the explanations (t(55) = 5.26, p < .001). In addition, a comparison of the two groups shows that the experimental group performed significantly better than the control group both in the true/false questions (F(1,91) = 16.84, $p < 0.$001) and in the explanations ($F($1,92) = 14.55, $p < 0.$001). Qualitative results: When observing the students, two different learning types were identified. Type 1 was labelled "research-oriented learning". Here, students adopted the tools, procedures and research practices in genetics in order to solve the tasks and showed interest in these practices. In type 2, "task-based learning", students were not constantly involved in the research steps and only generated a general understanding of the research objective. They used the tools and procedures superficially and were more concerned with finding the right answers to the guiding question rather than engaging with the research practices. According to the researchers, type 1 is better suited to using research simulations profitably, as they are designed to highlight gaps that students should identify independently.
Tolentino et al. (2009)	**N:** 136 (minus dropout n = 97) **uV:** Intervention (SmallLab), class type (honors, regular) **aV:** Conceptual knowledge, spatial rotation thinking **Method** Qualitative: coded video data, reformed teaching observation protocol Quantitative: concept knowledge pre-posttest (multiple choice and open question), spatial rotation thinking **Instruments used:** concept knowledge (own	In this study, students in the tenth and eleventh grades of a high school were examined when using multimodal interactive (simulation) software that is projected onto the floor. Both qualitative and quantitative measurement methods were used. Qualitative results: While working with the simulation, the students gradually concretized their hypotheses and quickly became accustomed to working with the new technology. In doing so, they deepened their *conceptual understanding*. There was also

	questionnaire), spatial rotation thinking (ETS according to Ekstrom et al., 1979) Qualitative: coded video data, reformed teacher observation protocol RTOP (according to Sawada et al 2002) **Duration:** 3x 50 min **Country:** USA **Grade level:** 10 and 11 **Subject context:** Chemistry, acids and alkalis **School type:** High school	evidence of collaboration and distributed cognition. The students made use of multimodality, for example by responding to sounds that indicate when indicator molecules change color. There was also evidence of playful learning in some exercises where children started titrating solutions or working against them. Quantitative results: Overall, all students performed significantly better in the pre-post comparison with regard to *concept knowledge* ($t(96) = 7.87$, $p < .001$, $ES^2 = 0.8$). While the two class groups did not differ significantly in the pretest, the results of the groups differed significantly in the posttest ($F(1,95) = 78.76$, $p < .001$). Although the regular classes also recorded significant increases in the posttest ($t(59) = 4.79$, $p < .001$, ES=0.13), the increases in the honor classes were greater (ES=1.12). Both groups also showed significant increases in *spatial rotation thinking* ($F(1,84) = 81.25$, $p < .001$, ES=0.7). However, the interaction of both groups showed a trend in favor of the Honor classes ($F(1,84) = 3.23$, $p = .07$).
Urhahne et al. (2009)	**N:** 155 (male n = 75, female n = 79, not specified n = 1) **uV:** Treatment (76 3D simulation, 79 2D images), chemistry grade, spatial imagination, prior knowledge, self-concept, cognitive load **aV:** conceptual knowledge, factual knowledge, **Method:** Pre-post test **Instruments used:** visual-spatial skills KFT (according to Heller & Perleth, 2000), self-assessment (according to Slater et al., 1997), self-concept questionnaire: domain-specific self-concept (5 items, $\alpha = .90$), self-assessment of prior knowledge ($\alpha = .77$), cognitive load (4 items, $\alpha = .67$), conceptual knowledge (26 items, $\alpha = .77$), factual knowledge (21 items, $\alpha = .67$) **Duration:** 25 min. **Country:** Germany **Grades**: 8 **Subject context:** Chemistry, crystal structure **School type:** Secondary schools	The researchers compared the effects of 3D simulations and 2D images on students' conceptual and factual knowledge. Other factors were also examined in order to investigate interactions. In terms of conceptual knowledge, the students with the 3D simulations showed significantly better results than the group of 2D images ($t(153) = 2.36$, $p < .01$). However, no difference was found in factual knowledge. *Factual knowledge* was associated with *prior knowledge*, *chemistry grade* and *self-concept* ($R^2 = 0.10$, $F(6,138) = 2.41$, $p < .05$). *Spatial imagination* had a significant influence on *conceptual knowledge* (B=0.11, SE B=0.04,$\beta = 0.04$, $p < .01$). A difference in *cognitive load* between the two groups was also not found.

[2] The effect size ES was calculated here from the difference between the two mean values (pre-post test) divided by the standard deviation.

Barab et al. (2006)	**N:** 28 (female n = 16, male n = 12) **uV:** Use of computer simulation (QuestAtlantis) **aV:** performance, understanding **Method**: Mixed method, pre-post test, video recording, mini-interviews, game (as performance control) **Instruments used:** own questionnaire (α = .8) Qualitative: content analysis on (narrative, research practices, and (de)constructed inscriptions) **Duration:** 2 weeks **Country:** USA **Grade level:** 9-12 **Subject context:** Biology **School type:** High School	This study investigated how well a simulation could be embedded into the existing (science-oriented) curriculum, replacing conventional lessons. The highly gifted pupils were videotaped during the intervention, and some were interviewed afterwards. The simulation game (Quest Atlantis) also served to monitor performance through built-in tests. The results showed significant increases in *understanding* in the pre-post comparison ($F(1,23) = 39.73$, $p < .001$). An increase in individual performance at a distal level could not be demonstrated significantly after the simulation ($F(1,23) = 2.57$, $p = .122$). Based on the qualitative analyses, the researchers believe they placed too strong a focus on the narrative. Nevertheless, the students developed a rich perceptual, conceptual, and ethical understanding of science by participating in this narrative.
Henderson et al. (2000)	**N:** 20 **uV:** Simulation (Message in a Fossil) **aV:** Thinking ability **Method:** Pre-post test, interview, questionnaire **Instruments used:** own questionnaire **Duration:** five weeks **Country:** USA **Grade level:** 2 **Age:** 7 **Subject context:** Geology, fossils **School type:** Primary school	This study investigated how the "Message in a Fossil" simulation affects the *thinking skills* of second graders. It was found that *thinking skills* improved as a result of using the simulation. Measured by the improvement in test results, boys improved by 22% in the pre-post comparison, while girls improved by as much as 27%. The simulation had a particularly strong effect on low students (poorer pupils) - here the intervention led to a 42 per cent improvement. The high students (better students) benefited comparatively less at 14%.