The Best of Two Worlds: A Systematic Review on Combining Real and Virtual Experiments in Science Education

Salome Wörner
Leibniz-Institut für Wissensmedien
Technische Universität Kaiserslautern
LEAD Graduate School and Research Network, University of Tübingen

Jochen Kuhn
Technische Universität Kaiserslautern

Katharina Scheiter
Leibniz-Institut für Wissensmedien Tübingen
LEAD Graduate School and Research Network, University of Tübingen

Conducting experiments fosters conceptual understanding in science education. In various studies, combinations of real (hands-on) and virtual (computer-simulated) experiments have been shown to be especially helpful for gaining conceptual understanding. The present systematic review, based on 42 experimental studies, focuses on the following: (1) What is the relative effectiveness of combining real and virtual experiments compared with a single type of experimentation? (2) Which sequence of real and virtual experiments is most effective? The results indicate that (1) in most cases combinations of real and virtual experiments promote conceptual understanding better than a single type of experimentation, and (2) there is no evidence for the superiority of a particular sequence. We conclude that for combining real and virtual experiments, apart from the individual affordances and the learning objectives of the different experiment types, especially their specific function for the learning task must be considered.

Keywords: conceptual understanding, inquiry learning, real experiments, systematic review, virtual experiments
Acquiring scientific literacy (i.e., knowledge and skills in science) is essential for successful participation in today’s knowledge society (Organisation for Economic Co-operation and Development [OECD], 2007, 2016). Scientific literacy is critical to form an opinion and make informed decisions (National Research Council, 2012). Consequently, fostering scientific literacy has become a fundamental aim of science education (National Research Council, 2012; OECD, 2007, 2016). However, acquiring scientific literacy seems to be challenging for students: Results of the PISA 2006 and PISA 2015 studies showed that around 20% of all students in OECD countries cannot perform tasks that require only minimal competencies (i.e., that are located at Level 2 of the science competency scale). This is the level of basic competencies that students should reach by the end of their compulsory education (OECD, 2006, 2016). Therefore, more research on how to effectively foster the acquisition of scientific literacy is needed.

To improve scientific literacy, especially guided inquiry-based learning activities can provide valuable learning opportunities for students (Edelson et al., 1999; Lazonder & Harmsen, 2016). Inquiry-based learning is an approach where students are asked to act like scientists when conducting experiments. For a long time, inquiry-based learning activities have been implemented in science classrooms in analogue forms, for instance, by asking students to perform real (hands-on) experiments to test their hypotheses. In recent years, such analogue forms of experimentation have been enhanced with, and sometimes even replaced by, digital technologies (e.g., Becker et al., 2020; Brinson, 2015; de Jong, 2006). In particular, students have been asked to conduct experiments using virtual experiments or simulations (as defined in the subsequent section), which have been claimed to foster learning (Chernikova et al., 2020; de Jong & van Joolingen, 1998; Geelan & Fan, 2014). A variety of benefits and drawbacks of either real or virtual experiments have been discussed in the literature, suggesting that both may contribute unique aspects to foster scientific literacy (de Jong et al., 2013). Accordingly, the question of how to design effective learning opportunities in science classrooms is probably best answered by looking into how to combine real and virtual experiments, making use of their unique affordances for learning. In line with this reasoning, several researchers such as Alkhaldi et al. (2016), Brinson (2015), de Jong et al. (2013), and Hofstein and Lunetta (2004) suggest that combinations of real and virtual experiments might be most effective for science learning—but they leave open how to sequence them.

Using this preliminary evidence as a starting point, the goal of the present systematic review was to have a closer look into whether combinations of real and virtual experiments are more effective than real or virtual experiments alone and how they should be sequenced to maximize students’ conceptual understanding. Conceptual understanding is only one part of scientific literacy, but it is one of the most important learning goals in science education. Because of its outstanding importance, conceptual understanding is one of the most frequently measured outcome variables in empirical studies in the field of educational research. Therefore, the focus of our review lies mainly on this outcome measure. For this review, we assume conceptual understanding and conceptual knowledge to be similar constructs, and for consistency of wording, we will only use the term “conceptual understanding” in this article. Conceptual understanding is relational
knowledge about the core concepts in a domain and their interrelations, including the understanding of the relation between observable (e.g., physical) phenomena and the underlying (abstract) invisible principles (Goldwater & Schalk, 2016; Schneider et al., 2011).

Implementing Inquiry Learning Using Real and Virtual Experiments

Inquiry-based learning is a common instructional approach in science education that often makes use of experiments. It can be described as an educational strategy where methods and practices frequently used by professional scientists are transferred to education and implemented with specific guidance to enable and facilitate knowledge construction (Keselman, 2003; Pedaste et al., 2015). Pedaste et al. (2012) define it as “a process of discovering new relations, with the learner formulating hypotheses and then testing them by conducting experiments and/or making observations” (p. 82). Thereby, students need to participate actively in and show responsibility for their learning process to discover relationships between variables and to construct (conceptual) knowledge that is new to them (de Jong & van Joolingen, 1998). During inquiry learning, students are self-directed and complete all the stages of scientific investigation, including hypothesis formulation, experiment design, data collection, and conclusion drawing (Keselman, 2003). Inquiry learning that is appropriately guided and that actively engages students in the learning process has been shown to be more effective for learning than other instructional approaches like passive, teacher-centered direct instruction or unassisted discovery in several meta-analyses and research syntheses (Alfieri et al., 2011; Furtak et al., 2012; Minner et al., 2010). This claim is also supported by more recent studies, as long as inquiry learning is combined with guidance and preceded by direct instruction (Aditomo & Klieme, 2020; Chen et al., 2017; Lazonder & Harmsen, 2016; Oliver et al., 2019). For example, Aditomo and Klieme (2020) examined whether inquiry learning was only successful when guided by teachers, using data from 151,721 students from 5,089 schools from the 10 highest and the 10 lowest science performers in PISA 2015. They performed exploratory and confirmatory factor analyses and structural equation modelling and found that inquiry learning led to higher learning outcomes when it incorporated teacher guidance and lower learning outcomes when it did not.

Inquiry learning shares similarities with the more generic process of problem solving and can hence be roughly divided into three phases: problem identification, problem solving, and knowledge consolidation (Bell et al., 2010; Pedaste et al., 2015). Experiments can play an important role in all three phases (de Jong, 2019).

Digital technologies provide new possibilities and perspectives for guided inquiry learning in science education. These technologies (e.g., virtual experiments) can adequately accompany and implement all processes of inquiry learning (Becker et al., 2020; Bell et al., 2010; de Jong, 2006, 2019; Mäeots et al., 2008). To clarify the wording used to describe experiments in this article, the terms “real experiment” and “virtual experiment” are defined as follows.

Real experiments (RE) are experiments that are performed with concrete, physical materials and (measuring) devices. These are traditionally carried out in science lessons. A classic example of an RE in physics education is an experiment about electric circuits where light bulbs or other resistors are connected in parallel
or series circuits. Students perform the experiment with actual light bulbs, wires, a power supply, and multi-meters to measure voltage and current. RE are sometimes also called “physical experiments” (e.g., in Pyatt & Sims, 2012; Smith & Puntambekar, 2010; Sullivan et al., 2017) or described as taking place in a “physical laboratory” (e.g., in de Jong et al., 2013; Husnaini & Chen, 2019) or “hands-on laboratory” (e.g., in Kapici et al., 2019; Toth et al., 2014). We do not use the term “physical” in this article to avoid confusion with the subject domain of physics. Furthermore, we do not use the term “laboratory” to avoid confusion with a proper laboratory room.

In contrast, virtual experiments (VE) are interactive computer simulations that can be performed on laptops or tablets (no AR-/VR-glasses are needed; e.g., in de Jong et al., 2014; Smith & Puntambekar, 2010; Sullivan et al., 2017). Specific variables can be manipulated, and the consequences of this manipulation are directly observable. An example of the corresponding VE to the previously described RE about electric circuits is the Circuit Construction Kit (https://phet.colorado.edu/en/simulations/circuit-construction-kit-dc) where students can build an electric circuit on their computers with virtual wires, batteries, light bulbs, and resistors. They can then also measure voltage and current with a virtual voltmeter and a virtual ammeter. Different types of VE include animated data visualization and scaffolds or feedback to different extents. In their most basic form, they show the phenomenon only (e.g., “OptiLab,” used by Olympiou & Zacharia, 2012; Olympiou et al., 2013). In other cases, a numerical value of the observed variable is displayed in addition to illustrating the actual phenomenon (e.g., “pulley simulation,” used by Chini et al., 2012; Smith & Puntambekar, 2010). In more extended cases the VE even allows to record data and create a table or a diagram of the measurement points (e.g., “heat exchanger,” used by Wiesner & Lan, 2004; “Thermolab,” used by Zacharia & Constantinou, 2008; Zacharia & Olympiou, 2011). VE can be found, for example, on the websites of PhET (https://phet.colorado.edu) or Go-Lab (https://www.golabz.eu). Other frequently used expressions for VE are “(computer) simulation” (e.g., in Jaakkola et al., 2011; Olympiou et al., 2013; Renken & Nunez, 2013) and “virtual laboratory” (e.g., in de Jong et al., 2013; Kapici et al., 2019; Toth et al., 2014). Also, the terms “physical and virtual manipulatives” are often used for RE and VE, respectively (e.g., in Chini et al., 2012; Olympiou & Zacharia, 2012; Wang & Tseng, 2018).

de Jong et al. (2013) suggest that the different types of experiments foster different scientific competencies and skills: On the one hand, students benefit from RE due to the experiments’ haptic components and their authenticity (see also Renken & Nunez, 2013; Zacharia et al., 2012). de Jong et al. (2013) also mention that RE promote motor skills for handling certain materials and (measuring) devices (see also Zacharia et al., 2012). In addition, the authors claim that in RE, scientific methods are practiced because careful planning, setting up, and executing measurements is necessary (see also Renken & Nunez, 2013; Toth et al., 2009). On the other hand, according to de Jong et al. (2013), VE have the advantage that abstract or invisible objects and constructs can be made observable (see also Deslauriers & Wieman, 2011; Jaakkola et al., 2011; Olympiou et al., 2013; Zacharia & Constantinou, 2008; Zhang & Linn, 2011). For VE neither expensive or dangerous materials nor large and complex measurement devices are required.
Combining Real and Virtual Experiments

Additionally, experiments can be accelerated and repeated quickly (see also Zacharia et al., 2008). In VE, multiple representations of a phenomenon can be integrated with one another and functional correlations can be represented directly (see also Kollöffel & de Jong, 2013; McElhaney & Linn, 2011; van der Meij & de Jong, 2006). Furthermore, in VE very accurate measurements are possible and the experiments can be simplified to improve the students’ focus on relevant conceptual aspects (see also Ford & McCormack, 2000; Pyatt & Sims, 2012; Trundle & Bell, 2010).

In general, inquiry learning with RE is described as constrained compared with traditional (instructional) science teaching (Edelson et al., 1999). Challenges of inquiry learning are, for instance, the limited variable space (e.g., for cost, safety, time, or material reasons), which constrains the set of variables that can be investigated. Other challenges are limited observation possibilities in RE and limited modelling opportunities. All these challenges can be met by additionally making use of VE and their advantages described above. Thus, VE and RE have complementary affordances (Alkhaldi et al., 2016; de Jong et al., 2013; Kapici et al., 2019; Rau, 2020).

Previous Research on Real and Virtual Experiments in Science Education

In previous research, using VE in science education has proven to be an effective tool for learning (Chernikova et al., 2020; de Jong, 2006; de Jong & van Joolingen, 1998; Geelan & Fan, 2014). de Jong (2006) as well as Geelan and Fan (2014) argue that VE themselves can serve as an effective tool for scaffolding the processes during inquiry learning. de Jong and van Joolingen (1998) in their review and Chernikova et al. (2020) in their meta-analysis emphasize the importance of additional guidance during the inquiry learning processes with experiments. They state that additional instructional support during science experimentation with VE can help overcome typical problems of inquiry-based learning (de Jong & van Joolingen, 1998) and facilitate the learning process with the simulation (Chernikova et al., 2020).

Apart from this body of literature focusing on the general use of VE, there are also numerous studies dealing with the question of whether VE can replace RE in science education (synthesized with different perspectives in Brinson, 2015; Husnaini & Chen, 2019; Ma & Nickerson, 2006; Rutten et al., 2012; Sypsas & Kalles, 2018; Zacharia, 2015). The results of these reviews show that in most cases, student achievement is equal or higher in VE versus RE (e.g., equal in Renken & Nunez, 2013; Zacharia & Constantinou, 2008; and higher in Finkelstein et al., 2005; Pyatt & Sims, 2012). However, in some cases RE have been shown to promote learning better than VE (e.g., Josephsen & Kristensen, 2006; Srinivasan et al., 2006). One example of these reviews is the article by Brinson (2015), who synthesized empirical studies that had their focus on direct comparisons of learning outcome achievement in RE versus VE (or remote laboratories). The main findings of this review were that in most studies he reviewed (n = 50, 89%), the students’ learning outcome achievement was equal or higher in VE than in RE across all the learning outcome categories: knowledge and understanding, inquiry skills, practical skills, perception, analytical skills, and social and scientific communication. Although most studies included in his review (n = 53, 95%) had their
focus on outcomes related to content knowledge (Brinson, 2015). In contrast to these reviews focusing on a comparison of RE versus VE, in this article we do not focus on this issue but rather on the effects of combining RE and VE.

The suggestion of combining RE and VE and using VE as an enhancement instead of a replacement of RE can be found in many of the aforementioned review papers (Brinson, 2015; Ma & Nickerson, 2006; Rutten et al., 2012; Sypsas & Kalles, 2018) and also in numerous other papers that review experiences with VE in science education (Alkhaldi et al., 2016; de Jong, 2019; de Jong et al., 2013; Hernández-de-Menéndez et al., 2019; Hofstein & Lunetta, 2004). In these papers, the suggestion to combine RE and VE is most often grounded on the unique and complementary affordances of RE and VE, respectively (e.g., Alkhaldi et al., 2016; de Jong et al., 2013). The combination offers students perspectives and learning experiences in an environment that draws from both the affordances of RE and the affordances of VE, which could not be likewise achieved by either RE or VE alone (Alkhaldi et al., 2016; de Jong et al., 2013). In line with this, Rau (2020) in her review article compared multiple theories about learning with physical and virtual representations to clarify whether there are conflicting or complementary effects. She concluded that for meaningful combinations of real and virtual representations, different representation modes offer complementary affordances and engage students in different learning processes. Some of the papers that argue for a combination even suggest a strategy for sequencing RE and VE: Rutten et al. (2012) and Sypsas and Kalles (2018) recommend using VE as a preparatory learning task before performing RE. Ma and Nickerson (2006), on the other hand, suggest implementing RE initially to establish the accuracy of simulations for later study.

There is no consensus in the literature concerning the question of how to specifically sequence the two experiment types; moreover, there are also researchers advocating for blending RE and VE instead of sequencing them (e.g., Olympiou & Zacharia, 2012). Ma and Nickerson (2006) emphasize that the different educational objectives that are associated with each experiment type need to be considered when sequencing RE and VE. In line with this, Olympiou and Zacharia (2012) suggest that students should use RE and VE during an experimentation task according to the different learning objectives of the specific task. They suggest a framework where the instructor should first identify the general and specific learning objectives of a specific experiment considering the target group’s characteristics such as prior knowledge and skills. Afterwards the instructor should match these objectives with the affordances that have been identified through literature review for RE and VE. From this basis and a review of the available RE and VE affordances, the students’ ability to switch between RE and VE, and the students’ required knowledge and skills for RE and VE use, the instructor can finally create a blended combination of RE and VE for each individual experiment.

Many researchers emphasize that the learning process with the experiments needs to be guided adequately (Alkhaldi et al., 2016; de Jong et al., 2013; Lazonder & Harmsen, 2016; Zacharia et al., 2015). This is an important reason for instructors to thoughtfully design the combinations of RE and VE so that the pursued learning objectives can be achieved. In their article, de Jong et al. (2013) compare
the individual affordances of RE and VE and conclude as one of the three open “Grand Challenges” that “although the best combination may vary based on circumstances, combining both virtual and physical investigation is likely to be optimal” (de Jong et al., 2013, p. 308). According to de Jong et al. (2013), students who learn with “well-designed” combinations of RE and VE in most cases outperform students who learn with either type of experiment alone, a claim that was backed up by referring to a handful of studies that had been available at the time the article was written (e.g., Kollöffel & de Jong, 2013; Olympiou & Zacharia, 2012; Zacharia et al., 2008). However, the article of de Jong et al. (2013) is no longer up to date, as many studies in this field have been published since then. Also, as the article by de Jong et al. (2013) was not aimed at offering a systematic review on this issue, a comprehensive overview of the field of studies investigating combinations versus single experiments and possible designs of combinations (e.g., the sequence of the experiments in a combination) is still lacking. Looking especially at possible designs of combinations has considerable practical implications: When combining RE and VE, instructors have to decide with which experiment their students should best start, as working with both experiments simultaneously would probably be very challenging for most learners. However, it is yet an open question whether there is one sequence that is more beneficial than another. As Brinson (2015) stated, “The results of blended lab studies are mixed and no consensus exists yet regarding best practices, so this is a fascinating and important avenue of further research” (p. 230). This is where this systematic review is positioned.

Conducting a meta-analysis is an alternative to conducting a systematic review. However, this is not a viable approach here because the studies we identified for inclusion in our systematic review differ highly in their designs and boundary conditions of implementation. Additionally, the number of primary studies that could be used to estimate an overall effect size was considered too small; and low-quality regarding reporting standards in the original studies leads to substantial information missing for a meta-analysis. Also, there are benefits of doing a systematic review instead, namely, that the content of the papers can be described, clustered, and discussed in more detail and the focus can be put on analyzing and summarizing the underlying arguments, perspectives, and theories of the included articles. Furthermore, this systematic review is the first of its kind and therefore an initial attempt to compile the results of this area of research; thus, this synthesis should help first understand the area of research better and not only quantify the results. The focus of a meta-analysis is different from the focus of a systematic review, and we decided for the latter in the case of this article to give an overview and an insight into the heterogeneous landscape of this area of research.

Objectives

Against the backdrop of the previously introduced literature, this systematic review shall give a comprehensive overview of the research on combinations of RE and VE conducted in the past 20 years. We focus on two research questions (RQs):
**Research Question 1:** What is the relative effectiveness of combining real and virtual experiments compared with a single type of experimentation for conceptual understanding in science?

**Research Question 2:** Which sequence of real and virtual experiments is most effective for conceptual understanding in science?

**Method**

For this systematic review, we followed the recommendations of the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) framework (Moher et al., 2009).

**Eligibility Criteria**

To reduce publication bias in our final sample of included papers, we considered not only journal articles for this review but also book chapters, dissertations, and peer-reviewed conference proceedings.

We included studies that met all of the following criteria: (1) articles that deal with science learning, (2) articles that report combinations of RE and VE, and (3) articles that report experimental studies with some objective measure for conceptual understanding. Students’ self-reports about their perceived learning gains were not counted as objective measures, studies needed to use some sort of test instrument, assessment, or other objective scoring of knowledge or understanding to be considered for inclusion. Moreover, there were separate eligibility criteria for RQ1 and RQ2. For RQ1, we looked for articles that reported an experimental study in one of the science domains physics, chemistry, or biology, with a study design that compared (1) at least one group of students learning with a combination of RE and VE to (2) at least one group of students learning with a single type of experiment (i.e., RE only or VE only). For RQ2, experimental studies were included that compared different sequences of RE and VE in science education, specifically in the same three science domains physics, chemistry, or biology.

**Search Query**

An initial nonsystematic search of articles in this research field yielded approximately 30 papers that showed naming inconsistencies between the different articles for the same ideas. Based on this insight, terms describing similar ideas were connected by the Boolean operator “OR” in the search query, whereas terms describing different ideas were connected by the Boolean operator “AND.” Words were abbreviated using the “*” to include a wider range of different versions of wording used in the previous literature (see Figure 1 for the composition of the search query used in this review).

The full search query therefore consisted of “(real OR hands-on OR physical) AND (online OR virtual OR simulat* OR computer OR interactive) AND (lab* OR experiment* OR manipulati* OR variable* OR environment*) AND (combin* OR blend* OR sequenc* OR together OR both) AND (learn* OR education*) AND (objective* OR outcome* OR science OR physics OR chemistry OR biology OR concept*).”
Information Sources and Search Restrictions

This full search query was used to execute the search in the databases ERIC, APA PsycInfo (via EBSCOhost), Scopus, and WebOfScience in September 2020. The search was restricted to the years 2000 to 2020 (except for ERIC, where 2000–2020 was not possible; therefore, there was no time restriction in the ERIC search query). This time frame was chosen to ensure that the technologies used in the single studies as well as students’ technical knowledge were not too different from each other; therefore, only studies from the last 20 years were included. Results were restricted to English and German articles only.

Study Selection

The steps we followed to select relevant publications are depicted in an adapted PRISMA flow diagram shown in Figure 2.

The initial database search led to \( n = 9,839 \) records, of which \( n = 607 \) were from ERIC, \( n = 675 \) were from APA PsycInfo, \( n = 4,437 \) were from Scopus, and \( n = 4,120 \) were from WebOfScience. These records were imported into the literature management tool Mendeley. They were checked for duplicates with the Mendeley “check for duplicates” algorithm; the suggested duplicates were removed by hand afterwards. After duplicates were removed, there were \( n = 7,546 \) records remaining that were then screened for eligibility. This screening was based on the eligibility criteria described above in a hierarchical order: articles were screened first for subject area (whether they described science learning), next for whether they reported combinations of RE and VE, and then for whether they were experimental studies. Last, we evaluated the study design for whether it addressed RQ1 and/or RQ2. We excluded \( n = 7,257 \) records based on a screening of the titles or abstracts because they either did not focus on science learning or did not describe a combination of RE and VE. The remaining \( n = 289 \) records were checked for full-text availability. We excluded \( n = 39 \) records because no full text was available. Another \( n = 149 \) records were excluded based on a full-text screening. We then assessed the \( n = 101 \) articles that were identified as potentially relevant for the review for eligibility by focusing on their methodology. For the classification and

*FIGURE 1. Composition of the search query for the review. The Boolean operators “OR” (within the boxes) and “AND” (between the boxes) were used.*
analysis of the full-text articles based on methodology, we used *Microsoft Excel* as a data management tool. The assessment of the 101 articles was done by three researchers individually (initial interrater agreement of 87.1%). All studies were then discussed until there was 100% agreement on which studies to include in the review. As a result, a total of $n = 70$ articles were excluded based on their methodology. These reasons for exclusion are depicted in Figure 2. A detailed overview of the articles excluded based on their methodology is presented in Supplemental Table S1 (available in the online version of this article). The $n = 31$ experimental studies from the database search that met all the inclusion criteria of this review were then used for a backward and forward search of their references. We also performed a backward and forward search on the $n = 11$ theoretical and overview articles that were excluded from this review earlier. With this additional search we identified another $n = 11$ experimental studies that matched the inclusion criteria. Thus, the resulting $N = 42$ empirical papers were included in this systematic review. Of these 42 articles, 24 articles addressed exclusively RQ1, six articles addressed RQ1 as well as RQ2, and 12 articles addressed exclusively RQ2. This led to a total of 30 papers for RQ1 and 18 papers for RQ2. The articles were then sorted by their study design to group them for the study analysis. The coding of the articles on the critical variables was done by two researchers individually, with an immediate interrater agreement of 100% as the variables we assessed could be drawn from the articles in an objective way.
Results

Description of the Empirical Studies Included in the Review

Table 1 gives an overview of descriptive features and outcomes of the 42 articles included in this review. A more detailed version of Table 1 is presented in online Supplemental Table S2.

Publication Year

Most of the articles included in the review (36 out of 42) were published between 2006 and 2020; the remaining six articles were published between 2000 and 2003. The continuous publication of 1 to 4 studies per year since 2006 in this field of research, with a peak of six publications in 2014, indicates a stable interest in research dealing with combinations of RE and VE in the past years.

Research Designs

We coded the research designs according to the categories suggested by Slavin and Lake (2008). Most of the included studies are randomized experiments (random assignment to conditions on student level; 19 out of 42) or randomized quasi-experiments (random assignment to conditions on class level; 13 out of 42). Three studies are matched studies where students were assigned to conditions based on prior testing; for seven studies, the research design was not specified in the article.

Study Designs

The study designs of the included studies show a lot of variance: For their experimental groups, studies assigned to RQ1 reported either an unspecified combination sequence of RE and VE (“RE + VE”) or a certain sequence of experimentation (e.g., “VE-RE”, “RE-VE,” or multistep combinations such as “RE-VE-RE” or “VE-RE-VE”). In half of the RQ1 studies (15 out of 30), the combination of RE and VE was tested against a single control group that used RE. Only one study reported using a VE as the control condition. The remaining 14 studies incorporated two control conditions, RE only and VE only. Each of the 24 studies exclusively assigned to RQ1 reported only one experimental condition with a combination of RE and VE. The six studies assigned to RQ1 as well as RQ2 reported at least two different sequences of combinations. For RQ2, 16 out of the 18 studies deployed a design based on only two groups where the sequences “RE-VE” and “VE-RE” were compared with each other. However, within the six studies assigned to both RQ1 and RQ2, two studies focused on multistep sequences with three experimentation phases (e.g., “RE-VE-RE”).

The number of participants per group varied substantially between the included studies from a minimum of 12 to a maximum of 139 participants in a group. This aspect can be considered to judge the quality of single studies by their statistical power.

Learning Domain and Topic

The vast majority of included papers (32 out of 42) dealt with learning in physics. Only five papers in biology and five papers in chemistry were identified as


| RQ | Author(s), (year) | Research design | Study design (n students per group) | Discipline | Sample/participants | Variables measured | Main findings |
|----|------------------|-----------------|----------------------------------|------------|---------------------|-------------------|--------------|
| RQ1 | Campbell et al. (2002) | RanEx | RE (n = 17) vs. RE + VE (n = 22) | Phy | Electric circuits | 39 | University, undergraduate (college) | 2nd year | USA | Conceptual understanding |
| RQ1 | Le (2015) | n/a | RE (n = 21–29) vs. RE + VE (n = 21–29)³ | Phy | Electric circuits | 42/58³ | University, undergraduate | 2nd or 3rd year | USA | Conceptual understanding, student satisfaction, student core skills |
| RQ1 | Kollöffel & de Jong (2013) | RanEx | RE (n = 23) vs. RE + VE (n = 20) | Phy | Electric circuits | 43 | Secondary vocational engineering education | 16–22 years, \( M = 19.17 \) years, \( SD = 1.39 \) | Netherlands | Conceptual understanding, procedural skills |
| RQ1 | Ronen & Eliahu (2000) | RanQEx | RE (n = 74) vs. RE + VE (n = 71) | Phy | Electric circuits | 145 | Middle school | 9th grade, about 15 years | Israel | Conceptual understanding |
| RQ1 | Huppert et al. (2002) | RanQEx | RE (n = 99) vs. RE + VE (n = 82) | Bio | Growth of microorganisms | 181 | High school | 10th grade | Israel | Conceptual understanding, procedural skills |
| RQ1 | Darrah et al. (2014) | RanEx | RE vs. VE vs. RE + VE (in both substudies)³ | Phy | Mechanics (dynamics) and ideal gas law | 224³ | University, undergraduate | 1st year | USA | Conceptual understanding |
| RQ1 | Farrokhnia & Esmailpour (2010) | n/a | RE (n = 30) vs. VE (n = 35) vs. RE + VE (n = 35) | Phy | Electric circuits | 100 | University, undergraduate | n/a | Tehran | Conceptual understanding, procedural skills |

(continued)
| RQ | Author(s), (year)                  | Research design           | Study design (n students per group) | Discipline | Sample/participants | Variables measured                                      | Main findings                                                                 |
|----|----------------------------------|---------------------------|-------------------------------------|------------|---------------------|--------------------------------------------------------|------------------------------------------------------------------------------|
| RQ1 | Gumilar et al. (2019)            | n/a                       | RE (n = 27) vs. VE (n = 27) vs. RE + VE (n = 27) | Phy        | Electric circuits   | N_{total} = 81, Educational level = Senior high school, Grade/age = 1st year, Country = Indonesia | Combined conceptual understanding and critical thinking skills                |
|     | Zacharia & Michael (2016)        | RanQEx                    | RE (n = 18) vs. VE (n = 18) vs. RE + VE (n = 19) | Phy        | Electric circuits   | N_{total} = 55, Educational level = Elementary school, Grade/age = 6th grade, Country = Cyprus | RE + VE > RE = VE                                                             |
|     | Olympiou & Zacharia (2012)       | RanEx                     | RE (n = 23) vs. VE (n = 23) vs. RE + VE (n = 24) | Phy        | Light and color     | N_{total} = 70, Educational level = University, undergraduate, Grade/age = 1st year, M = 18.3 years, SD = 0.87, Country = Cyprus | Conceptual understanding, RE-VE > RE = VE                                     |
|     | Olympiou & Zacharia (2014)       | RanEx                     | RE (n = 42) vs. VE (n = 36) vs. RE + VE (n = 36) (Study 2) | Phy        | Light and color     | N_{total} = 114, Educational level = University, undergraduate, Country = Cyprus | Conceptual understanding, RE-VE > RE = VE                                     |
|     | Zacharia (2007)                  | RanEx                     | RE (n = 43) vs. RE-VE (n = 45)       | Phy        | Electric circuits   | N_{total} = 88, Educational level = University, undergraduate, Grade/age = 20–22 years, Country = Cyprus | Conceptual understanding, RE-VE > RE = VE                                     |
|     | Zacharia et al. (2008)           | RanEx                     | RE (n = 31) vs. RE-VE (n = 31)       | Phy        | Heat and temperature | N_{total} = 62, Educational level = University, undergraduate, Country = Cyprus | Conceptual understanding, RE-VE > RE = VE                                     |
|     | Abdulwahed & Nagy (2009)         | MatchSt                   | RE (n = 30–32) vs. VE-RE (n = 18)   | Che        | Process control with surge tank system | N_{total} = 48/50, Educational level = University, undergraduate, Grade/age = 2nd year, Country = UK | Conceptual understanding, procedural skills, VE-RE > RE (for both variables) |
|     | Bortnik et al. (2017)            | n/a                       | RE (n = 25) vs. VE-RE (n = 25)       | Che        | Potentiometry and photoelectron-colorimetry | N_{total} = 50, Educational level = University, undergraduate, Grade/age = 3rd year, Country = Russia | Conceptual understanding, scientific literacy, VE-RE > RE (for both variables) |
|     | Climent-Bellido et al. (2003)    | n/a                       | RE (n = 139) vs. VE-RE (n = 135)    | Che        | Distillation and mixture separation problems | N_{total} = 274, Educational level = University, undergraduate, Grade/age = 1st year, Country = Spain | Conceptual understanding, procedural skills, VE-RE > RE (for both variables) |

(continued)
TABLE 1 (continued)

| RQ  | Author(s), (year) | Research design | Study design (n students per group) | Discipline | Sample/participants | Variables measured | Main findings |
|-----|-------------------|-----------------|-------------------------------------|------------|---------------------|-------------------|---------------|
| RQ1 | Manunure et al. (2020) | RanQEx | RE (n = 24) vs. VE-RE (n = 25) | Phy | Electric circuits | 49 | Secondary school | 12-14 years, M = 12.73 years, SD = 0.67 | Zimbabwe | Conceptual understanding | VE-RE > RE |
| RQ1 | Zacharia & Anderson (2003) | RanEx | RE (78 cases) vs. VE-RE (78 cases), self-control design with an alternating pattern of RE and VE-RE for each student | Phy | Mechanics, waves/ optics, thermal physics | 13 | University, postgraduate (4 in-service + 9 preservice science teachers) | 24-47 years, M = approx. 30 years | n/a | Conceptual understanding | VE-RE > RE |
| RQ1 | Makransky et al. (2016) | RanEx | RE (n = 94) vs. VE-RE (n = 95) | Bio | Streaking out bacteria, isolating them | 189 | University, undergraduate | M = 20.2 years | Scotland | Conceptual understanding, intrinsic motivation, self-efficacy | RE = VE-RE (for all variables) |
| RQ1 | Pineda (2015) | RanEx | RE (n = 18) vs. VE-RE (n = 17) | Phy | Induction | 35 | Community college | >18 years | USA | Conceptual understanding | RE > VE-RE, but students did not learn much from RE (in both groups) |
| RQ1 | Wang & Tseng (2018) | RanQEx | RE (n = 69) vs. VE (n = 69) vs. VE-RE (n = 70) | Che | Evaporation and condensation of water | 208 | Elementary school | 3rd grade, 8-9 years | Taiwan | Conceptual understanding, domain knowledge | VE-RE > VE > RE (for conceptual understanding) and VE-RE = VE > RE (for domain knowledge) |

(continued)
| RQ1 | Ünlü & Dökme (2011) | RE (n = 21) vs. VE (n = 18) vs. VE-RE (n = 27) | Phy | Electric circuits | 66 | Elementary school | Turkey | Conceptual understanding | VE-RE > VE = RE (for domain knowledge) and VE-RE > RE (for conceptual knowledge) |
| RQ1 | Jaakkola & Nurmi (2008) | RE (n = 22) vs. VE (n = 22) vs. RE-VE (n = 22) | Phy | Electric circuits | 66 | Elementary school | Finland | Conceptual understanding | VE-RE > VE > RE (for conceptual understanding) and VE-RE > RE (for domain knowledge) |
| RQ1 | Jaakkola et al. (2011) | VE_implicit (n = 12) vs. VE_explicit (n = 14) vs. RE-VE_implicit (n = 12) vs. RE-VE_explicit (n = 12) | Phy | Electric circuits | 50 | Elementary school | Finland | Conceptual understanding | VE-RE_implicit = VE-RE_explicit > VE_explicit > VE_implicit |
| RQ1 | Raman et al. (2014) | RE (n = n/a) vs. VE-RE (n = n/a) vs. RE-VE (n = n/a) vs. VE-RE (n = n/a) | Phy | Electric circuits | 246 | University, 1st year | India | Conceptual understanding, innovation attributes | RE-VE = VE-RE > RE (for conceptual understanding) and VE-RE > RE-VE (for innovation attributes) |
| RQ1 | Zacharia & Olympiou (2011) | RE (n = 56) vs. VE (n = 59) vs. RE-VE (n = 34) vs. VE-RE (n = 33) vs. no experiment, CG (n = 52) | Phy | Heat and temperature | 183 | University, 2nd year | Cyprus | Conceptual understanding | RE = VE = VE-RE > CG (but not for all experiments) and RE-VE > VE-RE (partly) and VE-RE > RE-VE (partly) |
| RQ1 | Atanas (2018) | RE (n = n/a) vs. VE (n = n/a) vs. RE-VE (n = n/a) vs. VE-RE (n = n/a) | Phy | Electromagnetism and optics | 183 | University, 2nd year | United Arab Emirates (UAE) | Conceptual understanding | RE = VE = RE-VE > CG (but not for all experiments) and VE-RE > RE-VE (partly) and RE-VE > VE-RE (partly) and VE-RE > RE-VE (partly) |
| RQ   | Author(s), (year)            | Study design | Discipline | Sample/participants | Variables measured | Main findings                                                                 |
|------|-----------------------------|--------------|------------|---------------------|-------------------|-------------------------------------------------------------------------------|
| RQ1  | Akpan & Andre (2000)        | RanQEx       | Bio        | Frog dissection     | 81/82             | Middle school 7th grade, 13–15 years | Conceptual understanding, procedural skills, attitudes                       |
| RQ1  | Kapici et al. (2019)        | RanQEx       | Phy        | Electric circuits   | 143               | Middle school 7th grade, 12–14 years | Conceptual understanding, inquiry skills                                      |
|      | Zacharia & de Jong (2014)   | RanEx        | Phy        | Electric circuits   | 194               | University, 2nd year, $M = 20.4$, $SD = 0.87$ | No difference between conditions. But interplay between manipulative and circuit type: Simple circuits: RE = VE; complex circuits: VE > RE, and VE before RE is advantageous in some cases |
| RQ2  | Chini et al. (2012)         | RanEx        | Phy        | Pulleys             | 121               | University, n/a undergraduate USA | Conceptual understanding RE-VE = VE-RE                                        |

(continued)
| RQ   | Author(s), (year)                  | Research design | Study design (n students per group) | Discipline | Sample/participants | Variables measured | Main findings                                      |
|------|-----------------------------------|-----------------|-------------------------------------|------------|---------------------|-------------------|----------------------------------------------------|
| RQ2  | Chini (2010)                      | RanQEx          | RE-VE vs. VE-RE (in all four substudies)  | Phy        | University, n/a, high school | Conceptual understanding | RE-VE = VE-RE (in all four substudies) |
|      | Myneni et al. (2013)              | RanEx           | RE-VE (n = 78) vs. VE-RE (n = 80) (Study 2) | Phy        | University, n/a, high school | Conceptual understanding | RE-VE = VE-RE |
| RQ2  | Sullivan et al. (2017)            | RanQEx          | RE-VE (n = 55) vs. VE-RE (n = 45)     | Phy        | Middle school, 8th grade | Conceptual understanding | RE-VE = VE-RE |
| RQ2  | Liu (2006)                        | RanQEx          | RE-VE (n = 18) vs. VE-RE (n = 15)     | Che        | High school, n/a, high school | Conceptual understanding | RE-VE = VE-RE (for conceptual understanding) |
| RQ2  | Saleh et al. (2014)               | RanEx           | RE-VE (n = 16) vs. VE-RE (n = 16)     | Phy        | University, n/a, high school | Conceptual understanding | RE-VE = VE-RE |
| RQ2  | Toth et al. (2009)                | RanEx           | RE-VE (n = 19) vs. VE-RE (n = 20)     | Bio        | University, 1st year, >18 years | Conceptual understanding | RE-VE = VE-RE |
| RQ2  | Gire et al. (2010)                | n/a             | RE-VE (n = 71) vs. VE-RE (n = 61)     | Phy        | University, n/a, high school | Conceptual understanding | RE-VE > VE-RE |
| RQ2  | Smith & Puntambekar (2010)        | RanQEx          | RE-VE (n = 43) vs. VE-RE (n = 17)     | Phy        | Middle school, 6th grade | Conceptual understanding | RE-VE > VE-RE |
| RQ2  | Tsihouridis et al. (2015)         | n/a             | RE-VE (n = 32) vs. VE-RE (n = 33)     | Phy        | High school, 3rd year | Conceptual understanding | RE-VE > VE-RE |
| RQ2  | Toth et al. (2014)                | RanQEx          | RE-VE (n = 12) vs. VE-RE (n = 16)     | Bio        | University, 1st year, >18 years | Combined conceptual and design knowledge, inquiry skills | VE-RE > VE-RE (for both variables) |
| RQ2  | Achuthan et al. (2017)            | RanEx           | RE-VE (n = 73) vs. VE-RE (n = 72)     | Phy        | University, n/a, high school | Conceptual understanding | VE-RE > VE-RE > CG |

Note. RQ = research question; RanEx = randomized experiment; RanQEx = randomized quasi-experiment; MatchSt = matched study; RE = real experiment; VE = virtual experiment; CG = control group; Phy = physics; Bio = biology; Che = chemistry; RE + VE = unspecified combination of RE and VE; RE-VE = sequence, RE followed by VE.

aThe number of participants could not clearly be determined from the paper due to missing information or inconsistent reporting.
matching the inclusion criteria. Half of the physics papers (16 out of 32) reported studies about learning in the domain of electric circuits. Another six papers dealt with learning in the domain of pulleys. That the same learning domains reappeared in multiple studies is reasonable as it is important to perform comprehensive research in one domain to make meaningful claims about the findings.

When comparing the use of RE and VE for understanding concepts in specific domains, it is interesting to evaluate whether RE and VE were both used to convey the same body of conceptual information within a certain domain or whether they focused on different variables and/or phenomena. In this review, all 42 included studies used RE and VE for the same topic, and sometimes the VE was even performed identically to the corresponding RE (e.g., Akpan & Andre, 2000; Zacharia & Michael, 2016). Sometimes the VE differed slightly in the variables tested or in the setup from the RE, but still referred to the same domain topic and concepts (e.g., Atanas, 2018; Salehi et al., 2014). This makes the studies and their results more comparable as the focus is not on the general method of combining the different modalities of RE and VE but on the learning of certain topics by using these different modes.

Simulations Used as VE

Despite the low variation in domains that are represented in the included studies, the simulations chosen as VE in the studies differed much more: Of the 16 papers dealing with electric circuits, three studies used the “PhET Circuit Construction Kit,” three studies used the “Virtual Labs Electricity software,” two studies used the “Electricity Exploration Tool,” and all the other eight simulations used for this topic differed from each other. However, for the six pulley studies it was always the “CoMPASS online hypertext system” (Concept Mapped Project-based Activity Scaffolding System) or “Virtual Physics System” (ViPS). The similar results obtained from different VE in studies dealing with the same learning domain show that there may be a certain generalizability of the results independent of the specific simulation.

Participants

The educational level, grade, and age of the participants in the different studies vary from third-grade elementary school students aged 8 to 9 years to postgraduate university preservice and in-service teachers aged 24 to 47 years. Most of the studies were conducted with postsecondary participants (27 out of 42) and much fewer with high school students (4 out of 42) or middle school and elementary school students (11 out of 42).

The studies included in this review were foremost conducted in the United States of America (15 studies out of 42); the second most frequent country of study was Cyprus (seven studies).

Time Point of RE and VE and Time on Task for the Individual Groups

In 17 of the 42 studies, the students conducted RE and VE on the same day in a direct sequence. Sixteen other studies used the different experiment types on different days, in subsequent sessions within one learning unit. Here weekly sessions were the most common session format. The nine remaining studies did not
Combining Real and Virtual Experiments

report on this aspect of timing. We found no pattern of results evolving out of the time point of RE and VE, which suggests that knowledge integration is no better whether the different experiment types are conducted in a direct sequence or not.

Among the papers assigned to RQ1, nine studies included a different time on task between the experimental and the control groups. We coded time on task by examining whether the students spent the same amount of time experimenting in any type of experiment whatsoever. Solving textbook problems related to the topic was not counted as time on the experiment. As with the time point of RE and VE, no pattern could be found that relates time on task to the outcome of the study. Therefore, we suppose that there is no direct influence of time on task on students’ conceptual understanding in these studies, where the focus was more on the different experimental modes that students were offered.

Variables Measured

The variable measured in all 42 included studies was conceptual understanding. Other variables measured in some of the studies were procedural experimentation skills (five studies), inquiry skills (three studies), domain knowledge (two studies), student satisfaction (one study), student core skills (one study), scientific literacy skills (one study), intrinsic motivation (one study), self-efficacy (one study), innovation attributes (one study), attitudes (one study), and understanding of models in science (one study). For more clarity, each of these variables is explained briefly in the following. Procedural experimentation skills refer to how well a student can conduct the specific experiment used in the study (or similar experiments). Inquiry skills address how well a student can plan, conduct, and discuss experiments in general. Domain knowledge is factual knowledge in a specific domain, which does not necessarily include knowledge about relations between core concepts in this domain and therefore is different from conceptual understanding. Student satisfaction concerns the question of how satisfied the students were with their learning experience with the experiments. Student core skills are defined as a cumulative construct consisting of design and professional skills, development of teamwork and social skills, development of analytical, report writing, and presentation skills (Le, 2015). Scientific literacy skills incorporate knowledge and skills in science but also procedural skills and the knowledge about scientific practices. Intrinsic motivation and self-efficacy are related to the respective subject (physics, biology, or chemistry) of the study. Innovation attributes are a subjective rating of the degree of innovation that students gave to their laboratory experience (Raman et al., 2014). Attitudes describe students’ attitudes toward the practice of dissection in this case (Akpan & Andre, 2000). Understanding of models in science refers to students’ general understanding of the function and use of models in science.

Potential Moderators and Boundary Conditions

Some potential moderators frequently mentioned for further research but not systematically investigated were learner characteristics (e.g., prior knowledge and prior conceptions, age and developmental level, gender, spatial ability, experience, interest), features of the learning material (e.g., difficulty of the learning material and complexity of the concept, goals and designed affordances of the
individual physical and virtual activities, directiveness, and fidelity of the simulation), and degree of guidance (e.g., scaffolds; lab manuals; worksheets; virtual hypertext; aid given by teachers, instructors, or technicians; video modeling or tutorials; introductory presentations; or training with the VE prior to the intervention).

All of the studies included guidance for the learners to some degree, but most did not report in detail the concrete boundary conditions of the learning setting with respect to guidance, and if they did, they most often did not take this factor into account when discussing their results. Only one of the studies (Jaakkola et al., 2011) systematically investigated effects of scaffolds, guidance, or instructional support.

Problems Frequently Reported

Some problems hindering learning were frequently mentioned in the included studies’ discussions and limitations. Those included participants’ low prior knowledge and skills in handling digital devices and especially VE, a very heterogeneous degree of prior knowledge in the topic between students, insufficient reliability of test instruments, and small numbers of participants in several studies.

Main Findings

Main Findings About Combined RE and VE Compared to a Single Experiment Type

RQ1 addresses whether combinations of RE and VE result in greater learning outcomes than RE or VE alone. Results for the 30 papers assigned to RQ1 are described in more detail in the following paragraphs. The single studies are therefore clustered by their study design. Studies that used similar designs are analyzed and compared directly in one subsection.

Findings in studies contrasting RE versus RE + VE. Campbell et al. (2002), Le (2015), Kollöffel and de Jong (2013), Ronen and Eliahu (2000), and Huppert et al. (2002) reported combinations or blends of RE and VE, where the sequence of the experiments was not specified in the paper. There might even be no clear sequence in these studies because the students used RE and VE simultaneously (e.g., Huppert et al., 2002) or in an individual sequence of their choice (e.g., Ronen & Eliahu, 2000) in the blended combination. These combinations of RE and VE were tested against RE only. Campbell et al. (2002) and Le (2015) compared two groups of college students, Kollöffel and de Jong (2013) secondary vocational engineering education students, and Ronen and Eliahu (2000) ninth-grade middle school students in their learning with electrical circuits in physics. All four studies found that the combined lab conditions scored significantly better on tests of conceptual understanding and had some advantages in acquiring certain procedural skills. In line with these findings, Huppert et al. (2002) also found an advantage of the combination group in their study. They compared 10th-grade students’ learning about the growth of microorganisms in biology. The explanation for the findings were similar across all five studies: Huppert et al. (2002) reasoned that the additional VE simplified the process for the students by displaying results visually and immediately, so that many simulations could be performed in a short
time but at the own pace of each student. Ronen and Eliahu (2000) argued that the addition of a VE to an RE helps students bridge the gap between theoretical idealized models, their formal representations, and reality and enhances students’ understanding of the underlying theoretical principles. According to Kollöffel and de Jong (2013), apart from VEs’ advantage of connecting reality and theoretical concepts, RE are also crucial in students’ education, so RE should not be replaced but rather enhanced with VE.

Findings in studies contrasting RE versus VE versus RE + VE. The second cluster of studies (Darrah et al., 2014; Farrokhnia & Esmailpour, 2010; Gumilar et al., 2019; Olympiou & Zacharia, 2012, 2014; Zacharia & Michael, 2016) compared a blended combination of RE and VE to a control group with RE as well as to a control group with VE. Darrah et al. (2014) reported no differences between the outcomes of their three groups. Their paper dealt with undergraduate university students’ learning of mechanics and the ideal gas law. In contrast, all the other five studies with this three-group design found that the use of a blended combination of RE and VE enhanced students’ conceptual understanding more than the use of either RE or VE alone. Farrokhnia and Esmailpour (2010), Gumilar et al. (2019), and Zacharia and Michael (2016) investigated students’ learning of electrical circuits in different educational levels (undergraduate university students, senior high school students, and sixth-grade elementary school students). Olympiou and Zacharia (2012, 2014) focused in their two studies on undergraduate university students’ learning in the domain of light and color. In all five aforementioned studies, the authors argued that RE and VE should be combined because of their unique affordances. In addition to the specific affordances of RE and VE that were already presented in the introduction of this article, they mentioned that RE allow for reflecting the true nature of science, including, for example, measurement errors. On the other hand, VE allow for a variety of measurement opportunities, immediate and observable feedback, and faster setting up and conducting of an experiment. This leads to more focus on the conceptual issues rather than procedural issues of the experiment (e.g., Gumilar et al., 2019; Olympiou & Zacharia, 2014; Zacharia & Michael, 2016).

Findings in studies contrasting RE versus a single sequence of RE and VE. All papers summarized in this subsection except for Zacharia (2007) and Zacharia et al. (2008) report a study where an RE is compared with a VE-RE sequence. In contrast, Zacharia (2007) and Zacharia et al. (2008) report studies where an RE is compared with an RE-VE sequence. In these two studies, undergraduate university students learned about the topic of electric circuits (Zacharia, 2007) or heat and temperature (Zacharia et al., 2008). Both studies found a significant advantage of learning with RE followed by VE compared with learning with RE alone. The explanation for these results was based on the reasoning that the experiment types had complementary benefits. Moreover, Zacharia et al. (2008) explained that the VE provides the learner with additional representations that contribute to and build on the learning experience with the RE, making this approach of RE-VE especially fruitful for learning. With the RE first, students were enabled to contextualize their learning experience with the RE, whereas with the VE, they
could expand their new knowledge and integrate it in their prior knowledge from class (Zacharia et al., 2008).

On the other hand, also the studies using a VE-RE sequence showed positive effects of this combination on different outcome variables of students’ achievement. Abdulwahed and Nagy (2009), Bortnik et al. (2017), and Climent-Bellido et al. (2003) conducted studies with undergraduate university students in the domain of chemistry. The domain topics included process control with a surge tank system (Abdulwahed & Nagy, 2009), potentiometry and photoelectrocolorimetry (Bortnik et al., 2017), and distillation (Climent-Bellido et al., 2003). Likewise, Manunure et al. (2020) reported an advantage of VE-RE compared with RE for secondary school students’ gain of conceptual understanding in the domain of electric circuits. In the study of Zacharia and Anderson (2003), the learning of preservice and in-service science teachers in the domains of mechanics, waves and optics, and thermal physics was evaluated in a self-control design with an alternating pattern of RE and VE-RE for each participant. The use of a VE in the combined experiments fostered conceptual change in the physics area studied compared with the single RE. The two authors explained this result with the advantage of VE to help students develop an insight about abstract physics concepts, which prepares students for the RE in inquiry-based learning.

The same line of argumentation was presented by Makransky et al. (2016). They investigated the effectiveness of using VE as preparation for RE for undergraduate university students’ learning about microbiology and streaking out as well as isolating bacteria. Even though the study did not find significant differences between the conceptual understanding of the VE-RE and the RE group, Makransky et al. (2016) concluded that the VE was an effective way to prepare students for the RE because the students gained the basic knowledge and the cognitive skills needed for the RE beforehand. The students practiced the technique of striking out bacteria and isolating them, and the VE allowed them to instantly observe the result. This allowed the students to direct all their cognitive resources toward the relevant activity in the RE afterwards.

In the doctoral thesis of Pineda (2015), learning with a VE preceding the RE in one group was compared to a second group that received an overview presentation preceding the RE. Pineda (2015) investigated community college students’ learning about induction in the domain of physics. Contrary to all the other findings presented in this systematic review, Pineda (2015) found an advantage for the students who learned without a simulation, but with an overview presentation before performing the RE. However, she reported that the RE did not seem to make a substantial contribution to the students’ learning in either group, but it was rather the overview presentation that aided conceptual understanding. In her explanation for this result, she mentions that the VE-RE group had trouble managing the complexity of the VE due to lack of time and lack of familiarity with computer simulations. On the other hand, the overview presentation provided the students in the RE group with a step-by-step multimedia-based explanation of the topic and guidance for the RE. This study provides multiple points for discussion: First, it includes only a very small sample consisting of 35 students split up into two groups and therefore has low statistical power. Second, the single experiment group received a structured introduction as guidance to the experiment which the
Combining Real and Virtual Experiments

combination group did not receive; the combination group was rather pushed into an experimenting situation with the VE. Third, the overview presentation even turned out to be the main source of learning for the students in the RE group. Therefore, the results of Pineda’s (2015) dissertation should be considered with reservations.

Findings in studies contrasting VE (vs. RE) versus a single sequence of RE and VE. Four studies (Jaakkola et al., 2011; Jaakkola & Nurmi, 2008; Ünlü & Dökme, 2011; Wang & Tseng, 2018) investigated whether the combination of VE-RE was more fruitful for learning compared to VE only, and in the case of Wang and Tseng (2018), Ünlü and Dökme (2011), and Jaakkola and Nurmi (2008) also to RE only. Whereas most of the studies mentioned in the preceding subsection explained the benefits of a sequence of RE and VE as being due to the affordances of the VE, those studies did not compare the combination to the use of a single VE. This is the focus of studies reviewed here. All four studies in this subsection incorporate participants from elementary schools. Wang and Tseng (2018) investigated conceptual understanding and domain knowledge in the topic of state changes of water. They found that using VE-RE or VE alone enhanced students’ knowledge gains more than the RE alone. The authors owe this to the possibility of making clear observations of invisible phenomena in the VE, which helps the students develop a conceptual model. Moreover, they found that VE-RE promoted students’ conceptual understanding more than either VE or RE alone, and VE alone was more beneficial for students’ conceptual understanding than RE alone. They explained that the VE first helped students understand the underlying mechanisms of complex phenomena, and the RE afterwards highlighted different aspects of the content and provided the opportunity to add micro details to the students’ understanding while providing the students with an authentic experience of the experiment. Therefore, the combination could bridge the gap between theory and reality for the students (Wang & Tseng, 2018).

The other three studies covered learning of electrical circuits. Ünlü and Dökme (2011) found that a VE-RE combination resulted in greater learning acquisition than RE or VE did alone, while there was no difference between RE or VE alone. Results were explained similarly to Wang and Tseng (2018). In line with these results, Jaakkola and Nurmi (2008) also found that a combination of VE-RE promoted conceptual understanding and domain knowledge better than either RE or VE alone, with the VE group acquiring more conceptual understanding than the RE group. Jaakkola et al. (2011) in a later study compared a VE-RE sequence to VE, with either implicit or explicit instructions for the experiments. It turned out that the combination of RE and VE led to better understanding even if it was not supported by explicit instructions compared with both VE with and without explicit instructions. Still, in the VE conditions, students with explicit instructions gained more conceptual understanding than the students with a VE and only implicit instructions.

Findings in studies contrasting multiple sequences of RE and VE. Finally, there are six studies addressing RQ1 as well as RQ2 that incorporated different combinations of RE and VE. This allows to examine whether one sequence compared with
another would lead to a higher learning outcome contrasted with RE or VE alone. These studies showed diverging results: Raman et al. (2014) investigated conceptual understanding and innovation attributes between three groups of undergraduate university students in the topics of magnetism, mechanics, and optics. Two groups learned with the sequences RE-VE or VE-RE, and the third group learned exclusively with RE. They reported a significant advantage of both sequences for conceptual understanding compared with the RE group. Likewise, Atanas (2018) found an advantage to undergraduate university students’ learning outcome in the domain of electromagnetism and optics when they performed experiments with different sequences of RE and VE compared with performing experiments with only RE or only VE. The sequence VE-RE or RE-VE did not matter. Kapici et al. (2019) investigated two sequences with three phases: VE-RE-VE and RE-VE-RE. Middle school students’ conceptual understanding of electric circuits and inquiry skills in the combination groups were higher in a posttest compared to the VE and the RE groups. Akpan and Andre (2000) compared middle school students’ learning with RE, VE, RE-VE, and VE-RE when performing a frog dissection. Different from the three studies described above within this paragraph, here the VE-RE group as well as the VE group performed significantly better than the RE-VE or the RE group in terms of conceptual understanding. Moreover, the VE-RE group outperformed all the other three groups when considering conceptual understanding and procedural skills together. In this case, the combination per se did not lead to better learning outcomes, but rather it was the aspect of performing the VE first, independent of whether an RE followed or not. For the two other studies comparing multistep combinations to single experiments, there were no significant differences in posttests of conceptual understanding between the groups (Zacharia & de Jong, 2014; Zacharia & Olympiou, 2011). Zacharia and Olympiou (2011) added a control condition where the university students did not perform an experiment at all when learning about heat and temperature. All the four experimenting conditions (i.e., RE, VE, RE-VE, and VE-RE) equally promoted students’ understanding, and these four conditions were better than the control condition. Zacharia and Olympiou (2011) concluded that manipulation per se rather than physicality during experimentation was important for students’ learning in this context. Zacharia and de Jong (2014) did not find significant differences in posttests between their four conditions (RE vs. VE vs. VE-RE-RE vs. RE-VE-RE vs. RE-RE-VE) either.

Summary of findings about combined RE and VE compared to a single experiment type. In summary, 25 of the 30 papers assigned to RQ1 reported a significant advantage of the experimental groups that used RE and VE in a combination, compared with control groups that used only one single experiment type for learning. Four studies reported no difference between the combination groups and the single experiment groups, and only one study reported an advantage of the single RE compared with the combination of a VE preceding the RE. So, the reviewed literature shows a clear trend toward combinations of RE and VE being superior to single experiments. Notably, these results are not biased or moderated by the aspects “research design” or “sample size.” When only considering the “gold standard” of research designs, namely, randomized experiment studies, within
Combining Real and Virtual Experiments

our sample, we found the same clear trend as described above. Also, the sample size (even though this number varies a lot between studies) is distributed equally between studies that show evidence for combinations of RE and VE being superior to single experiments and studies that do not find a difference between their experimental conditions. Moreover, the single study reporting an advantage of RE over VE-RE has a very small sample size and thus low power, which again supports our claim for the abovementioned trend.

Main Findings About Different Sequences of Combined RE and VE

Thirteen of the studies that were exclusively assigned to RQ1 reported the sequence of experiments in the combination condition. It is noticeable that 11 out of these 13 studies used a VE-RE sequence rather than an RE-VE sequence. As mentioned before, most of these studies reasoned that the VE needs to be conducted first, because it helps students gain abstract basic knowledge or procedural skills that they could later build on in the RE. Accordingly, starting with the VE in a sequence should support students’ conceptual understanding better than starting with the RE and performing the VE afterwards. Whether this really is the case was addressed in RQ2. The 18 papers (six already reviewed for RQ1 plus another 12 studies that investigated only combinations of RE and VE) assigned to RQ2 compared different sequences of RE and VE to each other. Most of these papers (16 out of 18 studies) simply compared the two sequences RE-VE and VE-RE to each other. The remaining two studies compared multistep combinations.

Findings in studies contrasting RE-VE versus VE-RE. Nine of the 16 studies did not find a significant difference in conceptual understanding for RE-VE and VE-RE. Two of these dealt with undergraduate university students’ learning of magnetism, mechanics, and optics (Raman et al., 2014), and heat and temperature (Zacharia & Olympiou, 2011). Four other studies reported about learning in the domain of pulleys with undergraduate university or middle school students (Chini, 2010; Chini et al., 2012; Myneni et al., 2013; Sullivan et al., 2017). Chini (2010) also investigated learning in the domain of inclined planes in physics. Liu (2006) investigated high school students’ learning of gas laws in chemistry, Salehi et al. (2014) studied undergraduate university students’ learning of electrical circuits in physics, and Toth et al. (2009) examined undergraduate university students’ learning of DNA-gel electrophoresis in biology.

The six studies that did reveal differences between sequences of RE and VE do not yield a clear pattern. Three studies found a superiority of RE-VE over VE-RE (Gire et al., 2010; Smith & Puntambekar, 2010; Tsihouridis et al., 2015). Gire et al. (2010) and Smith and Puntambekar (2010) investigated undergraduate university or middle school students’ conceptual understanding gain when experimenting with pulleys. Gire et al. (2010) explained their result with the salience of different concepts and experiment types. In RE some concepts (e.g., effort force) had a higher salience than others (e.g., work) and therefore captured more of the learners’ attention. In VE, however, the attention could be divided more evenly among the concepts to be learned due to the equivalence in salience. They argued that if the VE was done first, the initial equivalence in salience among different concepts may have lessened the impact of the subsequent kinesthetic experience
with the RE. Thus, the sequence RE-VE should be preferred for this context of learning. Smith and Puntambekar (2010) explained their result as follows: Students first learned the basic concepts of pulleys with the RE and were then able to test and refine their conceptions in the VE for situations that were either impossible or impractical to conduct in the RE. They concluded that the success of a particular sequence of RE and VE was influenced most importantly by the goals and designed affordances of the individual RE and VE. Tsihouridis et al. (2015) investigated high school students’ conceptual understanding of electric circuits and explained their result similarly to Smith and Puntambekar (2010): RE-VE is more beneficial for learning than VE-RE because the VE with its greater abstraction acts as a halfway step toward the formal abstraction of conceptual understanding.

Another three studies found evidence for the opposite pattern (i.e., VE-RE was superior to RE-VE; Achuthan et al., 2017; Akpan & Andre, 2000; Toth et al., 2014). Toth et al. (2014) performed a study similar to Toth et al. (2009); this time they found a significant difference between groups when learning about DNA-gel electrophoresis favoring a VE-RE sequence. They explained their findings by referring to the same underlying mechanisms as the three previously described studies advocating the sequence RE-VE, but with reversed arguments. According to Toth et al. (2014), starting with the VE helped the students learn basic concepts and skills that could then be successfully applied for knowledge synthesis in the following and more complex RE. Starting with the RE, in contrast, led to less deep and purposeful learning, and the students in the RE-VE condition had difficulties applying their knowledge during the VE. Additionally, Toth et al. (2014) mentioned that the sequence RE-VE did not convince students as they questioned the value of using the simplified VE after working with the RE. Thus, the epistemological value of the single experiment types should also be considered. Akpan and Andre (2000) in their study about middle school students’ learning with frog dissections explained their results as follows: Students in the VE-RE condition could refer to their episodic memory of the VE to make sense of the instructions in the more complex reality of an actual frog in the RE. Also, during the VE, students got valuable scaffolds on how to perform the actual dissection in the RE afterwards. On the other hand, students in the RE-VE condition were unable to form a good memory representation based on the RE because it was too complex. Performing the RE first also engaged the students in discovery learning during the dissection, which led to lower conceptual understanding. The authors suggested that the VE may have sufficiently simplified the complex anatomy of the frog and thus directly taught the students which procedures they should follow in the actual dissection, in the RE. Achuthan et al. (2017) reasoned similarly to Toth et al. (2014) and Akpan and Andre (2000) that VE allows an instructional preview to RE.

Importantly, all of the explanations for the results of the previously presented studies are post hoc explanations without any empirical backup.

As a last study comparing RE-VE and VE-RE, Atanas (2018) described mixed results: They found advantages for each of the sequences RE-VE and VE-RE for different experiments within their study on undergraduate university students’ learning in the domain of electromagnetism and optics.
Findings in studies contrasting sequences other than RE-VE versus VE-RE. The remaining two papers assigned to RQ2 compared multistep combinations, namely, RE-VE-RE versus VE-RE-VE (Kapici et al., 2019) and VE-RE-RE versus RE-VE-RE versus RE-RE-VE (Zacharia & de Jong, 2014). As already described for RQ1, both studies did not find significant overall differences between the sequences. However, Zacharia and de Jong (2014) found an interplay between experiment type and circuit type when comparing undergraduate university students’ learning of electric circuits. For simple circuits RE and VE were equal in promoting students’ understanding. For complex circuits, however, VE before RE promoted students’ understanding better than other conditions where VE was not before RE. Zacharia and de Jong (2014) reasoned that for complex circuits VE before RE helped students build an appropriate conceptual model of current flow that they could use later in the RE phase.

Other nonsignificant tendencies were found by several other studies, in favor of VE-RE (Chini et al., 2012; Sullivan et al., 2017; Toth et al., 2009) as well as in favor of RE-VE (Salehi et al., 2014). This shows again that there seems to be no clear direction in which the results of all these reviewed studies point.

Summary of findings about different sequences of combined RE and VE. The results of these 18 papers are very mixed: Three studies found an advantage of VE-RE (Achuthan et al., 2017; Akpan & Andre, 2000; Toth et al., 2014), whereas three other studies found an advantage of RE-VE (Gire et al., 2010; Smith & Puntambekar, 2010; Tsihouridis et al., 2015). The remaining 12 studies found no difference between different sequences. So far, no clear conclusion can be drawn about which sequence of experiments is the most effective for conceptual understanding in science. Again, as in the results for RQ1, a closer look into the articles when considering their research design and sample size does not reveal any bias in the mixed results for RQ2.

Discussion

In this review, we systematically collated and analyzed studies to answer the following questions: “What is the relative effectiveness of combining RE and VE compared to a single type of experimentation for conceptual understanding in science?” (RQ1) and “Which sequence of RE and VE is most effective for conceptual understanding in science?” (RQ2).

Summary of Evidence

For RQ1, there was overall converging evidence in 25 of 30 studies that a combination of RE and VE promotes science learning and leads to higher conceptual understanding for students at different educational levels and for different disciplines and learning domains compared with students learning with either RE or VE alone. This effect was mostly explained by the specific affordances that RE and VE possess. The authors frequently reasoned that one type of experiment (RE or VE) prepares the students for conducting the second experiment (VE or RE, respectively). Using a combination of RE and VE helps students bridge the gap between theory and practice, by providing two different levels of abstraction. Specifically, the VE provides students with insights that are closer to theory, whereas the RE is
closer to practice. We do not consider the single study with an opposite result (Pineda, 2015) because of its low statistical power and other limitations. The remaining four studies reported no differences between conditions, which could be due to the concrete boundary conditions of the studies. To clarify the exact boundary conditions under which combinations are superior to single experiments or not, further and especially systematic research on this topic is needed.

For RQ2, the studies showed mixed results: Each of the sequences (RE first or VE first) seems to provide different advantages for different learning objectives and subject domains depending on their individual affordances and the function that each of the experiments serves. The authors of these studies provided detailed explanations for why the sequence they found to be more beneficial for learning supported the learning process better than another sequence. Importantly, these explanations were generated only post hoc and no further empirical evidence to verify them was presented. Interestingly, the arguments of different advocates for one sequence of RE and VE or the other fit the characteristics of the different disciplines chemistry, biology, and physics very well. Almost all papers that included studies in the domains of chemistry and biology either reported only the sequence VE-RE or found an advantage for this combination compared to RE-VE. This aligns with the findings that the VE and the RE in these disciplines were most often identical, with the VE being a simplified version of the RE preparing the learners for the more complex reality of the RE. In physics with concepts like force that depend highly on physicality and experiments that slightly differ from RE to VE, the sequence RE-VE sometimes also provides an advantageous choice to provide the learners with an appropriate realistic experience of the experiment. Most of the studies reviewed for RQ2, however, reported no differences between different sequences of RE and VE. Thus, there is so far no evidence that one sequence generally promotes science learning better than another. Here more research is needed that considers the subject domains, learning objectives, and the sequence’s functions for learning systematically.

Work by Robert Slavin and colleagues (e.g., Slavin & Lake, 2008; Slavin & Smith, 2009) suggests that the sample size of each study included in the review and the studies’ research designs may have an effect on the results of research synthesis. Differences between research designs (i.e., randomized experiment, randomized quasi-experiment, matched study, and matched post hoc study) might lead to bias in the results of the study; randomized experiments are the “gold standard” within these different designs (Slavin & Lake, 2008). Slavin and Smith (2009) show that there is a negative correlation between studies’ sample size and effect size and that the differences in effect sizes between studies with small and large sample size are much greater than the differences between randomized and matched experiments. We evaluated the influence of different sample sizes on the results presented in this review and found that a special consideration of the sample sizes does not affect our results. The same holds true when considering the research design within the studies included in our review.

**Limitations**

There are some limitations of this review that need to be considered. First, even though we tried to reduce publication bias (by also including peer-reviewed
Combining Real and Virtual Experiments

combining conference proceedings, book chapters, and dissertations besides journal articles, there could still be unpublished studies that were not written up because they had failed to yield significant results. This is a general problem for reviews or meta-analyses.

Second, eight out of 11 papers that were identified through the backward and forward search of references would have been in the databases that we chose, but with our search query we did not find them. This shows that the search query was not ideal; however, with the very diverse wording used in the literature for our topic of interest it would have been difficult to construct a search query that covers all relevant papers. This is also why we did especially value the backward and forward search of references and performed the screening very carefully to find all the other papers that might also fit our topic of interest but that were not uncovered by our initial search query.

Third, it is obvious that many of the studies included in this review were significantly underpowered due to their small sample sizes. This means that the results of these studies always need to be handled with care. Also, in most papers the effect size was not specified, and in some cases, there was not even enough information reported for calculating the effect sizes retroactively.

Fourth, as reported in the Results section, only one study assigned to RQ1 used a VE as the only control condition, and all the other studies used an RE or both VE and RE as two control conditions. This can be explained by the long tradition of RE in science learning. Being the established tool for laboratory work and well known to students and instructors, the RE seems to be a convenient choice for the baseline of learning from a traditional laboratory.

Fifth, concerning RQ2, there were only two multistep sequences, all the other studies compared the sequences “RE-VE” and “VE-RE.” It is interesting that there has not been more variety in terms of research design concerning other possible sequences of RE and VE. One aspect that might play a role here is that switching between RE and VE always requires some transition time and therefore performing the RE and VE in consecutive blocks is more time efficient and controllable.

Sixth, most of the included studies dealt with learning in physics. This can be explained by the nature of experiments and the goals of inquiry learning in the different sciences physics, biology, and chemistry: Whereas in physics education, inductive learning procedures and the manipulation of variables to explore relations between these variables dominate, chemistry and biology learning is rather deductive and not as explorative. In chemistry and biology experiments, it is more often important to perform procedures in a correct manner to end up with the intended outcome of the experiment (e.g., Abdulwahed & Nagy, 2009; Makransky et al., 2016). This explanation also fits the arguments for certain sequences of RE and VE that were reported in the main findings. The most frequently reported learning domains were “electric circuits” and “pulleys.” This is reasonable due to the invisibility of electric current and the nonnegligible difference between the representational levels of this domain (bulbs and wires in a hands-on experiment, abstract symbolic representations in a schematic circuit diagram). Based on its advantages as described in the introduction, the VE can add value for learning in this domain by making electric current directly observable and by bridging the
gap between the representational levels with a VE that includes both hands-on experiment aspects and aspects of a schematic circuit diagram. “Pulleys” is also a domain where the advantages of a combination of RE and VE are undeniable: Hands-on experimentation with pulleys can be very time intensive due to the effortful process of setting up the experiment, which needs to be repeated for every manipulation that is tested in the experiment. VE are much less time consuming and are easy to handle; still the RE is valuable due to its haptic component of feeling the force needed to lift up a certain weight. Even though there are good reasons for the use of these learning domains, more variety would improve the generalizability of the results to other domains.

Seventh, the participants of the studies were foremost university students and from the United States of America or from Cyprus, and thus, our results may be slightly biased toward specificities of these countries and their educational systems. Also, considering the potential differences between different educational and developmental stages in learning with RE and VE and the small body of studies with participants younger than undergraduate university students, the conclusions of this review might be primarily applicable to secondary and postsecondary students. Haptic experiences and physicality might be more important for younger children as they have not accumulated as many hands-on experiences with their environment in their lives as older students have.

Eighth, due to space reasons, we focused specifically on RE and VE without also considering remote experiments in our review. Remote experiments combine some advantages of RE with some advantages of VE: They are digitally accessible RE that can be controlled and thus used for learning from remote, thereby showing a live video of the experiment and the authentic measurement values. The role of remote experiments compared to RE and VE is discussed in detail in various reviews of the literature (Alkhaldi et al., 2016; Brinson, 2015; Hernández-de-Menéndez et al., 2019; Ma & Nickerson, 2006; Zacharia et al., 2015).

Future Research

Future research should specifically address the following aspects: First, and most important, the type of instructional support and guidance during learning with combinations of RE and VE needs to be considered more systematically. Previous researchers emphasized that instruction, scaffolds, and guidance play an essential role for inquiry learning (e.g., Lazonder & Harmsen, 2016). It is noticeable that in the reviewed studies the role of instruction and guidance during inquiry learning was not considered to the extent that we expected it to be. For example, in Pineda (2015) the confounding role of the scaffolding overview presentation was not addressed. When conducting more studies to investigate the effectiveness of combined RE and VE, it should be carefully considered that the scaffolds, guidance, and instruction are the same between the different conditions to not bias the study. At this point, it is important to remark that individual scaffolds might be easier to implement in VE than in RE. Even if the aim of a future study is not to systematically investigate guidance as a factor of successful inquiry learning, the guidance provided to the students should at least be reported in more detail in future publications. Inspiration for future research concerning this first point can be drawn from Rau (2017), who suggested that interventions with
virtual representations may especially benefit from meta-cognitive support. Another idea comes from Hale-Hanes (2015), who concluded in her study that discussions with students where misconceptions are actively addressed, and student data are discussed, might be a helpful support for students’ learning with combinations of RE and VE. A third impulse might be the study by Jaakkola et al. (2011), which was also included in the papers assigned to RQ1. Their results suggest that explicit instruction rather than implicit instruction is an important scaffold for students under certain experimentation conditions. The idea that direct instruction can be an important form of guidance is also suggested by Chen et al. (2017) and Schneider and Preckel (2017). For more clarity about all these aspects of appropriately guiding combinations of RE and VE, more and especially systematic research is needed.

Second, when planning future studies about combinations of RE and VE, researchers might consider investigating in subject domains that have not yet been investigated in detail (i.e., topics other than electric circuits). Also, participants younger than university students (i.e., primary and secondary school students at various grade levels) should be involved in future studies on combining RE and VE, and if possible, research groups from more countries around the world should conduct studies about combinations of RE and VE to consider potential differences in the educational systems of different countries. For future research on this topic, the problems named in the results section “Problems Frequently Reported” should be carefully considered when planning a study.

Moreover, various methodological issues were noted in the studies included in this review, which should be overcome in future research. Future studies on combinations of RE and VE should report their effect sizes throughout and ideally also consider long-term results with a follow-up test several weeks after the intervention. These aspects were rarely reported in the studies included in this review. Moreover, access to test items used in the single studies should be granted, which was sometimes the case, but rather exceptional. Here it could be checked afterwards whether the result might be biased because the conceptual test used representations that also occurred in the combined or the VE condition, but not in the RE condition.

Third, the post hoc explanations that were provided by multiple authors of the studies included in RQ2 should be tested empirically. This would lead to much more robust findings about the best choice of sequence in a combination of RE and VE and strengthen the arguments presented in the studies assigned to RQ2.

On a content-related level, more systematic research on other boundary conditions for a successful combination than research on sequence alone would help for better insights, for example, when considering the subject domain and the learning objectives of the specific task. It would also be interesting to investigate whether the second representation offered additionally to the RE needs to be a VE to increase students’ learning, or whether it could also be an animation (i.e., dynamic, but not interactive visualization).

Last, it is stunning that time on task had no visible effect on learning in the settings reviewed in this article, whereas it has been shown to have significant influence on learning in general (Anderson, 1981). This also provides an interesting starting point for further research.
Theoretical and Practical Implications

Theoretical Implications

One of the most important insights of this systematic review is that apart from the individual affordances and the learning objectives of the different experiment types, especially their specific function within the learning task must be considered when combining RE and VE. This conclusion aligns well with what Ainsworth (2006) has proposed in the DeFT (Design, Functions, Tasks) framework for learning with multiple representations. This framework provides an approach to analyzing the effectiveness of learning with multiple representations by considering three dimensions, namely, multiple representations’ design parameters, their functions, and the cognitive tasks that must be undertaken by a student. The design aspect includes the number of representations, the way that information is distributed, the form of the representational system, the sequence of representations, and the support for translation between representations. For the functions of multiple representations, Ainsworth (2006) suggests the following roles: They can have complementary roles, they can constrain interpretation, and they can construct deeper understanding. These functions are not necessarily exclusive, and the different representations can simultaneously support more than one of these roles. Concerning the aspect of cognitive tasks that students need to perform when learning with multiple representations, Ainsworth (2006) emphasizes that it is important that learners can relate the different representations to each other to integrate the information presented in the different formats. This theoretical framework provides various points of reference that can be applied to learning with combinations of RE and VE in science education, which, so far, have not been considered. In particular, VE and RE can both be considered external representations that, when combined, can serve any of the functions mentioned within the DeFT framework. The design parameters specified in the DeFT framework could be used to analyze existing combinations of VE and RE and inform their design, as well as to develop instructional support that is tailored toward integrating knowledge from learning with either type of experimentation. Accordingly, applying the DeFT framework could provide new insights and perspectives on learning with combinations of RE and VE and also guide future research endeavors.

The findings of this review can also be analyzed from the perspective of research on “cumulative learning” (i.e., stepwise learning based on prior knowledge; Biemans, 1997). Biemans (1997) explains the importance of prior knowledge for students’ understanding of new information and construction of rich and useful mental representations. “If the learner has constructed representations of a certain domain based upon learning experiences in the past, s/he can use this prior knowledge when s/he has to study related material” (Biemans, 1997, p. 6). Similarly, Bransford and Schwartz (1999) argue that “the better prepared [students] are for future learning, the greater the transfer (in terms of speed and/or quality of new learning)” (p. 68).

This assumption can be applied to RE and VE: If the first experiment in the sequence ties in with the prior knowledge that the learner possesses already from prior lessons or everyday experiences, this can give students a good entry into the sequence of experiments. Additionally, if one type of experiment is used to
construct a representation of the learning topic first, the learners’ experiences with the second experiment can be based on this prior knowledge and thus help learners achieve deep understanding of the topic. For this reason, both sequences of RE and VE may contribute to cumulative learning as they either strengthen the link to students’ prior experiences when starting with a concrete experiment using real objects (i.e., in the sequence RE-VE) or enable the use of prior (conceptual) knowledge acquired from interacting with a VE when planning and conducting a real-world experiment (i.e., in the sequence VE-RE).

Both, the approach of cumulative learning and preparation for future learning, emphasize that learning is not a one-shot trial, but a continuous experience, where one learning activity leads to another and where one experience is used to facilitate and scaffold the next. Overall, designing effective transitions between learning experiences might lead to successful learning with different combinations of RE and VE. The questions of how to best activate prior knowledge before starting with the experiments (Zacharia et al., 2015) and how to design inquiry tasks using VE and RE so that they enable cumulative learning provide interesting avenues for future research.

Practical Implications

We encourage teachers to not dismiss either form of experimentation (RE or VE), as there is added value of combining them. Therefore, it is important to think carefully about how to best combine these experiments, with the single experiments’ functions and possible relations of each experiment to students’ prior knowledge in mind. Teachers play a crucial role in creating an effective learning environment with RE and VE, as they are the ones who are responsible for selecting and designing, as well as orchestrating the learning activities. The importance of science teachers as key enablers of effective inquiry learning was already earmarked by de Jong et al. (2013), who identified the question of how to promote teachers’ competence in this regard as one of the grand challenges to be solved in future research. In our view, the present review very well aligns with this position, since it highlights the fact that successful inquiry learning is based on orchestration of multiple learning activities rather than just assigning tools to students.

Conclusion

The goal of this systematic review was to have a closer look into whether combinations of real and virtual experiments are more effective than real or virtual experiments alone and how the combinations should be sequenced to maximize students’ learning. It can be concluded that combinations of experiments in this sample of studies have shown to be more effective for learning than real or virtual experiments alone. The advantage of the combinations is generic and based on synergetic effects of the complementary affordances of real and virtual experiments. In contrast to these broadly consistent findings regarding the benefits of combining the two types of experimentation, studies investigating the sequencing of real and virtual experiments reported mixed results. We conclude that the sequences should be designed with respect to the affordances of real and virtual experiments, with consideration of the learning objectives, and special consideration of the specific function each experiment should serve. This might determine
the individual sequence of experiments that best fits the learning topic and the learning objectives.

This systematic review highlights gaps in the existing body of research and proposes new avenues for future studies. Beyond emphasizing the need for studies with more methodological rigor, we propose to apply the DeFT framework for learning with multiple representations (Ainsworth, 2006) to research on combining RE and VE in order to more systematically identify the conditions under which they foster student achievement. The review provides science teachers with insights into the current scientific knowledge base that can inform their lesson design as it advocates orchestration of learning with virtual and real experiments rather than replacing real experiments with their virtual counterparts.

**ORCID iD**

Salome Wörner [ID](https://orcid.org/0000-0002-7049-9316)

**Note**

Salome Wörner is a doctoral student at the LEAD Graduate School & Research Network, which is funded by the Ministry of Science, Research and the Arts of the state of Baden-Württemberg within the framework of the sustainability funding for the projects of the Excellence Initiative II. Moreover, she is a junior fellow within the program “Kolleg Didaktik:digital” by the Joachim Herz Stiftung (Hamburg, Germany).

**References**

*References marked with an asterisk indicate studies included in the systematic review.*

*Abdulwahed, M., & Nagy, Z. K. (2009). Applying Kolb’s experiential learning cycle for laboratory education. *Journal of Engineering Education, 98*(3), 283–294. https://doi.org/10.1002/j.2168-9830.2009.tb01025.x

*Achuthan, K., Francis, S. P., & Diwakar, S. (2017). Augmented reflective learning and knowledge retention perceived among students in classrooms involving virtual laboratories. *Education and Information Technologies, 22*(6), 2825–2855. https://doi.org/10.1007/s10639-017-9626-x

Aditomo, A., & Klieme, E. (2020). Forms of inquiry-based science instruction and their relations with learning outcomes: Evidence from high and low-performing education systems. *International Journal of Science Education, 42*(4), 504–525. https://doi.org/10.1080/09500693.2020.1716093

Ainsworth, S. (2006). DeFT: A conceptual framework for considering learning with multiple representations. *Learning and Instruction, 16*(3), 183–198. https://doi.org/10.1016/j.learninstruc.2006.03.001

*Akpan, J. P., & Andre, T. (2000). Using a computer simulation before dissection to help students learn anatomy. *Journal of Computers in Mathematics and Science Teaching, 19*(3), 297–313. https://www.learntechlib.org/primary/p/8073/

Alfieri, L., Brooks, P. J., Aldrich, N. J., & Tenenbaum, H. R. (2011). Does discovery-based instruction enhance learning? *Journal of Educational Psychology, 103*(1), 1–18. https://doi.org/10.1037/a0021017

Alkhalidi, T., Pranata, I., & Athauda, R. I. (2016). A review of contemporary virtual and remote laboratory implementations: Observations and findings. *Journal of
Combining Real and Virtual Experiments

Anderson, L. W. (1981). Instruction and time-on-task: A review. Journal of Curriculum Studies, 13(4), 289–303. https://doi.org/10.1080/0022027810130402

*Atanas, J.-P. (2018). Is virtual-physical or physical-virtual manipulatives in physics irrelevant within studio physics environment? Athens Journal of Education, 5(1), 29–42. https://doi.org/10.30958/aje.5-1-2

Becker, S., Klein, P., Gößling, A., & Kuhn, J. (2020). Using mobile devices to enhance inquiry-based learning processes. Learning and Instruction, 69(October), 101350. https://doi.org/10.1016/j.learninstruc.2020.101350

Bell, T., Urhahne, D., Schanze, S., & Ploetzner, R. (2010). Collaborative inquiry learning: Models, tools, and challenges. International Journal of Science Education, 32(3), 349–377. https://doi.org/10.1080/09500690802582241

Biemans, H. J. A. (1997). Fostering activation of prior knowledge and conceptual change [Doctoral dissertation, Radboud University Nijmegen]. https://repository.ubn.ru.nl/bitstream/handle/2066/146341/mmubn000001_238538710.pdf

*Bortnik, B., Stozhko, N., Pervukhina, I., Tchernysheva, A., & Belysheva, G. (2017). Effect of virtual analytical chemistry laboratory on enhancing student research skills and practices. Research in Learning Technology, 25. https://doi.org/10.25304/rlt.v25.1968

Bransford, J. D., & Schwartz, D. L. (1999). Chapter 3: Rethinking transfer: A simple proposal with multiple implications. Review of Research in Education, 24(1), 61–100. https://doi.org/10.3102/0091732x024001061

Brinson, J. R. (2015). Learning outcome achievement in non-traditional (virtual and remote) versus traditional (hands-on) laboratories: A review of the empirical research. Computers & Education, 87(September), 218–237. https://doi.org/10.1016/j.compedu.2015.07.003

*Campbell, J. O., Bourne, J. R., Mosterman, P. J., & Brodersen, A. J. (2002). The effectiveness of learning simulations for electronic laboratories. Journal of Engineering Education, 91(1), 81–87. https://doi.org/10.1002/j.2168-9830.2002.tb00675.x

Chen, L., Dorn, E., Krawitz, M., Lim, C., & Moursched, M. (2017). Drivers of student performance: Insights from Asia. McKinsey & Company.

Chernikova, O., Heitzmann, N., Stadler, M., Holzberger, D., Seidel, T., & Fischer, F. (2020). Simulation-based learning in higher education: A meta-analysis. Review of Educational Research, 90(4), 499–541. https://doi.org/10.3102/0034654320933544

*Chini, J. J. (2010). Comparing the scaffolding provided by physical and virtual manipulative for students’ understanding of simple machines [Doctoral dissertation, Kansas State University]. https://krex.k-state.edu/dspace/handle/2097/6391

*Chini, J. J., Madsen, A., Gire, E., Rebello, N. S., & Puntambekar, S. (2012). Exploration of factors that affect the comparative effectiveness of physical and virtual manipulatives in an undergraduate laboratory. Physical Review Special Topics: Physics Education Research, 8(1), 010113. https://doi.org/10.1103/physrevst-per.8.010113

*Climent-Bellido, M. S., Martínez-Jiménez, P., Pontes-Pedrajas, A., & Polo, J. (2003). Learning in chemistry with virtual laboratories. Journal of Chemical Education, 80(3), 346. https://doi.org/10.1021/ed080p346

*Darrah, M., Humbert, R., Finstein, J., Simon, M., & Hopkins, J. (2014). Are virtual labs as effective as hands-on labs for undergraduate physics? A comparative study
Wörner et al.

at two major universities. *Journal of Science Education and Technology*, 23(6), 803–814. https://doi.org/10.1007/s10956-014-9513-9

de Jong, T. (2006). Technological advances in inquiry learning. *Science*, 312(5773), 532–533. https://doi.org/10.1126/science.1127750

de Jong, T. (2019). Moving towards engaged learning in STEM domains; There is no simple answer, but clearly a road ahead. *Journal of Computer Assisted Learning*, 35(2), 153–167. https://doi.org/10.1111/jcal.12337

de Jong, T., Linn, M. C., & Zacharia, Z. C. (2013). Physical and virtual laboratories in science and engineering education. *Science*, 340(6130), 305–308. https://doi.org/10.1126/science.1230579

de Jong, T., Sotiriou, S., & Gillet, D. (2014). Innovations in STEM education: The Go-Lab Federation of online labs. *Smart Learning Environments*, 1(1), 1–16. https://doi.org/10.1186/s40561-014-0003-6

de Jong, T., & van Joolingen, W. R. (1998). Scientific discovery learning with computer simulations of conceptual domains. *Review of Educational Research*, 68(2), 179–201. https://doi.org/10.3102/00346543068002179

Deslauriers, L., & Wieman, C. (2011). Learning and retention of quantum concepts with different teaching methods. *Physical Review Special Topics: Physics Education Research*, 7(1), 010101-1–010101-6. https://doi.org/10.1103/physrevstper.7.010101

Edelson, D. C., Gordin, D. N., & Pea, R. D. (1999). Addressing the challenges of inquiry-based learning through technology and curriculum design. *Journal of the Learning Sciences*, 8(3), 391–450. https://doi.org/10.1207/s15327809jls0803&4_3

*Farrokhnia, M. R., & Esmailpour, A. (2010). A study on the impact of real, virtual and comprehensive experimenting on students’ conceptual understanding of DC electric circuits and their skills in undergraduate electricity laboratory. *Procedia: Social and Behavioral Sciences*, 2(2), 5474–5482. https://doi.org/10.1016/j.sbspro.2010.03.893

Finkelstein, N. D., Adams, W. K., Keller, C. J., Kohl, P. B., Perkins, K. K., Podolefsky, N. S., Reid, S., & LeMaster, R. (2005). When learning about the real world is better done virtually: A study of substituting computer simulations for laboratory equipment. *Physical Review Special Topics: Physics Education Research*, 1(1), 010103-1–010103-8. https://doi.org/10.1103/physrevstper.1.010103

Ford, D. N., & McCormack, D. E. (2000). Effects of time scale focus on system understanding in decision support systems. *Simulation & Gaming*, 31(3), 309–330. https://doi.org/10.1177/104687810003100301

Furtak, E. M., Seidel, T., Iverson, H., & Briggs, D. C. (2012). Experimental and quasi-experimental studies of inquiry-based science teaching. *Review of Educational Research*, 82(3), 300–329. https://doi.org/10.3102/0034654312457206

Geelan, D. R., & Fan, X. (2014). Teachers using interactive simulations to scaffold inquiry instruction in physical science education. In *Science teachers’ use of visual representations* (pp. 249–270). Springer. https://doi.org/10.1007/978-3-319-06526-7_11

*Gire, E., Carmichael, A., Chini, J. J., Rouinfar, A., Rebello, S., Smith, G., & Puntambekar, S. (2010). The effects of physical and virtual manipulatives on students’ conceptual learning about pulleys. In *Proceedings of the 9th International Conference of the Learning Sciences (ICLS 2010)* (pp. 937–943). International Society of the Learning Sciences.

Goldwater, M. B., & Schalk, L. (2016). Relational categories as a bridge between cognitive and educational research. *Psychological Bulletin*, 142(7), 729–757. https://doi.org/10.1037/bul0000043
Combining Real and Virtual Experiments

*Gumilar, S., Ismail, A., Budiman, D. M., & Siswanto, S. (2019). Inquiry instructional model infused blended experiment: Helping students enhance critical thinking skills. *Journal of Physics: Conference Series, 1157*, 032009. https://doi.org/10.1088/1742-6596/1157/3/032009

Hale-Hanes, C. (2015). Promoting student development of models and scientific inquiry skills in acid–base chemistry: An important skill development in preparation for AP chemistry. *Journal of Chemical Education, 92*(8), 1320–1324. https://doi.org/10.1021/ed500814n

Hernández-de-Menéndez, M., Vallejo Guevara, A., & Morales-Menéndez, R. (2019). Virtual reality laboratories: A review of experiences. *International Journal on Interactive Design and Manufacturing, 13*(3), 947–966. https://doi.org/10.1007/s12008-019-00558-7

Hofstein, A., & Lunetta, V. N. (2004). The laboratory in science education: Foundations for the twenty-first century. *Science Education, 88*(1), 28–54. https://doi.org/10.1002/sce.10106

*Huppert, J., Lomask, S. M., & Lazarowitz, R. (2002). Computer simulations in the high school: Students’ cognitive stages, science process skills and academic achievement in microbiology. *International Journal of Science Education, 24*(8), 803–821. https://doi.org/10.1080/09500690110049150

Husnaini, S. J., & Chen, S. (2019). Effects of guided inquiry virtual and physical laboratories on conceptual understanding, inquiry performance, scientific inquiry self-efficacy, and enjoyment. *Physical Review Physics Education Research, 15*(1), 010119-1-010119-16. https://doi.org/10.1103/physrevphyseducres.15.010119

*Jaakkola, T., & Nurmi, S. (2008). Fostering elementary school students’ understanding of simple electricity by combining simulation and laboratory activities. *Journal of Computer-Assisted Learning, 24*(4), 271–283. https://doi.org/10.1111/j.1365-2729.2007.00259.x

*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

Josephsen, J., & Kristensen, A. K. (2006). Simulation of laboratory assignments to support students’ learning of introductory inorganic chemistry. *Chemistry Education Research and Practice, 7*(4), 266–279. https://doi.org/10.1039/b6rp90013e

*Kapici, H. O., Akcay, H., & de Jong, T. (2019). Using hands-on and virtual laboratories alone or together—Which works better for acquiring knowledge and skills? *Journal of Science Education and Technology, 28*(3), 231–250. https://doi.org/10.1007/s10956-018-9762-0

Keselman, A. (2003). Supporting inquiry learning by promoting normative understanding of multivariable causality. *Journal of Research in Science Teaching, 40*(9), 898–921. https://doi.org/10.1002/tea.10115

*Kollöffel, B., & de Jong, T. (2013). Conceptual understanding of electrical circuits in secondary vocational engineering education: Combining traditional instruction with inquiry learning in a virtual lab. *Journal of Engineering Education, 102*(3), 375–393. https://doi.org/10.1002/jee.20022

Lazonder, A. W., & Harmsen, R. (2016). Meta-analysis of inquiry-based learning: Effects of guidance. *Review of Educational Research, 86*(3), 681–718. https://doi.org/10.3102/0034654315627366
*Le, H. T. (2015). Guidance-based hybrid lab training method for enhancing core skills of EE students. In 2015 IEEE Power & Energy Society General Meeting. IEEE. https://ieeexplore.ieee.org/abstract/document/7285738

*Liu, X. (2006). Effects of combined hands-on laboratory and computer modeling on student learning of gas laws: A quasi-experimental study. *Journal of Science Education and Technology, 15*(1), 89–100. https://doi.org/10.1007/s10956-006-0359-7

Ma, J., & Nickerson, J. V. (2006). Hands-on, simulated, and remote laboratories: A comparative literature review. *ACM Computing Surveys (CSUR), 38*(3), 1–24. https://doi.org/10.1145/1132960.1132961

Mäeots, M., Pedaste, M., & Sarapuu, T. (2008, July). Transforming students’ inquiry skills with computer-based simulations [Conference]. Eighth IEEE International Conference on Advanced Learning Technologies, Santander, Spain. https://doi.org/10.1109/icalt.2008.239

*Makransky, G., Thisgaard, M. W., & Gadegaard, H. (2016). Virtual simulations as preparation for lab exercises: Assessing learning of key laboratory skills in microbiology and improvement of essential non-cognitive skills. *PLoS One, 11*(6), e0155895. https://doi.org/10.1371/journal.pone.0155895

*Manunure, K., Delserieys, A., & Castéra, J. (2020). The effects of combining simulations and laboratory experiments on Zimbabwean students’ conceptual understanding of electric circuits. *Research in Science & Technological Education, 38*(3), 289–307. https://doi.org/10.1080/02635143.2019.1629407

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

Minner, D. D., Levy, A. J., & Century, J. (2010). Inquiry-based science instruction—What is it and does it matter? Results from a research synthesis years 1984 to 2002. *Journal of Research in Science Teaching, 47*(4), 474–496. https://doi.org/10.1002/tea.20347

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine, 6*(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097

*Myneni, L. S., Narayanan, N. H., Rebello, S., Rouinfar, A., & Pumtambekar, S. (2013). An interactive and intelligent learning system for physics education. *IEEE Transactions on Learning Technologies, 6*(3), 228–239. https://doi.org/10.1109/tlt.2013.26

National Research Council. (2012). *A framework for K–12 science education: Practices, crosscutting concepts, and core ideas*. National Academies Press.

Oliver, M., McConney, A., & Woods-McConney, A. (2019). The efficacy of inquiry-based instruction in science: A comparative analysis of six countries using PISA 2015. *Research in Science Education, 51*(2), 595–616. https://doi.org/10.1007/s11165-019-09901-0

*Olympiou, G., & Zacharia, Z. C. (2012). Blending physical and virtual manipulatives: An effort to improve students’ conceptual understanding through science laboratory experimentation. *Science Education, 96*(1), 21–47. https://doi.org/10.1002/sce.20463

*Olympiou, G., & Zacharia, Z. C. (2014). Blending physical and virtual manipulatives in physics laboratory experimentation. In C. Bruguière, A. Tiberghien, & P. Clément (Eds.), *Topics and trends in current science education* (pp. 419–433). Springer. https://doi.org/10.1007/978-94-007-7281-6_26
Olympiou, G., Zacharia, Z. C., & de Jong, T. (2013). Making the invisible visible: Enhancing students’ conceptual understanding by introducing representations of abstract objects in a simulation. *Instructional Science, 41*(3), 575–596. https://doi.org/10.1007/s11251-012-9245-2

Organisation for Economic Co-operation and Development. (2006). *Assessing scientific, reading and mathematical literacy: A framework for PISA 2006*. Publications de l’OCDE. https://doi.org/10.1787/9789264026407-en

Organisation for Economic Co-operation and Development. (2007). *PISA 2006: Science competencies for tomorrow’s world. Volume 1: Analysis*. Publications de l’OCDE. https://www.oecd.org/unitedstates/39722597.pdf

Organisation for Economic Co-operation and Development. (2016). *PISA 2015: Excellence and equity in education: Results (Vol. 1)*. Publications de l’OCDE. https://doi.org/10.1787/9789264266490-en

Pedaste, M., Mäeots, M., Leijen, Ä., & Sarapuu, T. (2012). Improving students’ inquiry skills through reflection and self-regulation scaffolds. *Technology, Instruction, Cognition and Learning, 9*(1–2), 81–95. https://www.researchgate.net/publication/285309266_Improving_students’_inquiry_skills_through_reflection_and_self-regulation_scaffolds

Pedaste, M., Mäeots, M., Siiman, L. A., de Jong, T., van Riesen, S. A. N., Kamp, E. T., Manoli, C. C., Zacharia, Z. C., & Tsourlidaki, E. (2015). Phases of inquiry-based learning: Definitions and the inquiry cycle. *Educational Research Review, 14*(February), 47–61. https://doi.org/10.1016/j.edurev.2015.02.003

*Pineda, B. (2015). Using computer simulations as a pre-training activity in a hands-on lab to help community college students improve their understanding of physics [Doctoral dissertation, University of San Francisco]. https://repository.usfca.edu/diss/292/

Pyatt, K., & Sims, R. (2012). Virtual and physical experimentation in inquiry-based science labs: Attitudes, performance and access. *Journal of Science Education and Technology, 21*(1), 133–147. https://doi.org/10.1007/s10956-011-9291-6

*Raman, R., Achuthan, K., Nedungadi, P., & Ramesh, M. (2014). Modeling diffusion of blended labs for science experiments among undergraduate engineering students. In T. Bissyandé & G. van Stam (Eds.), *e-Infrastructure and e-services for developing countries* (pp. 234–247). Springer. https://doi.org/10.1007/978-3-319-08368-1_28

Rau, M. A. (2017). How do students learn to see concepts in visualizations? Social learning mechanisms with physical and virtual representations. *Journal of Learning Analytics, 4*(2), 240–263. https://doi.org/10.18608/jla.2017.42.16

Rau, M. A. (2020). Comparing multiple theories about learning with physical and virtual representations: Conflicting or complementary effects? *Educational Psychology Review, 32*(2), 297–325. https://doi.org/10.1007/s10649-020-09517-1

Renken, M. D., & Nunez, N. (2013). Computer simulations and clear observations do not guarantee conceptual understanding. *Learning and Instruction, 23*(February), 10–23. https://doi.org/10.1016/j.learninstruc.2012.08.006

*Ronen, M., & Eliahu, M. (2000). Simulation: A bridge between theory and reality: The case of electric circuits. *Journal of Computer Assisted Learning, 16*(1), 14–26. https://doi.org/10.1046/j.1365-2729.2000.00112.x

Rutten, N., van Jooolingen, W. R., & van der Veen, J. T. (2012). The learning effects of computer simulations in science education. *Computers & Education, 58*(1), 136–153. https://doi.org/10.1016/j.compedu.2011.07.017

*Salehi, S., Schneider, B., & Blikstein, P. (2014). The effects of physical and virtual manipulatives on learning basic concepts in electronics. In *Proceedings of the*
Extended Abstracts of the 32nd Annual ACM Conference on Human Factors in Computing Systems: CHI EA ’14 (pp. 2263–2268). Association for Computing Machinery. https://doi.org/10.1145/2559206.2581346

Schneider, M., & Preckel, F. (2017). Variables associated with achievement in higher education: A systematic review of meta-analyses. Psychological Bulletin, 143(6), 565–600. https://doi.org/10.1037/bul0000098

Schneider, M., Rittle-Johnson, B., & Star, J. R. (2011). Relations among conceptual knowledge, procedural knowledge, and procedural flexibility in two samples differing in prior knowledge. Developmental Psychology, 47(6), 1525–1538. https://doi.org/10.1037/a0024997

Slavin, R., & Lake, C. (2008). Effective programs in elementary mathematics: A best-evidence synthesis. Review of Educational Research, 78(3), 427–515. https://doi.org/10.3102/0034654308317473

Slavin, R., & Smith, D. (2009). The relationship between sample sizes and effect sizes in systematic reviews in education. Educational Evaluation and Policy Analysis, 31(4), 500–506. https://doi.org/10.3102/0162373709352369

*Smith, G. W., & Puntambekar, S. (2010). Examining the combination of physical and virtual experiments in an inquiry science classroom. In Proceedings of the Conference on Computer Based Learning in Science. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.685.2351&rep=rep1&type=pdf

Srinivasan, S., Pérez, L. C., Palmer, R. D., Brooks, D. W., Wilson, K., & Fowler, D. (2006). Reality versus simulation. Journal of Science Education and Technology, 15(2), 137–141. https://doi.org/10.1007/s10956-006-9007-5

*Sullivan, S., Gnesdilow, D., Puntambekar, S., & Kim, J.-S. (2017). Middle school students’ learning of mechanics concepts through engagement in different sequences of physical and virtual experiments. International Journal of Science Education, 39(12), 1573–1600. https://doi.org/10.1080/09500693.2017.1341668

Sypsas, A., & Kalles, D. (2018). Virtual laboratories in biology, biotechnology and chemistry education. In Proceedings of the 22nd Pan-Hellenic Conference on Informatics (pp. 70–75). Association for Computing Machinery. https://doi.org/10.1145/3291533.3291560

*Toth, E. E., Ludvico, L. R., & Morrow, B. L. (2014). Blended inquiry with hands-on and virtual laboratories: The role of perceptual features during knowledge construction. Interactive Learning Environments, 22(5), 614–630. https://doi.org/10.1080/10494820.2012.693102

*Toth, E. E., Morrow, B. L., & Ludvico, L. R. (2009). Designing blended inquiry learning in a laboratory context: A study of incorporating hands-on and virtual laboratories. Innovative Higher Education, 33(5), 333–344. https://doi.org/10.1007/s10755-008-9087-7

Trundle, K. C., & Bell, R. L. (2010). The use of a computer simulation to promote conceptual change: A quasi-experimental study. Computers & Education, 54(4), 1078–1088. https://doi.org/10.1016/j.compedu.2009.10.012

*Tsihouridis, C., Vavougios, D., Ioannidis, G. S., Alexias, A., Argyropoulos, C., & Poullos, S. (2015, September). The effect of teaching electric circuits switching from real to virtual lab or vice versa: A case study with junior high-school learners [Conference]. 2015 International Conference on Interactive Collaborative Learning (ICL), Firenze, Italy. https://doi.org/10.1109/icl.2015.7318102

*Ünlü, Z. K., & Dökme, I. (2011). The effect of combining analogy-based simulation and laboratory activities on Turkish elementary school students’ understanding of
Combining Real and Virtual Experiments

van der Meij, J., & de Jong, T. (2006). Supporting students’ learning with multiple representations in a dynamic simulation-based learning environment. *Learning and Instruction, 16*(3), 199–212. https://doi.org/10.1016/j.learninstruc.2006.03.007

*Wang, T.-L., & Tseng, Y.-K. (2018). The comparative effectiveness of physical, virtual, and virtual-physical manipulatives on third-grade students’ science achievement and conceptual understanding of evaporation and condensation. *International Journal of Science and Mathematics Education, 16*(2), 203–219. https://doi.org/10.1007/s10763-016-9774-2

Wiesner, T. F., & Lan, W. (2004). Comparison of student learning in physical and simulated unit operations experiments. *Journal of Engineering Education, 93*(3), 195–204. https://doi.org/10.1002/j.2168-9830.2004.tb00806.x

*Zacharia, Z. C. (2007). Comparing and combining real and virtual experimentation: An effort to enhance students’ conceptual understanding of electric circuits. *Journal of Computer Assisted Learning, 23*(2), 120–132. https://doi.org/10.1111/j.1365-2729.2006.00215.x

Zacharia, Z. C. (2015). Examining whether touch sensory feedback is necessary for science learning through experimentation: A literature review of two different lines of research across K-16. *Educational Research Review, 16*(October), 116–137. https://doi.org/10.1016/j.edurev.2015.10.001

*Zacharia, Z. C., & Anderson, O. R. (2003). The effects of an interactive computer-based simulation prior to performing a laboratory inquiry-based experiment on students’ conceptual understanding of physics. *American Journal of Physics, 71*(6), 618–629. https://doi.org/10.1119/1.1566427

Zacharia, Z. C., & Constantinou, C. P. (2008). Comparing the influence of physical and virtual manipulatives in the context of the Physics by Inquiry curriculum: The case of undergraduate students’ conceptual understanding of heat and temperature. *American Journal of Physics, 76*(4), 425–430. https://doi.org/10.1119/1.2885059

*Zacharia, Z. C., & de Jong, T. (2014). The effects on students’ conceptual understanding of electric circuits of introducing virtual manipulatives within a physical manipulatives-oriented curriculum. *Cognition and Instruction, 32*(2), 101–158. https://doi.org/10.1080/07370008.2014.887083

Zacharia, Z. C., Loizou, E., & Papaevripidou, M. (2012). Is physicality an important aspect of learning through science experimentation among kindergarten students? *Early Childhood Research Quarterly, 27*(3), 447–457. https://doi.org/10.1016/j.ecresq.2012.02.004

Zacharia, Z. C., Manoli, C., Xenofontos, N., de Jong, T., Pedaste, M., van Riesen, S. A. N., Kamp, E. T., Määets, M., Siiman, L., & Tsourlidaki, E. (2015). Identifying potential types of guidance for supporting student inquiry when using virtual and remote labs in science: A literature review. *Educational Technology Research and Development, 63*(2), 257–302. https://doi.org/10.1007/s11423-015-9370-0

*Zacharia, Z. C., & Michael, M. (2016). Using physical and virtual manipulatives to improve primary school students’ understanding of concepts of electric circuits. In *New developments in science and technology education* (pp. 125–140). https://doi.org/10.1007/978-3-319-22933-1_12

*Zacharia, Z. C., & Olympiou, G. (2011). Physical versus virtual manipulative experimentation in physics learning. *Learning and Instruction, 21*(3), 317–331. https://doi.org/10.1016/j.learninstruc.2010.03.001
Wörner et al.

*Zacharia, Z. C., Olympiou, G., & Papaevripidou, M. (2008). Effects of experimenting with physical and virtual manipulatives on students’ conceptual understanding in heat and temperature. *Journal of Research in Science Teaching, 45*(9), 1021–1035. https://doi.org/10.1002/tea.20260

Zhang, Z. H., & Linn, M. C. (2011). Can generating representations enhance learning with dynamic visualizations? *Journal of Research in Science Teaching, 48*(10), 1177–1198. https://doi.org/10.1002/tea.20443

**Authors**

SALOME WÖRNER is a PhD student at the Leibniz-Institut für Wissensmedien, Schleichstraße 6, 72076 Tübingen, Germany, and the Department of Physics/Physics Education Research Group, Technische Universität Kaiserslautern, Erwin-Schrödinger-Str. 46, 67663 Kaiserslautern, Germany; email: s.woerner@iwm-tuebingen.de. In her research, she deals with the question of how virtual and real experiments can be combined in an educationally meaningful way in science lessons.

JOCHEN KUHN is a full professor at the Department of Physics/Physics Education Research Group, Technische Universität Kaiserslautern, Erwin-Schrödinger-Str. 46, 67663 Kaiserslautern, Germany; email: kuhn@physik.uni-kl.de. His research focuses on learning with multiple representations in physics education using common and advanced multimedia technology.

KATHARINA SCHEITER is a full professor at the Leibniz-Institut für Wissensmedien, Schleichstraße 6, 72076 Tübingen, Germany; email: k.scheiter@iwm-tuebingen.de. Together with her research group, she investigates cognitive and meta-cognitive processes underlying learning from multiple representations as well as means of supporting these processes.