International Journal of Science and Mathematics Education
https://doi.org/10.1007/s10763-020-10147-w

# Comparison of the Quality of Written Scientific Arguments in Different Laboratory Environments

Hasan Ozgur Kapici[1] · Hakan Akcay[1] · Ece Ebrar Koca[2]

Received: 11 February 2020 / Accepted: 23 December 2020/
© Ministry of Science and Technology, Taiwan 2021

## Abstract

Although physical hands-on laboratory experiments are commonly used in schools, virtual laboratory environments are also being used in science classrooms. Their possibilities and impact on students' achievement, conceptual understanding, and inquiry skills have been investigated by other researchers. Yet, it is difficult to find studies about the effect of virtual learning environments on written argumentation in science classrooms. In the current study, the influence of hands-on and virtual laboratories on pre-service science teachers' scientific writing was investigated using the science writing heuristic. The participants were 52 pre-service science teachers who were assigned to two conditions. One group used the hands-on laboratory to design and implement experiments and used a paper-based worksheet to write their arguments; the other group did their investigations through a computer-based virtual-tool laboratory and wrote their arguments on the online learning platform. Pre-service science teachers' written arguments were evaluated using content analysis. The findings revealed that the overall quality of written arguments created by pre-service science teachers in both laboratory environments was at an intermediate level. Furthermore, pre-service science teachers who learned in the virtual laboratory posed better testable questions, presented strong and valid evidence, and used multiple representations better than their counterparts in the hands-on laboratory. An important conclusion of the current study is that the science writing heuristic can be used effectively in both virtual and hands-on laboratory environments.

**Keywords** Argumentation · Hands-on laboratory · Science writing heuristic · Virtual laboratory

✉ Hasan Ozgur Kapici
hokapici@yildiz.edu.tr

[1] Department of Mathematics and Science Education, Faculty of Education, Yildiz Technical University, Istanbul, Turkey

[2] Ministry of National Education, Ankara, Turkey

## Introduction

Writing to learn has a critical function in science education as a means to support the development and use of metacognitive skills such as reflection, critical and analytical thinking abilities (Leinonen, Keune, Veermans, & Toikkanen, 2016), which in turn lead to meaningful understanding (Chen, Park, & Hand, 2016b). Furthermore, conceptual understanding of core ideas of science can be achieved better when students write an argumentative text (Jang & Hand, 2017), because argumentative writing enables learners to construct explanations of their knowledge of the science topic (Hand, 2017).

Students can experience writing to learn in school science laboratories. For instance, use of the science writing heuristic (SWH) in school science laboratories is one of the most effective ways to engage students with inquiry-based science activities by focusing on written arguments (Stephenson & Sadler-McKnight, 2016), because it encourages students to use an argument structure by posing a testable question(s) about a topic, designing and implementing investigations, defining main outcomes in their collected data, asserting claims based on the outcomes, providing evidence to support their claims, and reflecting on their inquiry learning process (Choi, Hand, & Greenbowe, 2013).

Although hands-on laboratories (HLs) are common in schools, virtual laboratories (VLs) have also entered schools, with recent developments in educational technology (D'Angelo, Rutstein, & Harris, 2016). VLs are laboratories provided mainly by computer technology (de Jong, Linn, & Zacharia, 2013) in which students can perform their investigations with simulated (virtual) equipment (de Jong, Sotiriou, & Gillet, 2014). While many studies have investigated the effects of the SWH in HLs in terms of students' achievement or developing reasonable arguments (Stephenson & Sadler-McKnight, 2016; Vekli & Nazli, 2020; Yaman, 2018), there is little research directly investigating and comparing constructing arguments in hands-on and virtual laboratories. In this study, we aimed to investigate the characteristics of the arguments (e.g. quality of claims, questions, evidence, and multiple representations) created by preservice science teachers (PSTs) using the SWH approach in either a computer-based or hands-on laboratory environment.

### Theoretical Background

The study is grounded in a social constructivist perspective on science instruction. In this perspective, students are able to define their own research question(s) and interact with materials in laboratories to gather and analyze data and reach a conclusion, instead of verifying results taught in lectures (Walker & Sampson, 2013). The SWH approach is an instructional model that gives a central position to argumentation; the students' social construction of scientific knowledge is shaped by the role of argumentation in school science laboratories.

#### Written Argumentation in Science Education

Argumentation has a vital role in science education because it connects the hands-on activities in school science labs (claims and data) with the students' minds-on activities (justification or evaluation of claims through evidence) needed to develop and

understand scientific theories and ideas (Erduran, Ozdem, & Park, 2015; Osborne et al., 2016). Writing argumentative texts should be part of the science curriculum for two main reasons: Writing argumentative text is an important scientific practice, and the process of writing argumentative text may encourage students to enhance their conceptual understanding (Sampson, Enderle, Grooms, & Witte, 2013). The first reason is important because science is not possible without discussion and/or writing; they are the ways scientific knowledge is generated, shared, and communicated (Huerta & Garza, 2019). Therefore, students should be provided with opportunities to learn how to write argumentative texts in order to share their ideas. The latter reason is also vital, since writing argumentative texts requires mental activities that involve interpretation, reflection, justification, and synthesis, and that direct the learner to think deeply about the concept (Hand, 2017; Jang & Hand, 2017). Although there are potential advantages to integrating the writing of argumentative text into the science curriculum, many students have difficulty in sharing their ideas, supporting a claim with evidence, and coordinating evidence and claims (Chen, 2019). This shows that its value in improving their understanding of the topic might not be as strong, if students cannot actually carry it out properly (Hand, 2017).

### Writing in Science Classes: Writing to Learn or Learning to Write

There are many approaches that science teachers can use to get students to write in science classes (Sampson et al., 2013). Three common approaches are learning to write, writing to learn, and a combination of these two approaches (writing to learn by learning to write) (Chen, 2019). In the first approach, learning to write, the goal is improving students' skills that are required to write well in scientific contexts; the focus is on the process, in order to develop writing ability rather than domain knowledge (Sampson et al., 2013). In the second approach, writing to learn, students enhance their domain knowledge through a set of writing activities including arguing, reflecting, and explaining ideas, claims, and/or theories (Chen, 2019; Sampson et al., 2013). The third approach, writing to learn by learning to write, emphasizes that students should not only learn the writing genre, but should also use it as a tool to learn concepts (Chen, 2019). In the current study, we used the writing-to-learn approach, because our goal was not improving learners' writing skills, and we mostly focused on the content of the writing rather than development of the writing process.

The SWH approach is compatible with writing to learn, and was used in this study because it addresses claims, evidence, and the claims-evidence relationship as parts of the argument structure embedded in inquiry learning (Choi et al., 2013). In addition, the SWH approach has a criterion related to multiple representations as an indicator of the argument quality. In this regard, Kelly, Regev, and Prothero (2007) concluded that multiple representations were included in students' high-quality written arguments.

### Advantages and Drawbacks of Hands-on and Virtual Science Laboratories

Laboratories have a central function in science education, and it is clear that it is difficult to teach scientific knowledge without offering practical experiences in laboratories (Bretz, 2019). Developing students' psychomotor skills, enhancing their experience with how to design an experiment and collect data, and dealing with measurement

tools are some critical roles of laboratories in science education (Burkett & Smith, 2016; Kontra, Lyons, Fischer, & Beilock, 2015; Puntambekar, Gnesdilow, Tissenbaum, Narayanan, & Rebello, 2020; Zacharia, 2015). Nevertheless, students have to deal with inconsistency due to daily life variables such as air pressure, which results in measurement errors, and that makes learning difficult in laboratories (Bumbacher, Salehi, Wieman, & Blikstein, 2018; Pyatt & Sims, 2012; Sullivan, Gnesdilow, Puntambekar, & Kim, 2017). Furthermore, HLs may have some limitations for science teachers like quantities that cannot be measured or number of devices available (Nivalainen, Asikainen, Sormunen, & Hirvonen, 2010).

VLs can offer several solutions to deal with these restrictions, as they offer a safer environment, cost and time efficiency, dynamic and rapid data display, and portability (Achuthan & Murali, 2015; Puntambekar et al., 2020). They can also be adaptive with regard to specific topics (de Jong et al., 2013), for example, through transforming unobservable concepts such as electricity (Kollöffel & de Jong, 2013) or genes (Cakir & Dogan, 2015) into observable entities. It is also possible to simplify reality and eliminate more detailed knowledge in order to encourage students to focus on key concepts (Trundle & Bell, 2010). Although VLs have their own affordances (Puntambekar et al., 2020), they are not exact copies of real-world entities (Seoane, Arriassecq, & Greca, 2018), so this can cause the learner difficulty in understanding the real properties of the scientific concept (Burkett & Smith, 2016).

It is obvious that both hands-on and virtual laboratories have some common goals, such as enhancing students' conceptual knowledge, and developing inquiry skills or teamwork abilities (National Research Council [NRC], 2012). However, because each type of laboratory has its own affordances, it is important to understand how each type of laboratory environments supports students' learning (Puntambekar et al., 2020). In this respect, the main target of this study is to reveal how hands-on and virtual laboratories contribute to the quality of learners' written arguments.

## The Purpose of this Study

It is possible to find studies about online argumentation, the effects of technology on students' arguments, or how SWH works in VLs and how this type of laboratory environment affects students' written scientific arguments (Choi & Hand, 2020; Lamb, Etopio, Hand, & Yoon, 2019; Yaman, 2018; Zhu, Liu, & Lee, 2020). Most of these studies show that both HLs and VLs have their own advantages to support students' learning, and the conclusions of these studies try to explain why students might learn differently from HLs and VLs (Puntambekar et al., 2020). Yet, it is indisputable that learning in HLs and VLs is a complex process (Rau, 2020), so more studies are needed to reveal the relations between the properties of laboratory environments and students' learning. For example, in a recent study, Puntambekar et al. (2020) investigated how the two types of laboratory environment (HLs or VLs) encouraged students to share and discuss their ideas. In addition to students' talk, examining their written arguments can also be a way to uncover the relations between the properties of laboratory environments and students' learning. Therefore, in the current study, we investigated the characteristics of PSTs' argumentative writings in hands-on and virtual laboratory environments. In the related literature, Yaman (2018) investigated how a VL environment affected pre-service elementary teachers' written arguments in the context of

multiple representations or in the other study; she (2019) examined the effects of SWH on the quality of PSTs' argumentative writings in a HL. In this study, we examined the quality-related characteristics of arguments developed by using the SWH in a HL or in a VL, and compared them with each other. Furthermore, PSTs' written arguments were evaluated with regard to the quality of claims, questions, and evidence, as well as the use of multiple representations. The purpose of the current study is to reveal and compare the quality of argumentative writing developed by PSTs in HLs and VLs using the SWH approach. We thought that PSTs using a VL would create higher quality arguments, due to the positive contributions made by the educational technology. The main research questions for the study were the following:

- What are the quality-related characteristics of PSTs' written arguments based on a range of experiments in a VL?
- What are the quality-related characteristics of PSTs' written arguments based on a range of experiments in a real HL?
- What are the similarities and differences between PSTs' written arguments based on a range of experiments in a real HL or a VL?

## Method

The study is based on a quasi-experimental research design. There were two conditions in the current study. One group used the HL to design and implement experiments and used a paper-based worksheet to write their arguments; the other group did their investigations through a computer-based VL and wrote their arguments on the online learning platform.

### Participants

Participants in the study were 52 PSTs (40 females, 12 males; $M_{age} = 21$ years, $SD = 1.30$) who were enrolled in the laboratory application in science teaching course in a public university in Turkey. Convenience sampling, which is a type of non-probability sampling where the researcher includes people who are easy to reach, was used. Participants had the same background because they had all been exposed to the same curriculum. For example, they took the same courses, such as curriculum and planning in science education and teaching methods. They enrolled in laboratory courses such as general chemistry laboratory, general physics laboratory, and general biology laboratory. All of these courses were taught in HL environments. They also took a technology course that mainly focuses on core IT skills such as getting familiar with hardware and using some basic office programs. Furthermore, they all had similar scores on the national university entrance exam to be accepted to the department.

The PSTs were assigned to two groups, based on whether the last digit of their student identification number was even or odd. One group went to the class that worked in the control condition (HL), and the other group went to the class that worked in the experimental condition (VL). There were 28 PSTs in the control group, where PSTs designed and implemented their experiments in the HL, and they used a paper laboratory notebook to write their arguments. There were 24 PSTs in the experimental group, who completed the

course in the computer laboratory and designed and implemented their investigations in a VL platform, as well as writing their arguments in the online platform.

PSTs were chosen as the participants because their future career is to be science teachers in middle schools. Therefore, they should be taught about the importance of argumentation, the process of how to use argumentation in a science laboratory, and how to increase the quality of written arguments.

### Instrument

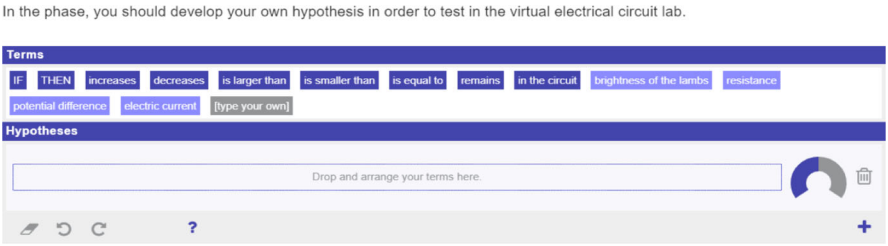
A writing template based on the SWH approach developed by Choi (2008) was used in the study. The template involves five main elements, as follows: preparation for experimentation, predictions, investigation, results, and discussion. In the first step, students write their initial thoughts and questions about the topic. For the predictions section, they write their hypothesis that will be investigated through experimentation. For the next section, students explain how and why they designed and implemented their experiments. In the results section, they present their findings based on their investigations. In the last section, the students write reflections about whether their initial ideas have changed or not. In the template, guiding questions were provided for each section to help students write their arguments.

### The VL Environment Platform: Go-Lab

Go-Lab is an EU-funded project that aims to use innovative educational technologies in STEM education, with a specific focus on VLs and inquiry learning approaches (Go-Lab Sharing and Authoring Platform, 2015). The Go-Lab platform involves three main parts: (i) laboratories, (ii) apps, and (iii) inquiry learning spaces. In terms of laboratories, there are more than 750 virtual and remote laboratories about physics, chemistry, biology, astronomy, engineering, environmental education, earth sciences, technology, and mathematics that enable students to simulate investigations (virtual labs) or offer an opportunity to experiment with real equipment from remote locations (remote labs). The apps in the Go-Lab platform are software tools developed to help students through inquiry learning processes such as developing a hypothesis, designing an experiment, or making predictions. For example, Fig. 1a shows the hypothesis scratchpads used in the current study. This tool helps students to formulate a hypothesis. It shows a given set of concepts about the domain and allows students to combine them to generate hypotheses by dragging and dropping. Students also have a chance to add their own concepts. The last section of the Go-Lab platform is the inquiry learning spaces (ILSs). They are based on an inquiry learning approach and mainly consist of the phases of orientation, conceptualization, investigation, conclusion, and discussion. They include a laboratory, apps, and other kinds of multimedia materials such as videos, photographs, or texts. The main goal of an ILS is to enable students to design and implement experiments using the inquiry learning cycle, with the help of guidance (Go-Lab Sharing and Authoring Platform, 2015).

### Science Writing Heuristic Approach

The science writing heuristic (SWH) approach is a type of scaffolding based originally on the notion of writing to learn and more recently, arguing to learn (Hand, Chen, &




a

**Terms**

- IF -THEN -increases -decreases -is larger than
- is smaller than - is equal to -remains -in the circuit
- the brightness of the lamp(s) - resistance -the potential difference - electric current

Write your hypothesis in the given box.

b

**Fig. 1** **a** Hypothesis scratchpad used in the VL. **b** Hypothesis scaffold used in the HL

Suh, 2020). The approach involves two sets of scaffolds, for both teachers and students in an inquiry-based laboratory environment (Burke, Greenbowe, & Hand, 2006). The set for students provides scaffolds in a written form to support metacognition about their laboratory experiments (Hohenshell & Hand, 2006), and the set for teachers enables them to design inquiry-based science laboratories.

The student writing template gives the titles of the section as follows: questions, claims, the question-claims relationship, evidence, the claims-evidence relationship, and reflection; this template helps students in developing their understanding through argumentation (Nam, Choi, & Hand, 2011). Students are prompted to pose their own questions about the topic, define patterns in their collected data, assert claims based on interpretation of the data, support their claims with evidence, and reflect on their investigations (Choi, 2008).

The teacher's role in the SWH approach is that of enabling students to think reasonably and to construct their knowledge through following an inquiry-based approach, instead of simply giving structured directions (Burke et al., 2006). The template for teachers involves a set of scaffolds that encourage students' reasoning and support their metacognition (Choi, 2008).

## Research Design and Implementation

At the beginning of the study, argumentation and the writing to learn approach were introduced to the PSTs. This process took 2 weeks. Then, because none of the PSTs was familiar with the VL environment, the PSTs who used the VL environment were introduced to the virtual platform. After that, the PSTs followed the course by using the

SWH approach, and working in either a VL or an HL to conduct their inquiry processes. We did not evaluate PSTs' domain knowledge.

PSTs' writing samples were collected over a period of 4 weeks. Every week, each PST designed and implemented their own experiments and wrote their own arguments for four different experiments, related to solutions, acids and bases, electricity, and buoyancy force, respectively. These experiments were chosen because they are all available to do in HLs and VLs. As far as the content of the laboratory investigations, the experiments were drawn from general physics and general chemistry laboratory courses. Table 1 shows the topics, the materials used (by group), and the concepts investigated in the PSTs' experiments.

An inquiry-based laboratory design was used in the current study. The main topic given to PSTs, and they decided their own initial questions, designed the inquiry process, and worked through the other parts of the SWH approach individually. Whereas guidance was provided through a laboratory worksheet for PSTs in the HL, online scaffolding tools were provided for PSTs in the VL. For example, Fig. 1 a and b show examples of the hypothesis generation stage of the third experiment for both conditions.

Both groups had the same amount of time for each experiment (90 min per experiment). The fundamental difference between the two groups was in terms of the laboratory environment. One group did their experiments in an HL environment and wrote their assignments on a paper-based laboratory worksheet. The other group did their experiments in the VL environment and wrote their assignments via an online learning platform. The same instructor taught both of the groups. Because the guidance was provided through either the laboratory worksheet or online tools, the main role of the instructor was providing procedural supports. For example, he helped PSTs in the VL environment only when they had a technical problem(s) due to internet connection or virtual learning platform. For PSTs in the HL environment, the instructor just provided the equipment. He did not give any feedback to PSTs in either of the laboratory environments. At the end of the study, the writing of the PSTs in both groups was gathered and analyzed to reveal the differences in the quality of the writing from the two different groups, both using the SWH approach.

## Data Analysis

Qualitative data collected during the study were transformed into quantitative data. The quality of PSTs' science writing was scored using the final rubric developed by Choi (2008), which was based on a literature review, input from science educators, and extensive use in evaluating student writing. The rubric (Appendix A in the Supplementary Information) uses scores from 1 (*weak writing quality*) to 5 (*very high writing quality*) for each section of the SWH writing template. The sections in the rubric are as follows: questions, claims, the questions-claims relationship, evidence, the claims-evidence relationship, multimodal representations, and reflection. For example, a score of 4 points was given if the questions section included at least one open-ended testable question, and a score of 1 point if there was a single closed-ended and not testable question. In terms of claims, 2 points were given if there were multiple claims but they did not appear to come from the investigation(s), and 5 points if there were multiple claims that were based on interpretation of the experimental data. For the

**Table 1** Topics, materials (by group), and concepts

| | Control group (hands-on lab.) | Experimental group (virtual lab.) | Concepts investigated in the experiment |
|---|---|---|---|
| 1st experiment—solutions | Chemicals ($K_2CrO_4$, $K_2Cr_2O_7$, $KMnO_4$, etc.), test tubes, beaker, Erlenmeyer flask, Florence flask, graduated cylinder, glass funnel, assay balance, spatula, wash bottle, pipet, glass stirring rod | Virtual lab (molarity) (https://www.golabz.eu/lab/molarity) | Mole, molarity, the concentration of a solution |
| 2nd experiment—acid and base | Chemicals (HCl, $H_2SO_4$, $NH_3$, NaOH, etc.), test tubes, buret, ring stand, ring clamp, Erlenmeyer flask, Florence flask, graduated cylinder, glass funnel, beaker | Virtual lab (acid-base solutions) (https://www.golabz.eu/lab/acid-base-solutions) | Concentrated and dilute solutions, strong/weak acid or base, pH |
| 3rd experiment—electricity | Cable, ammeter, voltmeter, bulbs, switches, resistors | Virtual lab (electrical circuit lab) (https://www.golabz.eu/lab/electrical-circuit-lab) | Serial/parallel circuits, Ohm's law, ammeter, voltmeter |
| 4th experiment—buoyancy force | Various liquids (pure water, saltwater, mineral water, sunflower oil, etc.), graduated cylinder, beaker, assay balance, small materials (e.g. eraser, rock) | Virtual lab (splash: virtual buoyancy laboratory) (https://www.golabz.eu/lab/splash-virtual-buoyancy-laboratory) | Archimedes' principle |

questions-claims relationship, 3 points were given if the question(s) and claim(s) fit together reasonably well, and 4 points if they fit together strongly. For the evidence section, 2 points were given if there was weak evidence that was just a description of the data and not very accurate and reliable, and 5 points if very accurate, reliable, and powerful evidence with interpretations based on the data gathered from the experiment(s) was presented. For the claims-evidence relationship, 2 points were given if the evidence supported the claim(s) loosely or inadequately, and 4 points if there was a strong relationship between the evidence and the claim(s). In terms of multimodal representations, 1 point was given if there was no representation or only text, and 3 points if there were three kinds of representations. Finally, for the reflection section, 2 points were given if there was a weak explanation for why ideas had changed or not changed and there was a weak link between PST's investigation(s) and the knowledge used in the explanation, and 5 points if there was a strong explanation about change (or no change) in the PST's ideas that involved real-life applications based on the experiment(s). A mean score for each group was calculated for every section. In the end, three main assumptions (independence, normality, and homogeneity of variance) were checked before conducting the independent samples *t* test used to compare the mean scores of the groups for each section.

Researcher triangulation (Patton, 1999) was used in the current study to ensure the reliability of the scoring. Two different researchers independently scored the 208 pieces of writing (52 PSTs individually wrote either paper-based or online reports for four experiments) created by the PSTs; the inter-rater reliability coefficients ($\kappa$) are shown in Table 2. This process aimed to avoid researcher bias as much as possible. In the end, their results were compared and they discussed any different scores given until consensus was reached.

## Results

In the current study, the characteristics and quality of argumentative writings produced by PSTs via the SWH approach in HLs and VLs were investigated and compared.

### The Characteristics of PSTs' Written Arguments Based on Experiments in a VL

The total mean scores for each section for PSTs in the VL environment are given in Table 3 (see Appendix B in the Supplementary Information for the means and standard

**Table 2** Inter-rater reliability coefficients ($\kappa$) for the PSTs' 208 written arguments

| | Q | C | E | M | R | Q-C | C-E |
|---|---|---|---|---|---|---|---|
| Lab 1 | 0.901 | 0.781 | 0.880 | 0.930 | 0.812 | 0.865 | 0.896 |
| Lab 2 | 0.846 | 0.820 | 0.803 | 0.927 | 0.789 | 0.885 | 0.818 |
| Lab 3 | 0.891 | 0.851 | 0.878 | 0.905 | 0.785 | 0.890 | 0.880 |
| Lab 4 | 0.921 | 0.890 | 0.820 | 0.923 | 0.715 | 0.903 | 0.850 |

*Q* questions, *C* claims, *E* evidence, *M* multimodal, *R* reflection, *Q-C* questions-claims relationship, *C-E* claims-evidence relationship

**Table 3** Overall mean (SD) for each section score per condition (max = 5)

| Groups | Q | C | E | M | R | Q-C | C-E |
|---|---|---|---|---|---|---|---|
| HL ($n$ = 28) | 2.90 (0.83) | 2.85 (0.94) | 3.35 (0.87) | 1.86 (1.32) | 2.71 (0.97) | 3.75 (0.79) | 3.85 (0.87) |
| VL ($n$ = 24) | 3.16 (0.67) | 2.88 (0.94) | 3.57 (0.56) | 2.34 (0.54) | 2.81 (0.67) | 3.65 (0.75) | 3.89 (0.84) |

*Q* questions, *C* claims, *E* evidence, *M* multimodal, *R* reflection, *Q-C* questions-claims relationship, *C-E* claims-evidence relationship; *HL* hands-on laboratory, *VL* virtual laboratory

deviations of each section score for each laboratory activity per condition). The section about the claims-evidence relationship had the highest mean score, while the lowest score was found for the multimodal section.

The mean score for the questions section was around 3 (intermediate) for the PSTs in the VL condition, which means that the questions generated by the PSTs were of average quality, multiple questions that were primarily closed-ended, with a few open-ended questions. Questions mostly matched the essence of the investigation, and they were mostly testable and meaningful questions. As an example of open-ended and testable questions, one PST in the VL environment stated his initial thought and posed a question about buoyancy force as follows: "Upward force is applied to the objects left in the liquid by the liquid. So, how does this force applied by liquids to the objects occur?"

The mean score for the claims section fell between 2 and 3 (weak and intermediate) for PSTs in the VL environment. This means that PSTs proposed single or multiple claims that were not adequately qualified. For instance, a PST in the VL condition proposed a claim about electricity: "If resistance is added to the circuit, it reduces the current in the circuit." This claim was coded as weak, due to some deficiencies; for instance, it is not clear what type of circuit is meant.

The mean score for the questions-claims relationship section fell between 3 and 4 (intermediate and strong connection). This means that the questions and claims proposed by the PSTs fit reasonably well. The claims focused on the questions but were loosely connected to them. The mean score for the evidence section fell around at 3 (intermediate). This means that the evidence mostly came from other sources (i.e. textbooks) with limited interpretation, but was generally valid and reliable evidence. For example, a PST presented her evidence as follows:

> A weak acid solution with a 0.001M concentration transmits electricity weakly. When I increased the concentration of the solution to 0.01 M, I observed that it conducted electricity better. When I increased the concentration of the solution to 0.1 M, I observed that the lamp burned stronger than the one in 0.01 M. When I applied the same process to strong acid, strong base, and weak base solutions, I observed that the results were the same.

The mean score for the claims-evidence relationship section fell between 3 and 4 (intermediate and strong relationship). This section had the highest score for the VL group. This score means that the evidence supported most of the proposed claims. The mean score for the multimodal section fell between 1 and 2 (very weak and weak). The PSTs mostly presented their data through one mode (just text) or, at most, two modes

(i.e. text and graph, text and formula (equation), or text and chemical representation). This section was the weakest one for the PSTs in the VL environment. The mean score for the reflection section fell around at 2 (weak). The PSTs presented weak explanations for why their ideas had changed or not. It was difficult for PSTs to link their own investigation to their existing knowledge. For example, one PST reflected about the inquiry process she followed:

> At first, I did brief research on the internet about preparing solutions. Then I examined the experimental setup and thought about how I could generate a hypothesis. In the experimentation phase, I tested the accuracy of my hypothesis and recorded my observations. Based on the findings of the experiments, I made a generalization and conclusion.

### The Characteristics of PSTs' Written Arguments Based on Experiments in an HL

The PSTs' written arguments in the HL environment were generally of intermediate quality; they had relatively higher mean scores for evidence, the questions-claims relationship, and the claims-evidence relationship. Table 3 shows the total mean scores for each section for PSTs in the HL environment (see Appendix B in the Supplementary Information for the means and standard deviations of each section score for each laboratory activity per condition). Similar to the PSTs in the VL condition, the highest mean score for those in the HL condition was for the section about the claims-evidence relationship and the lowest mean score was for multimodal representations. The quality of questions posed by PSTs in the HL condition was mainly at the intermediate level. Although several questions were posed, they were mostly closed-ended and difficult to test. For example, one PST asked: "What does the density of an object depend on?"

For the claims section, PSTs proposed single or multiple claims, but they were usually unclear. "If the number of bulbs in the circuit is increased, the brightness decreases" is an example of a claim by a PST from the HL condition. In terms of the questions-claims relationship, there were relatively strong relations between them; the claims mostly depended on the questions. For the evidence section, the evidence given was usually valid and reliable, but there was limited interpretation based on the evidence. For instance, a PST presented her evidence as: "I observed that when I increased the concentrations of acid solutions, their pH decreased; on the other hand, when I increased the concentrations of basic solutions, their pH also increased." The claims-evidence relationship had the highest mean score, and the evidence strongly supported the claims. On the other hand, the section related to multimodal representations had the lowest mean score. This shows that PSTs mostly used only text when presenting their claim(s), evidence, or reflection. The PSTs also made weak reflections. It was difficult for them to explain how their prior knowledge and investigations were related to each other. The following example shows that a PST from the HL did not reflect on the process, but reflected on the solution experiments he did:

> In my experiments, I tested the effects of the type, amount and temperature of the substances on solutions and concluded that these are very important variables. I also observed that mixing affects the speed of dissolution.

### Comparison of PSTs' Written Arguments Based on Experiments in a Real HL and a VL

For the next step, the total mean score for each section for the PSTs in the HL and VL environments was compared in independent samples *t* tests. Table 4 shows the results.

As shown in Table 4, significant differences were found for the sections on questions, evidence, and multimodal representations. For all three of these sections, students who were taught in a VL environment reached significantly higher mean scores than their counterparts who were taught in a real HL environment. The claims-evidence relationship section had the highest mean score for PSTs in both conditions, and the multimodal section had the lowest mean score. PSTs in the HL environment achieved a better score than their counterparts only in the section on the questions-claims relationship, but this minor difference was not statistically significant. For all other sections, the written arguments of PSTs in the VL environment were slightly better than those of the PSTs in the HL environment.

## Discussion

In the current study, PSTs' written arguments created in hands-on and virtual laboratory environments were analyzed and compared. Analysis of written arguments created in the HL and VL environments revealed that use of the SWH approach in an inquiry-based science laboratory enabled PSTs to create moderately well-written arguments. The PSTs generated questions, designed and implemented investigations, gathered and analyzed data, proposed claims, provided evidence, and reflected on their inquiry process at an intermediate level.

In terms of the characteristics of the written arguments created in the VL, the generated questions were mainly testable but there were also some closed-ended questions. The claims made did not usually come from the investigations. PSTs typically provided accurate and reliable evidence with explanations based on the data gathered through their investigations. However, they rarely used multimodal representations, and mostly used text and chemical representations in their arguments. They reflected on their own inquiry moderately well, meaning that they were usually able to

**Table 4** Pairwise comparisons of PSTs' writing scores in the two conditions, by section

| Paired differences between VL and HL | Mean | *t* | df | Cohen's *d* |
|---|---|---|---|---|
| Questions | 0.254 | 2.391* | 206 | 0.34 |
| Claims | 0.027 | 0.204 | 206 | 0.03 |
| Evidence | 0.215 | 2.093* | 206 | 0.30 |
| Multimodal representations | 0.478 | 3.318** | 206 | 0.47 |
| Reflection | 0.098 | 0.835 | 206 | 0.12 |
| Questions-claims relationship | − 0.104 | − 0.969 | 206 | 0.13 |
| Claims-evidence relationship | 0.028 | 0.237 | 206 | 0.05 |

Note: $*p < .05$, $**p = .01$

explain why their ideas did or did not change. In terms of the relations between questions and claims, PSTs made connections moderately well. For the relations between claims and evidence, the evidence provided mostly strongly supported the claims made. Overall, it can be said that PSTs produced reasonable written arguments in the VL environment.

In terms of the characteristics of the written arguments created in the HL, PSTs posed questions that were mainly not testable. They usually made multiple claims that were not based on their experiments. They provided valid and reliable evidence, but with limited interpretation. Supporting claims with multiple representations was the weakest section for the PSTs in the HL environment. They mostly preferred one mode (texts) in their arguments. In PSTs' reflections on their own inquiry process, it was difficult for them to link their own investigations with their existing knowledge. Their questions and claims usually fit together strongly, and there were strong connections between their claims and evidence.

Overall, the characteristics of the written arguments created in HLs and VLs show that inquiry-based laboratory activities and use of the SWH approach have the potential to support learners in constructing reasonable arguments. Similar results were also found by other researchers (Chen, Hand, & Park, 2016a; Vekli & Nazli, 2020; Yaman, 2018, 2019). The basic difference between the current study and Yaman's (2018) work is that PSTs in this study used multiple representations only occasionally for providing evidence to support claims (see Appendix B in the Supplementary Information). A similar conclusion was also reached by Vekli and Nazli (2020). Both of the studies (Vekli & Nazli, 2020; Yaman, 2018) stated that the SWH approach's template enabled PSTs to improve their representational competence. Yet, the PSTs in the current study did not achieve this. It might be related to the topics covered in this study.

PSTs who were taught in the VL environment wrote relatively better arguments than the students in the HL environment, based on the descriptive data. Significant differences based on the groups' mean scores for each section were found for the questions, evidence, and multimodal representations sections.

Developing testable and open-ended questions in inquiry-based laboratories is crucial because such questions sustain the lessons (Marbach-Ad & Claassen, 2001). Students' questions have the potential to initiate argumentation, to form explanatory hypotheses, and to examine the evidence to determine if it supports the hypotheses (Chin & Osborne, 2010). Posing proper questions also has an impact on students' higher-order thinking skills (Kuhn, 2009). Chin and Osborne (2008) stated that posing questions during the learning process in one of the most powerful indicators of engaged learning. Questioning is clearly an important step in inquiry learning. In our case, although PSTs in both laboratory environments had similarly intermediate mean scores for the questions section, statistical comparison revealed that PSTs in the VL environment posed better questions than their counterparts in the HL environment, probably due to use of computer technology, because PSTs in the VL had the opportunity to design more complicated experiments. Being able to access and use the internet may also have been an advantage for PSTs in the VL. Hu and Chiou (2012) stated that virtual environments allow students to think of, ask, and answer more questions because such environments alleviate the pressure on students due to the classroom atmosphere. For example, students sometimes feel uncomfortable asking questions in the classroom, because they think that their questions can cause negative reactions from

their classmates and/or teachers (Dillon, 1988). Virtual environments may offer a solution in such cases. Students are also able to access, update, revise and manage their questions in a virtual environment (Yu, Liu, & Chan, 2005), so they can develop their questions during the learning process. These could all be reasons that PSTs in the VL posed more testable and open-ended questions.

Another significant difference was found regarding evidence. In the current study, we assumed that PSTs had similar content knowledge about the topic, because they had all taken the same courses and passed the same university entrance exam. The data showed that PSTs in both groups presented moderately strong evidence, with explanations based on data gathered through experiment(s) to support their claims (Choi et al., 2013), but statistical analyses revealed a significant difference between the groups, with a better result for the PSTs taught in a VL environment. Many students have difficulty in supporting their claims with strong evidence (Kelly & Takao, 2002; Sandoval, 2003) and a lack of well-developed evidence causes the purpose of argumentation to fail (Gleason, 1999). For students to support their claims with strong and valid evidence based on their data, interpretation is an important step, because Chen, Hand, and McDowell (2013) concluded that the strength of the relationship between claims and evidence predicts success on achievement tests. In the VL environment, PSTs were able to use scaffolding tools that enabled them to relate their findings with their questions or claims. The scaffolds, in this case, were online tools available to support PSTs' performance beyond their capacities (Cho & Jonassen, 2002). Such tools assist students to develop an idea based on the data they gathered or reached in the light of other evidence from their classroom (Clark & Sampson, 2007). Several studies (e.g. Cho & Jonassen, 2002; McNeill, Lizotte, Krajcik, & Marx, 2006; Oh & Jonassen, 2007) in the literature have also shown that scaffolding tools have a positive impact on argumentation skills, especially on presenting evidence. For example, McNeill et al. (2006) concluded that scaffolds lead to creating reliable and valid explanations. In our study, providing questions or claims with the corresponding findings via the scaffolding tools may have facilitated PSTs in presenting valid evidence in the VL environment.

Another important difference was seen in terms of multimodal representations. Multimodal representations should be seen as key elements of argumentation to support the quality of evidence and the relation between claim and evidence (Hand et al., 2020). Therefore, it is important to be able to use them during the learning process. PSTs in VL environments were able to show data in graphs or tables. Yet, it was difficult for the other group of PSTs to use multiple representation levels on their paper-based worksheets, because they were mainly focused on writing, and did not focus much on drawing graphs or tables. VLs have the potential to present data in different forms, such as diagrams, graphics, animations, videos (Ainsworth & van Labeke, 2002; Blake & Scanlon, 2007), and interactive dynamic visuals (van der Meij & de Jong, 2006). The other advantage of a VL environment is that it can provide information in real and theoretical forms simultaneously (McFarlane & Sakellariou, 2002), which helps learners to relate to and translate between different representations (van der Meij & de Jong, 2006). In the current study, although PSTs in the VL environment did not take advantage of using multiple representations as much as expected (the mean score for the multimodal section was 2.34, which is weak), their overall mean score was still significantly better than for PSTs in the HL.

Lastly, it is interesting to consider the outcomes of the current study concerning with regard the participants' age. Although some studies have been done with PSTs (e.g. Hand, Nam, & Choi, 2012; Yaman, 2018), most studies related to the SWH approach have been done with K-12 students. Yet, the outcomes of these studies are mainly coherent, regardless of the age of the participants. In other words, the SWH approach works well with both K-12 students and PSTs. Two critical points related to improving the quality of students' written arguments may be time and practice, rather than their age (Chen et al., 2016a).

## Conclusion

The outcomes of this study revealed that PSTs in HLs and VLs were able to construct written arguments of intermediate or moderate quality, respectively, when they carried out the inquiry process using the SWH approach. PSTs in a VL had more advantages, due to the computer technology, than their counterparts in the HL. An important conclusion of the current study is that the SWH approach can be used as well in VL environments as in HL environments. Furthermore, some studies have shown that using virtual and in-class discussion together (Choi & Hand, 2020) or virtual and textbook instruction in combination (Lamb & Etopio, 2019) yields better written products from students. Therefore, investigating the effects on students' written arguments of using HLs and VLs in combination could be a topic for future research. Based on the current study, it is difficult to reach a definite conclusion about which type of laboratory environment is better. We can conclude that each type of laboratory environment has its own advantages. For example, PSTs were able to focus on key points about the topic in the VL environment, and this may have helped them to spend more time on writing arguments. A similar conclusion was reached by Puntambekar et al. (2020). Moreover, while it is easier in the VL environment to use multimodal representations when presenting evidence or to do more experiments in a limited time to gather more evidence, the investigations students can do depend on the limits of the VL. On the other hand, students in the HL environment can design and implement experiments based on their research questions and touch sensory input is provided in this type of laboratory. Yet, time and cost inefficiency and the difficulty of observing submicroscopic or huge entities (e.g. molecules or celestial bodies) are some limitations of HLs.

Consequently, the findings of the current study have several implications for science teachers. First, when VLs are considered as an alternative for HLs due to their practical advantages, the SWH approach can be integrated into the learning process to embed argumentation in science classes. Studies have shown that VLs enable students to focus more on the relations between variables and/or claims and evidence, rather than setting up an experiment or measurement and calculations (Puntambekar et al., 2020), which helps students achieve meaningful understanding. Second, an open inquiry process was followed in the study, so science teachers can provide more structured scaffolding to enhance the quality of students' arguments. There was a relatively low use of multiple representations in the current study. Teachers can also guide their students to focus on multimodal representations when providing evidence to support their claims (Vekli & Nazli, 2020). Further studies are needed to reach more certain conclusions because this

study had a relatively small number of participants and the length of the intervention was relatively brief. Furthermore, how the instructor implemented the SWH approach was not investigated, but this can be an important element affecting how students develop an argument (Akkus, Gunel, & Hand, 2007). Lastly, an interview with the participants is needed in future studies to access their perspectives about argumentation based on a VL; similar studies can also be done with primary, middle, or high school students to investigate how the SWH approach works in VLs.

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1007/s10763-020-10147-w.

## References


Achuthan, K., & Murali, S. S. (2015). A comparative study of educational laboratories from cost and learning effectiveness perspective. In R. Silhavy, R. Senkerik, Z. K. Oplatkova, Z. Prokopova, & P. Silhavy (Eds.), *Software engineering in intelligent systems* (pp. 143–153). Cham, Switzerland: Springer Publishing.

Ainsworth, S., & van Labeke, N. (2002). *Using a multi-representational design framework to develop and evaluate a dynamic simulation environment.* Paper presented at the Dynamic Information and Visualization Workshop, Tübingen, Germany.

Akkus, R., Gunel, M., & Hand, B. (2007). Comparing an inquiry-based approach known as the science writing heuristic to traditional science teaching practices: Are there differences? *International Journal of Science Education, 29*(14), 1745–1765.

Blake, C., & Scanlon, E. (2007). Reconsidering simulations in science education at a distance: Features of effective use. *Journal of Computer Assisted Learning, 23*(6), 491–502.

Bretz, S. L. (2019). Evidence for the importance of the laboratory courses. *Journal of Chemical Education, 96*(2), 193–195.

Bumbacher, E., Salehi, S., Wieman, C., & Blikstein, P. (2018). Tools for science inquiry learning: Tool affordances, experimentation strategies, and conceptual understanding. *Journal of Science Education and Technology, 27*(3), 215–235.

Burke, K. A., Greenbowe, T. J., & Hand, B. (2006). Implementing the science writing heuristic in the chemistry laboratory. *Journal of Chemical Education, 83*(7), 1032–1038.

Burkett, V. C., & Smith, C. (2016). Simulated vs. hands-on laboratory position paper. *Electronic Journal of Science Education, 20*(9), 8–24.

Cakir, M., & Dogan, O. K. (2015). Teaching Mendelian genetics with a virtual Drosophila laboratory. *School Science Review, 97*(358), 10–15.

Chen, Y. C. (2019). Writing as an epistemological tool: Perspectives from personal, disciplinary, and sociocultural landscapes. In V. Prain & B. Hand (Eds.), *Theorizing the future of science education research* (pp. 115–132). Cham, Switzerland: Springer.

Chen, Y. C., Hand, B., & McDowell, L. (2013). The effects of writing-to-learn activities on elementary students' conceptual understanding: Learning about force and motion through writing to older peers. *Science Education, 97*(5), 745–771.

Chen, Y. C., Hand, B., & Park, S. (2016a). Examining elementary students' development of oral and written argumentation practices through argument-based inquiry. *Science & Education, 25*(3–4), 277–320.

Chen, Y. C., Park, S., & Hand, B. (2016b). Examining the use of talk and writing for students' development of scientific conceptual knowledge through constructing and critiquing arguments. *Cognition and Instruction, 34*(2), 100–147.

Chin, C., & Osborne, J. (2008). Students' questions: A potential resource for teaching and learning science. *Studies in Science Education, 44*(1), 1–39.

Chin, C., & Osborne, J. (2010). Supporting argumentation through students' questions: Case studies in science classrooms. *Journal of the Learning Sciences, 19*(2), 230–284.

Cho, K.-L., & Jonassen, D. H. (2002). The effects of argumentation scaffolds on argumentation and problem solving. *Educational Technology Research and Development, 50*(3), 5–22.

Choi, A. (2008). *A study of written argument using the science writing heuristic approach in inquiry-based freshman general chemistry laboratory classes* (Doctoral dissertation). Retrieved from https://search.proquest.com/docview/304609585

Choi, A., & Hand, B. (2020). Students' construct and critique of claims and evidence through online asynchronous discussion combined with in-class discussion. *International Journal of Science and Mathematics Education, 18*(6), 1023–1040.

Choi, A., Hand, B., & Greenbowe, T. (2013). Students' written arguments in general chemistry laboratory investigations. *Research in Science Education, 43*(5), 1763–1783.

Clark, D. B., & Sampson, V. D. (2007). Personally-seeded discussions to scaffold online argumentation. *International Journal of Science Education, 29*(3), 253–277.

D'Angelo, C. M., Rutstein, D., & Harris, C. J. (2016). Learning with STEM simulations in the classroom: Findings and trends from a meta-analysis. *Educational Technology, 56*(3), 58–61.

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

de Jong, T., Sotiriou, S., & Gillet, D. (2014). Innovations in STEM education: The Go-Lab federation of online labs. *Smart Learning Environments, 1*(3), 1–16.

Dillon, J. T. (1988). The remedial status of student questioning. *Journal of Curriculum Studies, 20*(3), 197–210.

Erduran, S., Ozdem, Y., & Park, J. Y. (2015). Research trends on argumentation in science education: A journal content analysis from 1998-2014. *International Journal of STEM Education, 2*(1), 1–12.

Gleason, M. M. (1999). The role of evidence in argumentative writing. *Reading & Writing Quarterly, 15*(1), 81–106.

Go-Lab Sharing and Authoring Platform. (2015). https://www.golabz.eu Accessed 27 January 2020.

Hand, B. (2017). Exploring the role of writing in science: A 25-year journey. *Literacy Learning: The Middle Years, 25*(3), 16–23.

Hand, B., Chen, Y. C., & Suh, J. K. (2020). Does a knowledge generation approach to learning benefit students? A systematic review of research on the science writing heuristic approach. *Educational Psychology Review*, 1–43. https://doi.org/10.1007/s10648-020-09550-0.

Hand, B., Nam, J., & Choi, A. (2012). Argument-based general chemistry laboratory investigations for pre-service science teachers. *Educación Química, 23*(1), 96–100.

Hohenshell, L., & Hand, B. (2006). Writing-to-learn strategies in secondary school cell biology: A mixed method study. *International Journal of Science Education, 28*(2–3), 261–289.

Hu, H.-W., & Chiou, G.-F. (2012). The types, frequency and quality of elementary pupils' questions in an online environment. *The Turkish Online Journal of Educational Technology, 11*(4), 325–335.

Huerta, M., & Garza, T. (2019). Writing in science: Why, how, and for whom? A systematic literature review of 20 years of intervention research (1996-2016). *Educational Psychology Review, 31*, 533–570.

Jang, J. Y., & Hand, B. (2017). Examining the value of a scaffolded critique framework to promote argumentative and explanatory writings within an argument-based inquiry approach. *Research in Science Education, 47*(6), 1213–1231.

Kelly, G. J., Regev, J., & Prothero, W. (2007). Analysis of lines of reasoning in written argumentation. In S. Erduran & M. P. Jimenez-Aleixandre (Eds.), *Argumentation in science education: Perspectives from classroom-based research* (pp. 137–158). Dordrecht, the Netherlands: Springer.

Kelly, G. J., & Takao, A. (2002). Epistemic levels in argument: An analysis of university oceanography students' use of evidence in writing. *Science Education, 86*(3), 314–342.

Kollöffel, B., & de Jong, T. (2013). Conceptual understanding about 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.

Kontra, C., Lyons, D. J., Fischer, S. M., & Beilock, S. L. (2015). Physical experience enhances science learning. *Psychological Science, 26*(6), 737–749.

Kuhn, D. (2009). Do students need to be taught how to reason? *Educational Research Review, 4*(1), 1–6.

Lamb, R. L., & Etopio, E. (2019). Virtual reality simulations and writing: A neuroimaging study in science education. *Journal of Science Education and Technology, 28*(5), 542–552.

Lamb, R. L., Etopio, E., Hand, B., & Yoon, S. Y. (2019). Virtual reality simulation: Effects on academic performance within two domains of writing in science. *Journal of Science Education and Technology, 28*(4), 371–381.

Leinonen, T., Keune, A., Veermans, K., & Toikkanen, T. (2016). Mobile apps for reflection in learning: A design research in K-12 education. *British Journal of Educational Technology, 47*(1), 184–202.

Marbach-Ad, G., & Claassen, L. A. (2001). Improving students' questions in inquiry labs. *The American Biology Teacher, 63*(6), 410–419.

McFarlane, A., & Sakellariou, S. (2002). The role of ICT in science education. *Cambridge Journal of Education, 32*(2), 219–232.

McNeill, K. L., Lizotte, D. J., Krajcik, J., & Marx, R. W. (2006). Supporting students' construction of scientific explanations by fading scaffolds in instructional materials. *Journal of the Learning Sciences, 15*(2), 153–191.

Nam, J., Choi, A., & Hand, B. (2011). Implementation of the science writing heuristic (SWH) approach in 8th grade science classrooms. *International Journal of Science and Mathematics Education, 9*(5), 1111–1133.

Nivalainen, V., Asikainen, M. A., Sormunen, K., & Hirvonen, P. E. (2010). Preservice and inservice teachers' challenges in the planning of practical work in physics. *Journal of Science Teacher Education, 21*(4), 393–409.

National Research Council (NRC). (2012). *A framework for k-12 science education: Practices, crosscutting concepts, and core ideas*. Washington, DC: National Academy Press.

Oh, S., & Jonassen, D. H. (2007). Scaffolding online argumentation during problem solving. *Journal of Computer Assisted Learning, 23*(2), 95–110.

Osborne, J., Henderson, J. B., MacPherson, A., Szu, E., Wild, A., & Yao, S. Y. (2016). The development and validation of a learning progression for argumentation in science. *Journal of Research in Science Teaching, 53*(6), 821–846.

Patton, M. Q. (1999). Enhancing the quality and credibility of qualitative analysis. *Health Services Research, 34*(5, part 2), 1189–1208.

Puntambekar, S., Gnesdilow, D., Tissenbaum, C. D., Narayanan, N. H., & Rebello, N. S. (2020). Supporting middle school students' science talk: A comparison of physical and virtual labs. *Journal of Research in Science Teaching*. https://doi.org/10.1002/tea.21664.

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.

Rau, M. A. (2020). Comparing multiple theories about learning with physical and virtual representations: Conflicting or complementary effects? *Educational Psychology Review, 32*, 297–325.

Sampson, V., Enderle, P., Grooms, J., & Witte, S. (2013). Writing to learn by learning to write during the school science laboratory: Helping middle and high school students develop argumentative writing skills as they learn core ideas. *Science Education, 97*(5), 643–670.

Sandoval, W. A. (2003). Conceptual and epistemic aspects of students' scientific explanations. *Journal of the Learning Sciences, 12*(1), 5–51.

Seoane, M. E., Arriassecq, I., & Greca, I. M. (2018). Epistemological debate underlying computer simulations used in science teaching: The designers' perspective. In M. E. de Brzezinski Prestes & C. C. Silva (Eds.), *Teaching science with context* (pp. 405–417). Cham, Switzerland: Springer.

Stephenson, N. S., & Sadler-McKnight, N. P. (2016). Developing critical thinking skills using the science writing heuristic in the chemistry laboratory. *Chemistry Education Research and Practice, 17*(1), 72–79.

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.

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.

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.

Vekli, G. S., & Nazli, C. (2020). Examining the effect of reporting based on science writing heuristic approach on written argument quality in general biology laboratory. *Journal of Biological Education*, 1–15. https://doi.org/10.1080/00219266.2020.1785924.

Walker, J. P., & Sampson, V. (2013). Learning to argue and arguing to learn: Argument-driven inquiry as a way to help undergraduate chemistry students to learn how to construct arguments and engage in argumentation during a laboratory course. *Journal of Research in Science Teaching, 50*(5), 561–596.

Yaman, F. (2018). Investigation of multiple levels of representations in students' written argument using virtual chemistry laboratory. *Elementary Education Online, 18*(1), 207–225.

Yaman, F. (2019). Effects of the science writing heuristic approach on the quality of prospective science teachers' argumentative writing and their understanding of scientific argumentation. *International Journal of Science and Mathematics Education, 16*(3), 421–442.

Yu, F. Y., Liu, Y. H., Chan, & T. W. (2005). A web-based learning system for question-posing and peer assessment. *Innovations in Education and Teaching International, 42*(4), 337–348.

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*, 116–137.

Zhu, M., Liu, O. L., & Lee, H. S. (2020). The effect of automated feedback on revision behavior and learning gains in formative assessment of scientific argument writing. *Computers & Education, 143*, 103668. https://doi.org/10.1016/j.compedu.2019.103668.