DEVELOPMENT ARTICLE

Providing guidance in virtual lab experimentation: the case of an experiment design tool

Charalampos Efstathiou[1] · Tasos Hovardas[1] · Nikoletta A. Xenofontos[1] · Zacharias C. Zacharia[1] · Ton deJong[2] · Anjo Anjewierden[2] · Siswa A. N. van Riesen[2]

Abstract The present study employed a quasi-experimental design to assess a computer-based tool, which was intended to scaffold the task of designing experiments when using a virtual lab for the process of experimentation. In particular, we assessed the impact of this tool on primary school students' cognitive processes and inquiry skills before and after the study's treatment, using pre- and post-tests. Our research design involved a group of students who used the computer-based tool/scaffold to design the study's experiments (experimental condition) and a group of students who used a paper-and-pencil worksheet as a scaffold to design the same experiments (control condition). The primary finding of the study was that the use of the computer-based experiment design tool had a more positive effect on students' inquiry skills related to identifying variables and designing investigations than the paper-and-pencil one. This might be attributed to the functionalities provided only by the computer-based experiment design tool, which enabled students to focus their attention on crucial aspects of the task of designing experiments through (1) maintaining values for constant variables when planning experimental trials and (2) the provision of instant feedback when classifying variables into independent, dependent and controlled variables. Moreover, students in the two conditions displayed differing patterns of interactions among cognitive process and inquiry skills. Implications for designing and assessing similar computer-based scaffolds are discussed.

Keywords Experimental design · science education · Virtual labs · Inquiry skills

✉ Zacharias C. Zacharia
   zach@ucy.ac.cy

1   University of Cyprus, Nicosia, Cyprus

2   University of Twente, Enschede, The Netherlands

Introduction

Experimentation has been at the forefront of science education and at the heart of inquiry-based learning (van Joolingen and Zacharia 2009). It is highly valued because of its positive impact on students' science learning (Hofstein and Lunetta 2004; Zacharia 2015; Zacharia and de Jong 2014). However, due to its complex and demanding nature, it poses several challenges to students, especially during the design phase (Zacharia et al. 2015). In the experimental design phase, students are expected to identify which are the independent and dependent variables, specify which dependent variable will be investigated in relation to which independent variables in each round of experimentation, and follow the principles for ensuring that a fair experiment is conducted (i.e., by varying one independent variable at a time while keeping all other variables the same). Needless to say, failure to implement an appropriate experimental design will lead to failure in addressing the hypotheses under investigation (Arnold et al. 2014; Pedaste et al. 2015). As a result, teachers are puzzled about how to support their students during the experimental design phase (Furtak 2006; Kirschner et al. 2006). The latter becomes even a greater challenge, if you consider that there are several researchers arguing that the familiarization with experimental design should begin as early as primary school (e.g., Klahr and Nigam 2004).

In this study, we aimed to contribute to this line of research. In particular, we examined how a computer-based tool for designing experiments ("Experiment Design Tool", henceforth "EDT") could support primary school students' learning, particularly their cognitive processes (i.e., "to remember", "to understand", "to apply" and "to think critically and creatively") and certain inquiry skills (i.e., "identifying variables", "stating hypotheses", "operationally defining, and designing investigations"), when using virtual labs in a computer-supported inquiry learning (CoSIL) environment. We used the Go-Lab CoSIL environment (Go-Lab Sharing and Authoring Platform 2015), which includes, among other resources, virtual labs and a tool for designing experiments. For assessing the impact of the computer-based EDT, we implemented a quasi-experimental design that involved two conditions. The first one made use of the computer-based EDT and the second one made use of a paper-and-pencil worksheet, which was designed to scaffold the process of designing an experiment. The two conditions differed in two aspects offered only by the computer-based EDT, namely, the provision of feedback concerning the classification of variables into independent, dependent and controlled variables, and in maintaining values for controlled variables when planning experimental trials while requesting handling the independent variable at hand. Both of these affordances aimed at further problematizing the design of fair experiments for the students of the experimental condition. The subject domain of buoyancy was selected for our participants to study, because it is a topic with which students usually face problems when experimenting, including when designing a fair experiment (Hsin and Wu 2011; Marschner et al. 2012; Meindertsma et al. 2014).

This paper is organized into four parts. In the first part we discuss the value of experimentation in science education and report some of the problems that students face when experimenting, especially when designing a fair experiment. Right after, we report on the support mechanisms (e.g., scaffolds, heuristics) found in the literature for supporting students to overcome these problems, particularly when working with virtual labs in CoSIL environments. In so doing, we want to provide some background on what has been done so far to support the design of an experiment process when using virtual labs in CoSIL environments. We also situate this latter discussion within the buoyancy context to provide the reader with the background information on what prior researchers have found

concerning the effect of CoSIL guidance tools on student's learning of buoyancy concepts. Furthermore, we discuss the difficulty in designing effective scaffolding software due to the need to reconcile two rather contradictory requirements, namely, *structuring* learning tasks (e.g., decreasing the complexity of a learning task by breaking it down to smaller, simpler, more manageable sub-tasks) versus *problematizing* learning for students (e.g., show to the students problematic aspects of their learning and ask them to handle them). The idea behind this discussion was also to set the basis for discussing our study's results later on, given that our study's conditions differ in terms of the level of problematizing the design of fair experiments for students (structuring was the same for both conditions). In the second part of this article, the methodology followed is presented. In the third part, information on the results of the study is detailed. Finally, in the fourth part, the findings of the study are discussed.

Theoretical background

Several researchers have argued that experimentation is an indispensable part of science education for promoting student knowledge and skills, while it also presents pedagogical challenges best suited for student-centered learning approaches (e.g., Kremer et al. 2014; Minner et al. 2010). Experimentation can help students confront their prior knowledge with evidence derived from activities planned and undertaken by students themselves. Such a confrontation is highly valued in learning by inquiry, because it facilitates students' ability to self-regulate their own learning (de Jong 2006). Indeed, designing an experiment might be seen as the organizing principle of an entire inquiry cycle, namely, a complete sequence of phases for concluding an inquiry in science education (Pedaste et al. 2015). In this regard, the first set of activities involves developing an experimental design by identifying the variables involved in an experiment, and formulating hypotheses. Running the experiment, as well as gathering and interpreting data, follows after experimental design. When focusing on experimental design alone, a number of steps are necessary to design a fair experiment. Due to the complexity of this serial processing of tasks and the high probability of making an error along this linear chain of learning activities, this part of the inquiry cycle has been characterized as the most demanding phase (e.g., Lin and Lehman 1999). Designing fair experiments has proven to be quite difficult for students (Arnold et al. 2014; de Jong 2006; Furtak 2006; Kirschner et al. 2006), and considerable effort has been undertaken to develop and assess computer-based tools and scaffolds for guiding experimental design (Zacharia et al. 2015).

A very useful strategy in learning how to conduct fair experiments is to vary only one variable at a time in successive experimental trials (i.e., the "vary one thing at a time" or "VOTAT" heuristic; see, for instance, Glaser et al. 1992; Loucks-Horsley and Olson 2000; Tsirgi 1980; Veermans et al. 2006). This is strongly related to the ability to control variables (i.e., "Control of Variables Strategy" or "CVS", see, for instance, Klahr and Nigam 2004; Lin and Lehman 1999). Both the VOTAT heuristic and CVS are crucial for inferring causal relationships between variables and for planning unconfounded experiments (Arnold et al. 2014). This is because varying one variable at a time and keeping all other variables constant allows the experimental outcome for the dependent variable to be attributed to the variable that was varied. Obviously at least two successive experimental trials will be needed for the VOTAT heuristic to yield a conclusion (Marschner et al. 2012).

A number of researchers have recognized the need to provide support to learners through a tool for designing an experiment, including designing an experiment while using

virtual labs in CoSIL environments. The aforementioned guidance tools used for supporting the process of designing an experiment, namely the VOTAT and CVS heuristics, have also been implemented in CoSIL environments (Zacharia et al. 2015). A predominant purpose of software scaffolding has been to structure tasks (de Jong et al. 2012; Simons and Klein 2007). This guidance often goes together with re-focusing learner attention on more demanding or productive aspects of learning activities (Reiser 2004) and is highly compatible with a typical definition of scaffolding, where the objective is to assist learners in moving towards higher levels of cognitive processes or inquiry skills, which would not be accessible in the absence of that guidance (Saye and Brush 2002). A recent review in the field of science education has revealed that the phase in which students design their experiments and perform their investigations has received the most support from software scaffolds, including the use of the VOTAT heuristic (Zacharia et al. 2015). However, the results for the effect of this heuristic on student performance have been mixed (Zacharia et al. 2015). For example, providing prompts for experimentation based on the VOTAT heuristic was reported to favor reasoning skills (Chang et al. 2008). On the other hand, embedding the VOTAT heuristic in another computer-based tool did not have any noteworthy effect on student learning (Marschner et al. 2012).

The indeterminate findings reported for software scaffolds equipped with the VOTAT heuristic have also been seen in the context of buoyancy, which is the topic addressed by our participants in this study. Marschner et al. (2012) reported that feedback and prompts promoting the VOTAT heuristic in a computer-supported learning environment did not yield gains in knowledge and skill. According to these authors, the enactment of feedback and prompts might have interrupted the learning activity sequence, which could have further reoriented student attention away from the domain; all of these consequences might have backfired by having an undesirable effect on student knowledge and skills. Conversely, non-technological instructional support based on the VOTAT heuristic has been effective in promoting understanding of sinking and floating among preschool and primary school children (Hardy et al. 2006; Havu-Nuutinen 2005; Hsin and Wu 2011; Rappolt-Schlichtmann et al. 2007).

With regard to sinking and floating, the primary difficulty to overcome across age cohorts and educational levels has been learners' spontaneous heuristic of focusing on a single property. Most often, students concentrate on an object's weight to predict whether this object would sink or float in a fluid (Hsin and Wu 2011; Loverude et al. 2003; Meindertsma et al. 2014). The VOTAT heuristic has been promoted to address this learning difficulty in primary school (Heron et al. 2003), since it is expected to scaffold learners to consider multiple properties. Learners need to keep either mass or volume constant to investigate the sinking or floating of different objects in a given fluid. Further, learners need to keep either object density or fluid density constant to study the outcome of immersing objects of different density in fluids of different density.

Structuring learning tasks versus "problematizing" student inquiry for students in CoSIL environments

A substantial part of the difficulty in designing effective scaffolding software could arise from the need to reconcile two apparently contradictory requirements, namely, structuring learning tasks versus problematizing inquiry for students.

Structuring learning tasks aims at decreasing the complexity of a learning task by breaking it down to smaller, simpler, more manageable (sub-)tasks and/or by re-focusing learners' attention on to the important aspects of the task at hand (Quintana et al. 2004). Of

course, there is always the threat of oversimplifying a task for students, which could lead to unproductive learning paths (De Boer et al. 2014). In this case, computer-based scaffolding should provide structure to a level that still allows the learner to follow a productive path. A classic example of providing structure is to partition complex tasks and have them be processed serially (Clarke et al. 2005; Kalyuga 2007; Pollock et al. 2002). With regard to experimental design, this would mean that a tool might segregate the task in subsequent steps. Specifically, it might first guide students to outline different categories of variables (i.e., dependent variables, independent variables, controlled variables), then assign values to independent and controlled variables, and finally, plan all necessary experimental trials to carry out an experiment (van Joolingen et al. 2011).

Problematizing student inquiry requires from teachers or the CoSIL environment itself to reveal to the students the problematic aspects of their inquiry enactment. Learning occurs when the students consider and treat these identified problematic aspects as problems to be solved (e.g., turn an initial (unfair) design of an experiment into a design of a fair experiment). The level of difficulty of these problems is affected by the level of structuring and the level of guidance provided to the students. The more structuring and guidance provided, the less problematizing instances will occur. Hence, problematizing inquiry learning for students requires handling the structuring of the learning task in a way that it allows the student to be challenged (Reiser 2004, Molenaar et al. 2010). With regards to designing an experiment, which is at the center of this study, an instance of problematizing this process for students is to leave students on their own to figure out how to categorize the variables (dependent, independent and controlled) involved, after the CoSIL environment has pointed to the students that they have misclassified certain variables or have not classified some of them at all.

The tricky part of problematizing inquiry learning for students is to identify the problematic aspects of a student's learning and redirect them back to her/him for further consideration. In the case of CoSIL environments, such a process is much easier than any other traditional method, as long as the CoSIL used could analyse and evaluate students' progress and draw immediately students' attention to any resulted problematic aspects (Reiser 2004). Such a feedback is vital, since crucial aspects of the student inquiry might remain unattended by students, if the software scaffolds of the CoSIL environment are not designed to provide proper feedback accordingly (e.g., Zacharia et al. 2015; Reiser 2004). Providing prompts or timely feedback to students by software scaffolds and linking prompts or feedback to student behaviour seems to be the preferred strategy for problematizing inquiry learning for students (Zacharia et al. 2015).

A delicate balance between structuring and problematizing would mean that learners are supported in undertaking a demanding activity sequence and at the same time remain actively engaged and be challenged in the learning process. Whilst structuring scaffolds would directly support students and provide regulation to maintain them along productive learning trajectories, problematizing scaffolds would be less directive and they would aim at prompting students to consider alternative options (De Backer et al. 2016; Molenaar et al. 2011; Molenaar et al. 2014). Hence, reaching to an optimum balance between structuring and problematizing is a challenge. The literature is really poor on this particular topic. No frameworks exist to portray how structuring and problematizing could balance out. This is an aspect that we also wanted to examine in this study.

This study

In this study, we aimed at implementing a well-structured treatment, by breaking the design of an experiment process into smaller (more easily manageable) sub-processes, and examine whether further problematizing these sub-processes for students, through the provision of targeted feedback from the computer-based EDT, would have a different impact on their learning (i.e., cognitive processes and inquiry skills) from students who did not receive such feedback. In so doing, we compared two conditions (computer-based EDT versus paper-and-pencil worksheet) that involved the same structuring, in term of the design an experiment process, but differed in terms of the level of problematizing the design of the experiment process for students.

Specifically, the present study aimed at assessing the Go-Lab EDT, a computer-based scaffold for supporting students in designing experiments in the Go-Lab platform (for details see Sect. 2.2.2), by comparing two conditions, one involving the use of the Go-Lab EDT (experimental condition) and another that involved the use of a paper-and-pencil worksheet (control condition). The paper-and-pencil worksheet was used to scaffold the control group's participants when designing a fair experiment. The Go-Lab EDT was constructed to offer structure when students of the experimental condition were designing an experiment, as well as to "problematize" the task for students, when necessary. We integrated this tool into a computer-supported inquiry learning environment on relative density and monitored its influence on the cognitive processes (i.e., "to remember", "to understand", "to apply" and "to think critically and creatively") and certain inquiry skills (i.e., "identifying variables", "stating hypotheses", "operationally defining", and "designing investigations") of primary school students. Specifically, through the use of a quasi-experimental approach, we aimed at answering the following questions:

(1) Does the use of the Go-Lab EDT impact primary school students' cognitive processes and inquiry skills in a different manner than the use of a paper-and-pencil worksheet, when designing a fair experiment?
(2) Are there any interactions between students' cognitive processes and inquiry skills when designing a fair experiment through the use of the Go-Lab EDT or through the use of a paper-and-pencil worksheet? If yes, how do these interactions compare between the two conditions?

Methods

Participants

The sample included 26 fifth graders (10–11 years old) from two classes in a public elementary school in Limassol (Cyprus), which was randomly selected (i.e., all schools were first numbered and then one school was chosen by means of a random number generator). The two classes were randomly assigned to a condition. One class (5 boys, 9 girls) served as the experimental condition (i.e., use of the Go-Lab EDT), while the other class (7 boys, 5 girls) served as the control condition (i.e., use of a paper-and-pencil worksheet). Both classes included students of mixed ability (i.e., a class with students of varying competences, skills and knowledge levels in science). All students had basic computer skills and were able to carry out learning activities in the computer-based

AECT

learning environment. Additionally, all students had already been taught about sinking and floating in the previous school year, which had involved putting objects of different densities in containers of water. The learning environment in this study addressed an advanced buoyancy context including sinking and floating, for objects of different densities in fluids of different densities.

Materials

The learning environment

An Inquiry Learning Space (ILS) was developed by means of the authoring tool of the Go-Lab Project (de Jong et al. 2014). The ILS addressed the context of relative density (Inquiry Learning Space on Relative Density 2015) and presented a sequence of learning activities arranged in different phases of an inquiry learning cycle (Pedaste et al. 2015). In the first phase (orientation phase), students were introduced to the topic under study through the use of a driving question and then explored the basic terminology of the subject domain. A learning activity followed that required from students to make use of their prior knowledge, and then, they watched a video. Next, they outlined variables they were going to use and formulated their hypotheses (conceptualization phase). The upcoming phase involved experimentation in a virtual laboratory (investigation phase). This laboratory is called "Splash Lab" (Fig. 1) and it was developed in the Go-lab platform (Go-Lab – Learning by Experience 2015) to study sinking and floating and relative density, as well as buoyancy and the Archimedes principle (Splash: Virtual Buoyancy Laboratory 2015). Students had the opportunity to investigate objects put in containers with fluids. They were able to set values for the mass and volume of objects, as well as for the density of the fluid through the use of sliders (see Fig. 1). Then they ran the experiment and observed whether the object sank, or floated. The results of their experiments were presented in a table on the right side of the Splash Lab interface. Before students performed their experiments, they had to design them. For this purpose, students in the experimental group used the Go-Lab EDT, which was the focus of the present study (see Sect. 2.2.2). Students in the control
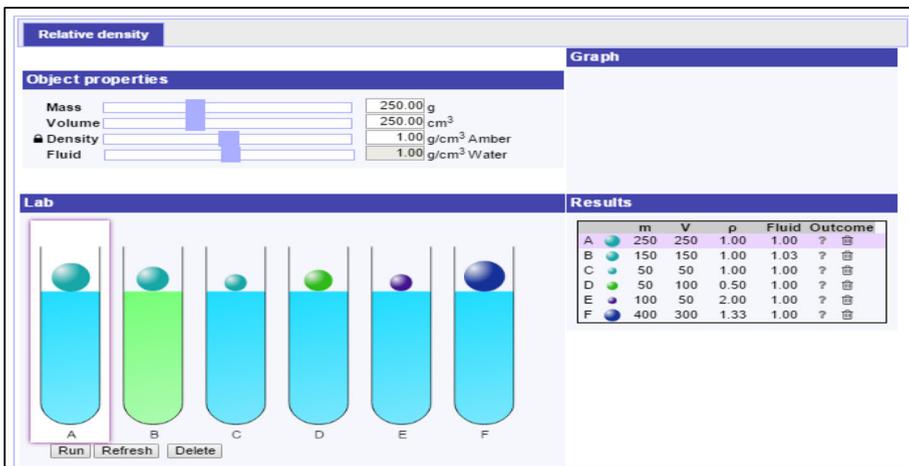

Fig. 1 Splash Lab interface (http://www.golabz.eu/lab/splash-virtual-buoyancy-laboratory)

group used a worksheet developed for the same purpose (see Sect. 2.2.3). After having gathered their data, students evaluated their hypotheses and reached a conclusion (conclusion phase). During the last phase of the inquiry cycle (discussion phase), students reflected upon the learning activity sequence in whole-class discussions guided by the teacher.

The Go-Lab Experiment Design Tool

The Go-Lab EDT was used only by students in the experimental group (Experiment Design Tool 2015). It was constructed in order to structure the task of experimental design as well as problematize the task for students, when necessary.[1] Specifically, the tool structured the process of experiment design as a three-step serial sub-processes: Students first had to identify the independent, controlled, and dependent variables; then they needed to assign values to their variables; and, finally, they had to set up their experimental trials. The tool included a set of predefined properties and measures, which were listed on the left side of its interface (Fig. 2). The provision of these predefined variables was another structuring feature of the tool.

On top of structuring the process of designing experiments, the Go-Lab EDT aimed at problematizing the design of an experiment process and sub-processes for students. For instance, any mistakes made by the students, at any of the aforementioned sub-processes of the design of an experiment process, were spotted by the EDT and were redirected back to the students for fixing them. The problematizing affordance of the Go-Lab EDT was offered through two specific functionalities. The first one appeared if students had not classified or had misclassified any variable. In those cases, feedback was presented to students to highlight the gap or misclassification. This setup aimed at further problematizing the design of experiment for students, since an invalid classification would interrupt the serial sequence of processing learning tasks and initiate remedial action taken up by the student in order to re-address classification of variables. Indeed, such a remedial action might be traced back to the considerations formulated by Wood et al. (1976) in their ground-breaking contribution to scaffolding. The feedback provided by the tool upon an invalid classification would initiate a regressive move backwards to tasks already encountered but not yet settled. Once again, this problematizing functionality would increase complexity locally to re-focus learner attention and re-orient it towards previous steps in the inquiry process (Reiser 2004).

The second functionality was linked to assigning values for variables and setting up experimental trials. The tool itself maintained the value of the controlled variable as constant across experimental trials and students had to set values for the independent variable. This latter functionality of the tool aimed at facilitating the VOTAT heuristic (e.g., keeping the value of the controlled variable constant through at least two successive experimental trials). In this case structuring was intertwined with problematizing the design of an experiment for students. On the one hand, maintaining the value of the controlled variable might have simplified the task by adding structure. On the other hand, students would still need to assign values for the independent variable, which was "problematizing" the learning route for them by directing their attention towards "a

---

[1] Previous research has largely conceptualized structuring and problematizing strategies of software scaffolds as distinct, namely, one could have the one or the other type (e.g., Kukkonen et al. 2016; Molenaar et al. 2011). Our design incorporated structuring and problematizing strategies in one and the same software scaffold.
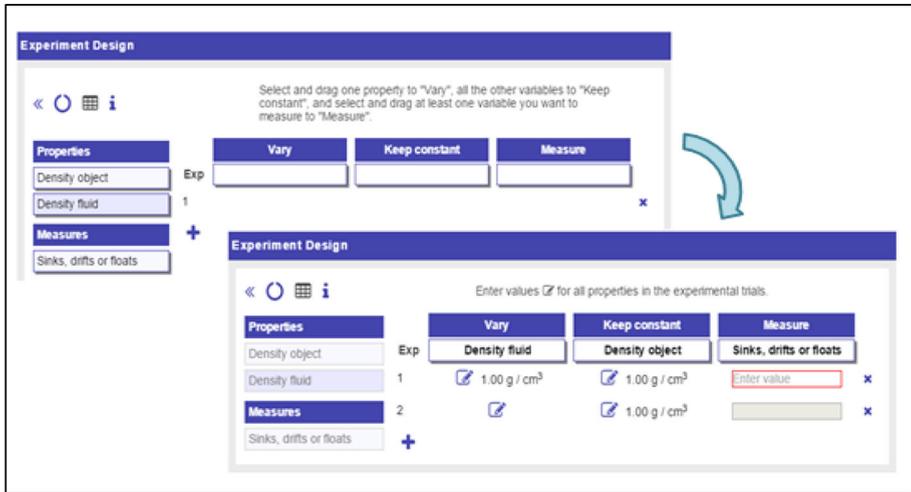
Fig. 2 The Experiment Design Tool (http://www.golabz.eu/apps/experiment-design-tool)

situation that needed resolution" (e.g., what values should be assigned to the independent variable?, How many values are needed?) (Reiser 2004, p. 287).

The paper-and-pencil worksheet

Students in the control condition used a paper-and-pencil worksheet which was designed to scaffold the process of designing an experiment (Appendix 1). Students were instructed to assign the given variables to three different columns of a table in the worksheet, which corresponded to independent, controlled, and dependent variables, respectively. Students were also instructed to record next to each variable its value for each of their experimental trials (for which they used Splash Lab, as did the experimental group). Finally, students were reminded that they might need multiple experimental trials to complete their experiment and they were urged to add more rows to the table if necessary. As mentioned above, the paper-and-pencil worksheet differed from the Go-Lab EDT in that it did not offer any feedback to the learners during the design of an experiment. As a result, the two conditions differed in that the experimental condition problematized more the design of an experiment process for students than the control condition did.

Procedure

Before and after the educational intervention, all students completed a cognitive processes test and an inquiry skills test (see Sect. 2.4.1). Students had 20 min to complete each test. Students in both conditions used a computer to go through the Inquiry Learning Space (ILS) and needed about 80 min to complete the learning activity sequence. This was done in the school's computer laboratory. With regard to the investigation phase, which is where the procedure of designing an experiment had been incorporated, students in the experimental and control groups had the same time to go through the learning activities. In fact, it was observed that this time sufficed for all students to complete all planned tasks (i.e., about 30 min dedicated to the investigation phase). During the learning activities, the

teacher provided only technical assistance to students, for instance, about entering or exiting the learning environment, when this was necessary. Students in both conditions were taught by the same teacher, who had received training on using the Go-lab platform and who followed the same protocol in both the experimental and control group. Overall, the teacher intervention was kept to the minimum for both conditions.

Data collection

Tests for cognitive processes and inquiry skills

We used two tests to examine students' cognitive processes and inquiry skills, respectively. Both instruments were administered to students before and after the educational intervention. The cognitive processes test has been based on Bloom's (1956) taxonomy of educational objectives as revised by Anderson and Krathwohl (2001) and as further adjusted by de Jong (2014) and Zervas (2013). Specifically, the test was structured according to four cognitive processes arranged along a gradient towards higher-order thinking (Table 1). The most basic process referred to students' ability to recognize or recall information encountered during a course, such as material, terminology, or procedures ("to remember"). At the next level, students were able to organize and arrange information so that it would become intelligible to them ("to understand"). This level addressed understanding of concepts and definitions. When students were able to apply information to reach an answer, they had moved on to the next process ("to apply"). In that case, students were capable of applying acquired knowledge to work through a new task, which was framed within a new learning context. The highest level in the gradient of cognitive processes corresponded to adapting acquired knowledge to address a novel context ("to think critically and creatively"). This involved screening background knowledge and information to select aspects relevant to the novel context as well as being able to combine these aspects and produce original reasoning. Overall, the first two processes, namely, "to remember" and "to understand", pertained to working within a single context, while the latter two, "to apply" and "to think critically and creatively", signified a partial or full ability of transfer across learning contexts, respectively.

We included one item for each process from the taxonomy in the instrument, so the cognitive processes test had four items, overall (Appendix 2). Two of the items (1 and 4) were open-ended and two of them were close-ended (2 and 3). All items referred to the subject domain of the study, namely, relative density. The test was evaluated prior to the

Table 1 Cognitive processes

Process	Description
To remember (Item 1 in Appendix 2)	To help the learner recognize or recall information
To understand (Item 2 in Appendix 2)	To help the learner organize and arrange information mentally
To apply (Item 3 in Appendix 2)	To help the learner apply information to reach an answer
To think critically and creatively (Item 4 in Appendix 2)	To help the learner think about causes, predict, make judgments and create new ideas

Processes are aligned along a gradient ranging from the most basic (i.e., "To remember") to higher-order processes (i.e., "To think critically and creatively"). This is a revised version of the Anderson and Krathwohl (2001) framework [for details see de Jong (2014) and Zervas (2013)]

educational intervention for content validity and face validity. A four-member expert panel (one academic, one post-doctoral fellow and two in-service primary school teachers) provided feedback concerning the content of the test. Ten fifth graders who did not take part in the study also provided feedback in terms of wording of items and presentation of tasks. After analysing the feedback and making all necessary adjustments, the final version of the cognitive processes test was ready for use. For the post-test, the order of items was changed.

For assessing inquiry skills, we selected 12 out of the 36 multiple choice items of the TIPSII test that was developed by Burns et al. (1985) (Appendix 3). The instrument focused on four different skills, namely, "identifying variables" (items 30, 31, 32, and 36 in TIPSII), "stating hypotheses" (items 6, 27, and 35 in TIPSII), "operationally defining" (items 2, 23, and 26 in TIPSII), and "designing investigations" (items 10 and 21 in TIPSII). If we order these four skills in the typical timeline of an experimental design procedure (e.g., Hofstein et al. 2005; Kremer et al. 2014; Minner et al. 2010), we can distinguish between an "entry" skill, intermediate skills and an "exit" skill. Specifically, "identifying variables" precedes as an "entry" skill. Next, variables are interrelated through hypotheses ("stating hypotheses") and then values are assigned to variables so that measurements can be made ("operationally defining"). Finally, the "exit" skill is to define the different experimental trials needed to complete the experimentation, addressed by "designing investigations". The two teachers, the two postgraduate students and the ten fifth graders who were requested to provide feedback for the cognitive process test were also asked to do the same for the inquiry skills test. After receiving teacher and student feedback, we prepared the final version of the inquiry skills test. The order of items was changed in the post-test.

Data analyses

Coding of items in the instruments

For the cognitive processes test, student responses for the open-ended items 1 and 4 were coded by two independent coders to indicate scientifically acceptable or unacceptable reasoning. For determining the scientific accuracy of students' reasoning, a rubric table was used for each item, which indicated the possible scientifically accurate parts needed to comprise a complete and scientifically correct reasoning. For example, for item 4, which involved a challenge of selecting a type of wood to construct a wooden raft so that it would float in water (see Appendix 2), we coded as scientifically valid reasoning the one that acknowledged different densities for different types of wood, and discussed sinking or floating of the raft to be constructed with reference to this varying density and the density of the water. Each one of the required parts of the reasoning, as specified in the evaluation rubric, received a point. Right after, student performance for items 1 and 4 was adjusted to range between 0 and 1. The idea behind this adjustment was to get a sense of how students' performance on an item compares to another item. Inter-rater reliability for a subset of 20% of all student responses for those two items was acceptable (Cohen's Kappa = 0.92), while mismatches between coders were settled through discussion. For the close-ended items 2 and 3, one point was awarded for each correct reply. However, students' performance for these two items was also adjusted to range between 0 and 1 (same scale as in the case of items 1 and 4). In addition, a composite score for the whole cognitive processes test was also calculated and adjusted to range between 0 and 1.

For the inquiry skills test, each student received one point for each correct response to each of the inquiry test's close-ended items. Students' performance for each skill was later weighed to range between 0 and 1 (same scale as the one used for the items of the cognitive processes test). Finally, a composite score for the entire skills test was also calculated and adjusted to range between 0 and 1.

Overall, there were four cognitive processes ("to remember"; "to understand"; "to apply"; "to think critically and creatively"), with their composite score, and four inquiry skills ("identifying variables"; "stating hypotheses"; "operationally defining"; "designing investigations"), with their composite score, all ranging between 0 and 1.

Reliability and validity of instruments

Both tests revealed acceptable reliability indices. A split-half reliability analysis showed that *Cronbach's alpha* was 0.58 and 0.65 for the cognitive processes test before and after the educational intervention, respectively. *Cronbach's alpha* was 0.59 and 0.61 for the inquiry skills instrument at pre- and post-test, respectively. Further, we calculated *Spearman's rank* correlation indices for items within dimensions as well as *Kendall's tau b* correlation indices among dimension scores for both instruments. Items within dimensions had significant Spearman's rank correlation indices ($p < 0.05$), while there were no significant *Kendall's tau b* correlations among dimension scores ($p > 0.05$).

Nonparametric statistics

To examine trends in the data, we employed nonparametric tests (i.e., *Mann–Whitney* and *Wilcoxon signed ranks* tests) and nonparametric correlations (*Kendall's tau b* correlations). This option was taken since our data were non-normally distributed. For all analyses we used IBM SPSS Statistics 21.

Results

Does the use of the Go-Lab EDT impact primary school students' cognitive processes and inquiry skills in a different manner than the use of a paper-and-pencil worksheet, when designing a fair experiment? (research question 1)

We performed *Mann–Whitney* tests to determine significant differences between the experimental and control conditions in cognitive processes and inquiry skills before and after the educational intervention. We performed *Wilcoxon signed ranks* tests to account for significant trends within each group by contrasting their pre-test and post-test performance. All tests were computed for composite scores as well as for separate cognitive processes and inquiry skills.

Students' overall performance for cognitive processes and inquiry skills did not differ significantly between conditions on the pre-test (*Mann–Whitney* tests reported for pre-tests in Tables 2 and 3, respectively). Additional *Mann–Whitney* tests computed for each cognitive process and skill separately also revealed no significant differences between the conditions on the pre-test. After the intervention, there was a marked improvement for both conditions in students' overall performance for cognitive processes (*Wilcoxon signed ranks* test *Zs* in Table 2). Overall, these analyses showed that all cognitive processes improved

Table 2 Students' overall performance on the cognitive processes test

	Experimental group	Control group	*Mann–Whitney* test Z
Pre-test	0.28	0.20	− 1.04 ns
Post-test	0.71	0.69	− 0.18 ns
Wilcoxon signed ranks test Z	− 3.31**	− 3.06**	

ns non-significant

** $p < 0.01$

Table 3 Students' overall performance on the inquiry skills test

	Experimental group	Control group	*Mann–Whitney* test Z
Pre-test	0.33	0.26	− 1.94 ns
Post-test	0.68	0.44	− 2.75**
Wilcoxon signed ranks test Z	− 3.31**	− 2.03*	

ns non-significant

* $p < 0.05$; ** $p < 0.01$

significantly for both conditions ($p < 0.05$ for all *Wilcoxon signed ranks* test Zs). There was no significant difference between conditions on the post-test for either overall performance or any single cognitive process ($p > 0.05$ for all *Mann–Whitney* test Zs).

Students' overall performance also improved for inquiry skills in both the experimental (Table 3, *Wilcoxon signed ranks* test $Z = - 3.31$; $p < 0.01$) and control conditions (Table 3, *Wilcoxon signed ranks* test $Z = - 2.03$; $p < 0.05$). In this case, however, conditions differed significantly on the post-test, with the experimental group performing better than the control group (Table 3, *Mann–Whitney* test $Z = - 2.75$; $p < 0.01$).

According to these analyses, it was found that the experimental group's scores increased considerably across all inquiry skills on the post-test. The control group performed better at "stating hypotheses" and "operationally defining" but not at "identifying variables" and "designing investigations". These developments were reflected in significant differences between conditions for "identifying variables" (*Mann–Whitney* test $Z = - 3.77$; $p < 0.001$) and "designing investigations" (*Mann–Whitney* test $Z = - 2.91$; $p < 0.01$) after the educational intervention.

Are there any interactions between students' cognitive processes and inquiry skills when designing a fair experiment through the use of the Go-Lab EDT or through the use of a paper-and-pencil worksheet? If yes, how do these interactions compare between the two conditions? (research question 2)

We subtracted pre-test from post-test scores for each cognitive process and inquiry skill and calculated *Kendall's tau b* correlations among these differences. This would indicate dimensions that tended to correlate during the intervention. A significant positive correlation between a cognitive process and an inquiry skill would denote that both dimensions tended to increase as a result of the experimental or control treatment.

Table 4 *Kendalls tau b* correlations computed for differences between pre- and post-test scores for cognitive processes and inquiry skills in the experimental group

Cognitive processes	Inquiry skills			
	Identifying variables	Stating hypotheses	Operationally defining	Designing investigations
To remember	ns	ns	ns	ns
To understand	ns	ns	ns	ns
To apply	0.69**	ns	ns	ns
To think critically and creatively	0.64*	ns	ns	ns

ns non-significant

* $p < 0.05$; ** $p < 0.01$

Table 5 *Kendalls tau b* correlations computed for differences between pre- and post-test scores for cognitive processes and inquiry skills in the control group

Cognitive processes	Inquiry skills			
	Identifying variables	Stating hypotheses	Operationally defining	Designing investigations
To remember	ns	ns	ns	ns
To understand	ns	0.58*	0.66*	ns
To apply	ns	ns	ns	ns
To think critically and creatively	ns	ns	ns	ns

ns non-significant

* $p < 0.05$

Table 4 shows that in the experimental condition, "identifying variables" correlated with two cognitive processes, namely "to apply" (*Kendalls tau b* = 0.69; $p < 0.01$) and "to think critically and creatively" (*Kendalls tau b* = 0.64; $p < 0.05$). For the control condition, correlations revealed a different pattern (Table 5). The two inquiry skills that improved for the group after the intervention, namely "stating hypotheses" and "operationally defining", correlated with students' ability to organize and arrange information mentally (i.e., "to understand"; *Kendalls tau b* = 0.58; $p < 0.05$, and *Kendalls tau b* = 0.66; $p < 0.05$, respectively).

Discussion

The primary finding of the present study was that the Go-Lab EDT facilitated development of "entry" and "exit" inquiry skills for the experimental group, which was not the case for the students' of the control group, who used a paper-and-pencil worksheet designed to scaffold their design process of their experiment. Namely, students in the experimental condition who used the tool improved their inquiry skills related to "identifying variables"

AECT

("entry" skill) and "designing investigations" ("exit" skill). These effects could be linked to the feedback offered by the Go-Lab EDT, as well as to the increased level of problematizing the design of an experiment process and sub-processes for the students of the experimental group, which was lacking in the paper-and-pencil worksheet that was given to students in the control group. For instance, feedback provided to the students of the experimental group for missing or wrong classification of variables, along with the automatic importation of stable values for controlled variables, could have positively catalysed the inquiry skill of "identifying variables". Further, the feedback on maintaining the values for controlled variables constant across the experimental trials might have shown to the learner the importance of controlling certain variables in an experimental design in order for a fair experiment to be planned and executed. Such an anchoring might have also worked towards enhancing the skill of "designing investigations" for the experimental condition. Needless to say, looking into the feedback provided by EDT tools, such as the one of this study, should be a priority for future research. As conjectured right above, it appears that the feedback provided by an EDT tool positively affects students' learning and inquiry skills. Of course, we are missing all sort of information concerning this relationship, such as, identifying the type, quantity and quality of the feedback needed for optimizing students' learning and inquiry skills. In addition, predicting when the students would be in need for a particular type of feedback, as well as synchronizing the provision of feedback with the emerging needs of students, are top priorities for this research domain.

The automatic importation of stable values for controlled variables might have further acted as a prompt that could have helped students in the experimental condition to deal with production deficiency in their experimental designs (see, for instance, Veenman et al. 2006, who have noted that production deficiency may be encountered by students who may fail to use their metacognition due to task difficulty). This would have been an insightful example of hitting two birds with one stone; namely, re-allocating student attention from an aspect of the task that was being supported, namely keeping controlled variables constant, to assigning values for independent variables. Reiser (2004) has underlined that software scaffolds may take over some aspects of learning tasks and re-focus student attention to more productive or demanding parts of the same learning tasks.

An additional aspect of the Go-Lab EDT, which might explain its effectiveness in supporting development of the entry skill of "identifying variables", could have been its provision of timely assessment of gaps and misclassifications of variables (see in this regard Kalyuga 2007). Such a functionality might provide a rather simple act of formative assessment taken over by the tool itself, but its direct and indirect effects might be more valuable for configuring a learner-tailored environment, even at such a confined scale. Because tool configurations that account for "problematizing" the task for students have not been frequently investigated by previous research (Reiser 2004), this rather uncomplicated software tool provision might prove quite promising for analogous endeavours in the future. Future configurations of the Go-Lab EDT might also support additional and more sophisticated functionalities of "problematizing" the task for students, for instance, providing feedback for assigning values to independent variables so that causal inferences relating independent and dependent variables can be readily determined. This potential link between "problematizing" the task for students and providing formative assessment should be more thoroughly examined in designing software scaffolds. One fruitful direction would be that software tools might take over aspects of formative assessment, which might enable the teacher to concentrate on other instructional issues that are not easily monitored by software tools. For instance, identification of variables may be largely taken over by the software scaffold, while the teacher may concentrate on more demanding aspects of the

experimental design, such as planning experimental trials or configuring the values for the independent variables in those experimental trials. Given the usual time constraints in delivering effective formative assessment, which includes a timely diagnosis of learner performance and timely provision of teacher feedback to students, the division of labour mentioned right above may prove crucial.

Although there was no significant difference between the two conditions in cognitive processes, we observed different patterns in how cognitive processes and inquiry skills correlated during the study's treatments. In the experimental group, these interactions involved the inquiry skill of "identifying variables", which was markedly enhanced by the intervention, and which correlated with cognitive processes that referred to partial or full inter-contextual transfer (i.e., "to apply" and "to think critically and creatively", respectively). This might imply that reinforcement of the "entry" skill in experimental design (i.e., "identifying variables") could have also triggered some inter-contextual transfer for students in the experimental condition. It was quite interesting that cognitive processes related to transfer (i.e., "to apply" and "to think critically and creatively") were linked to the "entry" skill (i.e., "identifying variables") and not to the "exit" skill (i.e., "designing investigations"). In the same direction, Veermans et al. (2006) have suggested that heuristics, such as the VOTAT heuristic, might amplify transfer. Additionally, there have been indications that feedback such as that provided by the Go Lab EDT might initiate self-regulation, i.e., an individual's competence to guide his/her own learning (Marschner et al. 2012). For instance, self-regulation, in an inquiry context such as the one of this study, relates to the competence of the student to be able to plan and implement an inquiry task, such as designing an experiment, on his/her own. In other words, it is expected that students, who used the Go Lab EDT in several inquiry cycles, will gradually stop using it because they would be in position to design an experiment on their own (without the use of an EDT scaffold). A methodological approach more focused on these effects might enable closer elaboration of the impact of software scaffolds.

Our study also attempted to shed some light on the learning process as it played out for the control condition, as well. In this case, there were different patterns of significant interactions between cognitive processes and inquiry skills as compared to those detected for the experimental condition. Students in the control group showed correlations of the two inquiry skills that improved for them (i.e., "stating hypotheses" and "operationally defining") with a cognitive process concentrating on a single context ("to understand"). It could be that differences in the correlations of cognitive processes and inquiry skills between the experimental and control conditions might reflect differences in prioritizing or allocating student effort and attention. In this vein, the Go-Lab EDT might have directed student concentration on "entry" and "exit" inquiry skills. On the other hand, the use of the paper-and-pencil worksheet might have triggered more active involvement of students in the control condition with aspects of learning that fostered skills with an "intermediate" position in the inquiry process, which might all have been confined within the learning context addressed. In any case, our findings imply that future research employing quasi-experimental approaches for assessing software scaffolds might benefit from detailed monitoring of cognitive processes and inquiry skills in both experimental and control conditions.

Moreover, we urge other researchers to continue the investigations on when and how to use scaffolds, which target different aspects of inquiry enactments through CoSIL environments. The idea is to reach to a solid framework on when and how to offer scaffolding to the students and when and how to fade it out. Currently, such a framework is missing (see, for a thorough review, Zacharia et al. 2015).

Limitations of the current study and further implications for future research

The exploratory character of the current study was based on a rather confined sample of students that has necessitated the use of non-parametric statistics due to deviations from normality in data distribution. This might have also increased the odds of Type-I Error. In addition, the reliability figures we have estimated for our instruments have been acceptable but rather low. Future research needs to recruit larger samples to address the effectiveness of the Go-Lab EDT.

Moreover, future research needs to address another limitation of the current study. We have only examined the effect of the tool on cognitive processes and inquiry skills of students in one cycle of inquiry, which restricted us from following any effects of "fading". The Go-lab EDT could be adequately configured to partially or fully remove the support offered to students (i.e., through "fading"). For instance, the set of variables given to students might decrease (partial removal of support), or be removed as a whole (full removal of support).[2] Such an arrangement would also influence the planning of experimental trials in the tool. A given and closed set of variables in the tool would restrict the range of possible experimental trials, while an open list of variables would necessitate an advanced complexity in experimental designs. Another domain for future research could be the study of the use of different virtual labs (i.e., other than the Splash Lab, which we have embedded in the learning environment of the current study) and how structuring and problematizing varies as the students move from designing an experiment in one lab to another. The latter could also be combined with a "fading" mechanism in order to identify if certain skills and practices are transferable from one context to another.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

---

[2] The option of "fading" would mean that the Go-Lab EDT cannot be conceived as a tool allowing for "working smart" within the frame of distributed intelligence (see the rationale introduced by Pea 2004, p. 443). However, certain functionalities of the Go-Lab EDT, namely the sequential arrangement in planning experimental designs, the table for classifying variables and the provision of the value for constant variables, might be considered as aspects of "working smart". The latter means that these functionalities could enhance the work of an experienced user of the tool without affecting his/her ability to design a valid experiment.

Appendix 1: Experimental design by the control group

Please complete the following table so that the given properties (density of the object; density of the fluid) and outcomes (sinks, drifts/hovers or floats) are placed in the proper column.

Please insert next to each variable the value used when running your experimental trial. Please also insert the result of your experimental trial under "Measure".

Please note that you might need multiple experimental trials to conclude your experiment. You can add more rows to your table if you need to do so.

Properties:
Density of the object
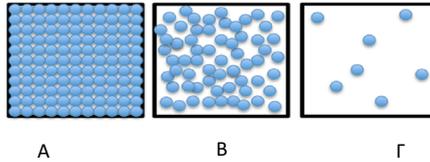Density of the fluid

Measures:
Sinks, drifts or floats

Table to be completed:

Vary	Keep constant	Measure

Name:………………………………………………….. Class:……

Appendix 2: Cognitive processes test

1. Three different quadrants are shown below. How do their contents compare? Please explain your reasoning.


A   B   Γ

2. Which words would you circle so that each sentence would be scientifically valid?
2a. When two objects of different material have the same/different volume, then the object with the greater/lesser mass will display a higher/lower density.
2b. When two objects of different material have the same/different mass, then the object with the greater/lesser volume will display a higher/lower density.

3. Densities of various solids and liquids are listed in the table below. Using the data depicted in the table, please indicate whether the following statements are right or wrong.

Solids	Density (g/cm^3)	Liquids	Density (g/cm^3)
Birch	0.63	Acetone	0.79
Paraffin	0.85	Olive oil	0.92
Ice	0.90	Water	1
Amber	1.05	Sea water	1.20
Ebony	1.22	Glycerine	1.26
Aluminium	2.70	Bleach	1.49

3a. Birch would float in bleach   Right   Wrong
3b. Paraffin would sink in glycerine   Right   Wrong
3c. Aluminium would float in olive oil   Right   Wrong
3d. Ice would sink in water   Right   Wrong
3e. Birch and paraffin would sink in acetone   Right   Wrong
3f. Paraffin and ice would sink in sea water   Right   Wrong
3g. Amber would float in sea water, glycerine, and bleach   Right   Wrong
3h. All solids in the table would float in bleach   Right   Wrong

4. Newton and Hypatia want to construct a wooden raft that will float in water. However, they disagree whether they can use any type of wood to construct this raft.

We can use any type of wood to construct our raft.

No, we cannot use any type of wood.
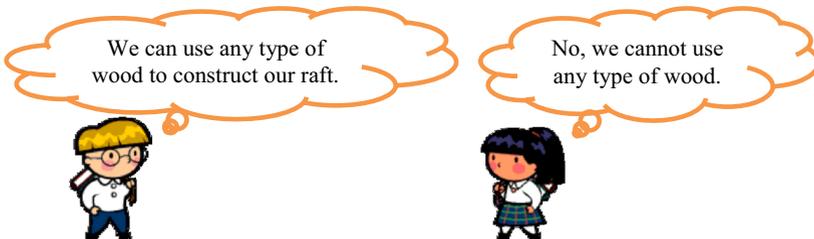
Who is right? Why?

Appendix 3: Inquiry skills test [for each item, its corresponding item number in the TIPSII test by Burns et al. (1985) is given in parentheses]

1. (2) A study of car efficiency is done. The hypothesis tested is that a gasoline additive will increase car efficiency. Five identical cars each receive the same amount of gasoline but with different amounts of Additive A. They travel the same track until they run out of gasoline.
   The research team records the number of kilometers each car travels. How is car efficiency measured in this study?

   A. The time each car runs out of gasoline.
   B. The distance each car travels.
   C. The amount of gasoline used.
   D. The amount of Additive A used.

2. (6) A police chief is concerned about reducing the speed of cars. He thinks several factors may affect automobile speed.
   Which of the following is a hypothesis he could test about how fast people drive?

   A. If the drivers are younger, then they are likely to drive faster.
   B. If the number of cars involved in an accident is larger, then it will be less likely people that are to get hurt.
   C. If more policemen are on patrol, then the number of car accidents will be fewer.
   D. If the cars are older, then they are likely to be in more accidents.

3. (10) Jim thinks that if there is more air pressure in a basketball, then it will bounce higher. To investigate this hypothesis he collects several basketballs and an air pump with a pressure gauge. How should Jim test his hypothesis?

   A. Bounce basketballs with different amounts of force from the same height.
   B. Bounce basketballs having different air pressures from the same height.
   C. Bounce basketballs having the same air pressure at different angles from the floor.
   D. Bounce basketballs having the same amount of air pressure from different heights.

4. (26) A biologist tests this hypothesis: the greater the amount of vitamins given to rats the faster they will grow. How can the biologist measure how fast rats will grow?

   A. Measure the speed of the rats.
   B. Measure the amount of exercise the rats receive.
   C. Weigh the rats every day.
   D. Weigh the amount of vitamins the rats will eat.

5. (21) A greenhouse manager wants to speed up the production of tomato plants to meet the demands of anxious gardeners. She plants tomato seeds in several trays. Her hypothesis is that the more moisture seeds receive the faster they sprout. How can she test this hypothesis?

   A. Count the number of days it takes seeds receiving different amounts of water to sprout.
   B. Measure the height of the tomato plants a day after each watering.
   C. Measure the amount of water used by plants in different trays.
   D. Count the number of tomato seeds placed in each of the trays.

6. (23) Lisa wants to measure the amount of heat energy a flame will produce in a certain amount of time. A burner will be used to heat a beaker containing a 1liter of cold water for ten minutes. How will Lisa measure the amount of heat energy produced by the flame?

   A. Note the change in water temperature after ten minutes.
   B. Measure the volume of water after ten minutes.
   C. Measure the temperature of the flame after ten minutes.
   D. Calculate the time it takes for the liter of water to boil.

7. (27) Some students are considering variables that might affect the time it takes for sugar to dissolve in water. They identify the temperature of the water, the amount of sugar and the amount of water as variables to consider. What is a hypothesis the students could test about the time it takes for sugar to dissolve in water?

   A. If the amount of sugar is larger, then more water is required to dissolve it.
   B. If the water is colder, then it has to be stirred faster to dissolve.
   C. If the water is warmer, then more sugar will dissolve.
   D. If the water is warmer, then it takes the sugar more time to dissolve.

A study was done to see if leaves added to soil had an effect on tomato production. Tomato plants were grown in four large tubs. Each tub had the same kind and amount of soil. One tub had 15 kg of rotted leaves mixed in the soil and a second had 10 kg. A third tub had 5 kg and the fourth had no leaves added. Each tub was kept in the sun and watered the same amount. The number of kilograms of tomatoes produced in each tub was recorded.

8. (30) What is a controlled variable in this study?

   A. Amount of tomatoes produced in each tub.
   B. Amount of leaves added to the tubs.
   C. Amount of soil in each tub.
   D. Number of tubs receiving rotted leaves.

9. (31) What is the dependent or responding variable?

   A. Amount of tomatoes produced in each tub.
   B. Amount of leaves added to the tubs.
   C. Amount of soil in each tub.
   D. Number of tubs receiving rotted leaves.

10. (32) What is the independent or manipulated variable?

   A. Amount of tomatoes produced in each tub.
   B. Amount of leaves added to the tubs.
   C. Amount of soil in each tub.
   D. Number of tubs receiving rotted leaves.

11. (35) Ann has an aquarium in which she keeps goldfish. She notices that the fish are very active sometimes but not at others. She wonders what affects the activity of the fish. What is a hypothesis she could test about factors that affect the activity of the fish?

   A. If you feed fish more, then the fish will become larger.
   B. If the fish are more active, then they will need more food.

C. If there is more oxygen in the water, then the fish will become larger.

D. If there is more light on the aquarium, then the fish will be more active.

12. (36) Mr. Bixby has an all-electric house and is concerned about his electricity bill. He decides to study factors that affect how much electrical energy he uses. Which variable might influence the amount of electrical energy used?

A. The amount of television the family watches.

B. The location of the electricity meter.

C. The number of baths taken by family members.

D. A and C.

References

Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Addison Wesley Longman.

Arnold, J. C., Kremer, K., & Mayer, J. (2014). Understanding students' experiments—What kind of support do they need in inquiry tasks? *International Journal of Science Education, 36,* 2719–2749.

Bloom, B. S. (1956). *Taxonomy of educational objectives. Handbook I: The cognitive domain*. New York: David McKay.

Burns, J., Okey, J., & Wise, K. (1985). Development of an integrated process skill test: TIPS II. *Journal of Research in Science Teaching, 22,* 169–177.

Chang, K. E., Chen, Y. L., Lin, H. Y., & Sung, Y. T. (2008). Effects of learning support in simulation-based physics learning. *Computers & Education, 51,* 1486–1498.

Clarke, T., Ayres, P., & Sweller, J. (2005). The impact of sequencing and prior knowledge on learning mathematics through spreadsheet applications. *Educational Technology Research and Development, 53,* 15–24.

De Backer, L., Van Keer, H., & Valcke, M. (2016). Eliciting reciprocal peer-tutoring groups' metacognitive regulation through structuring and problematizing scaffolds. *The Journal of Experimental Education, 84,* 804–828.

De Boer, G. E., Quellmalz, E. S., Davenport, J. L., Timms, M. J., Herrmann-Abell, C. F., Buckley, B. C., et al. (2014). Comparing three online testing modalities: Using static, active, and interactive online testing modalities to access middle school students' understanding of fundamental ideas and use of inquiry skills related to ecosystems. *Journal of Research in Science Teaching, 51,* 523–554.

de Jong, T. (2006). Computer simulations—Technological advances in inquiry learning. *Science, 312,* 532–533.

de Jong, T. (Ed.). (2014). *Preliminary inquiry classroom scenarios and guidelines*. D1.3. Go-Lab Project (Global Online Science Labs for Inquiry Learning at School). Retrieved from http://www.go-lab-project.eu/sites/default/files/files/deliverable/file/Go-Lab%20D1.3.pdf.

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

de Jong, T., Weinberger, A., van Joolingen, W. R., Ludvigsen, S., Ney, M., Girault, I., et al. (2012). Designing complex and open learning environments based on scenarios. *Educational Technology Research & Development, 60,* 883–901.

Experiment Design Tool. (2015). http://www.golabz.eu/apps/experiment-design-tool. Accessed 25 January 2018.

Furtak, E. M. (2006). The problem with answers: An exploration of guided scientific inquiry teaching. *Science Education, 90,* 453–466.

Glaser, R., Schauble, L., Raghavan, K., & Zeitz, C. (1992). Scientific reasoning across different domains. In E. de Corte, M. Linn, H. Mandl, & L. Verschaffel (Eds.), *Computer-based learning environments and problem solving* (pp. 345–373). Berlin: Springer.

Go-Lab – Learning by Experience. (2015). http://www.go-lab-project.eu/. Accessed 25 January 2018.

Go-Lab Sharing and Authoring Platform. (2015). http://www.golabz.eu/. Accessed 25 January 2018.

Hardy, I., Jonen, A., Möller, K., & Stern, E. (2006). Effects of instructional support within constructivist learning environments for elementary school students' understanding of ''floating and sinking''. *Journal of Educational Psychology, 98,* 307–326.

Havu-Nuutinen, S. (2005). Examining young children's conceptual change process in floating and sinking from a social constructivist perspective. *International Journal of Science Education, 27,* 259–279.

Heron, P. R. L., Loverude, M. E., Shaffer, P. S., & McDermott, L. C. (2003). Helping students develop an understanding of Archimedes' principle. II. Development of research-based instructional materials. *American Journal of Physics, 71,* 1188–1195.

Hofstein, A., & Lunetta, V. N. (2004). The laboratory in science education: Foundations for the twenty-first century. *Science Education, 88,* 28–54.

Hofstein, A., Navon, O., Kipnis, M., & Mamlok-Naaman, R. (2005). Developing students' ability to ask more and better questions resulting from inquiry-type chemistry laboratories. *Journal of Research in Science Teaching, 42,* 791–806.

Hsin, C.-T., & Wu, H.-K. (2011). Using scaffolding strategies to promote young children's scientific understandings of floating and sinking. *Journal of Science Education and Technology, 20,* 656–666.

Inquiry Learning Space on Relative Density. (2015). http://graasp.eu/ils/546b3398e9934012b7c65c65/?lang=el (in Greek). Accessed 25 January 2018.

Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review, 19,* 509–539.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41,* 75–86.

Klahr, D., & Nigam, M. (2004). The equivalence of learning paths in early science instruction: Effects of direct instruction and discovery learning. *Psychological Science, 15,* 661–667.

Kremer, K., Specht, C., Urhahne, D., & Mayer, J. (2014). The relationship in biology between the nature of science and scientific inquiry. *Journal of Biological Education, 48,* 1–8.

Kukkonen, J., Dillon, P., Kärkkäinen, S., Hartikainen-Ahia, A., & Keinonen, T. (2016). Pre-service teachers' experiences of scaffolded learning in science through a computer supported collaborative inquiry. *Education and Information Technologies, 21*(2), 349–371. https://doi.org/10.1007/s10639-014-9326-8.

Lin, X., & Lehman, J. D. (1999). Supporting learning of variable control in a computer-based biology environment: Effects of prompting college students to reflect on their own thinking. *Journal of Research in Science Teaching, 36,* 837–858.

Loucks-Horsley, S., & Olson, S. (Eds.). (2000). *Inquiry and the national science education standards: A guide for teaching and learning.* Washington, DC: National Academies Press.

Loverude, M. E., Kautz, C. H., & Heron, P. R. L. (2003). Helping students develop an understanding of Archimedes' principle. I. Research on student understanding. *American Journal of Physics, 71,* 1178–1187.

Marschner, J., Thillmann, H., Wirth, J., & Leutner, D. (2012). Wie lässt sich die Experimentierstrategie-Nutzung fördern? Ein Vergleich verschiedener gestalteter Prompts. *Zeitschrift für Erziehungswissenschaft, 15,* 77–93.

Meindertsma, H. B., van Dijk, M. W. G., Steenbeek, H. W., & van Geert, P. L. C. (2014). Stabilty and variability in young children's understanding of floating and sinking duyring one single-task session. *Mind, Brain, and Education, 8,* 149–158.

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

Molenaar, I., van Boxtel, C. A. M., & Sleegers, P. J. C. (2010). The effects of scaffolding metacognitive activities in small groups. *Computers in Human Behavior, 26,* 1727–1738.

Molenaar, I., van Boxtel, C. A. M., & Sleegers, P. J. C. (2011). Metacognitive scaffolding in an innovative learning arrangement. *Instructional Science, 39,* 785–803.

Molenaar, I., Sleegers, P., & van Boxtel, C. (2014). Metacognitive scaffolding during collaborative learning: A promising combination. *Metacognition and Learning, 9,* 309–332.

Pea, R. D. (2004). The social and technological dimensions of scaffolding and related theoretical concepts for learning, education, and human activity. *Journal of the Learning Sciences, 13,* 423–451.

Pedaste, M., Mäeots, M., Siiman, L. A., de Jong, T., van Riesen, S. A., Kamp, E. T., et al. (2015). Phases of inquiry-based learning: Definitions and the inquiry cycle. *Educational Research Review, 14*, 47–61.

Pollock, E., Chandler, P., & Sweller, J. (2002). Assimilating complex information. *Learning and Instruction, 12,* 61–86.

Quintana, C., Reiser, B. J., Davis, E. A., Krajcik, J., Fretz, E., Duncan, R. G., et al. (2004). A scaffolding design framework for software to support science inquiry. *Journal of the Learning Sciences, 13,* 337–386.

Rappolt-Schlichtmann, G., Tenenbaum, H. R., Koepke, M. F., & Fischer, K. W. (2007). Transient and robust knowledge: Contextual support and the dynamics of children's reasoning about density. *Mind, Brain, and Education, 1,* 98–108.

Reiser, B. J. (2004). Scaffolding complex learning: The mechanisms of structuring and problematizing student work. *Journal of the Learning Sciences, 13,* 273–304.

Saye, J. W., & Brush, T. (2002). Scaffolding critical reasoning about history and social issues in multimedia-supported learning environments. *Educational Technology Research and Development, 50,* 77–96.

Simons, K. D., & Klein, J. D. (2007). The impact of scaffolding and student achievement levels in a problem-based learning environment. *Instructional Science, 35,* 41–72.

Splash: Virtual Buoyancy Laboratory. (2015). http://www.golabz.eu/lab/splash-virtual-buoyancy-laboratory. Accessed 25 January 2018.

Tsirgi, J. E. (1980). Sensible reasoning: A hypothesis about hypotheses. *Child Development, 51,* 1–10.

van Joolingen, W. R., Giemza, A., Bollen, L., Bodin, M., Manske, S., Engler, J., et al. (2011). *SCY cognitive scaffolds and tools (DII.2).* Twente: SCY Consortium.

van Joolingen, W. & Zacharia, Z. C. (2009). Developments in inquiry learning. In N. Balacheff, S. Ludvigsen, T. de Jong, A. Lazonder, & S. Barnes (Eds.), *Technology-enhanced learning: A Kaleidosope view* (pp. 21–37). Dordrecht: Springer.

Veermans, K., van Joolingen, W. R., & de Jong, T. (2006). Use of heuristics to facilitate scientific discovery learning in a simulation learning environment in a physics domain. *International Journal of Science Education, 28*(4), 341–361.

Veenman, M. V. J., van Hout-Wolters, B. H. A. M., & Afflerbach, P. (2006). Metacognition and learning: conceptual and methodological considerations. *Metacognition and Learning, 1,* 3–14.

Wood, D., Bruner, J., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 17,* 89–100.

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.

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.

Zacharia, Z. C., Manoli, C., Xenofontos, N., de Jong, T., Pedaste, M., van Riesen, S. A., et al. (2015). Identifying potential types of guidance for supporting student inquiry when using virtual and remote labs in science: A literature review. *Educational Technology Research & Development, 63*(2), 257–302.

Zervas, P. (Ed.). (2013). *The Go-Lab inventory and integration of online labs—Labs offered by large scientific organisations.* D2.1. Go-Lab Project (Global Online Science Labs for Inquiry Learning at School). Retrieved from http://www.go-lab-project.eu/sites/default/files/files/deliverable/file/Go-Lab-D2.1.pdf.

Charalampos Efstathiou is a research collaborator of the Research in Science and Technology Education Group at the University of Cyprus. His research interests focus on STEM education.

Tasos Hovardas is a research associate of the Research in Science and Technology Education Group at the University of Cyprus. His research interests focus on integrating computer supported inquiry learning in the domain of environmental education.

Nikoletta A. Xenofontos is a PhD candidate in the Department of Educational Sciences at the University of Cyprus. Her research interests focus on enhancing students' conceptual understanding through the use of computer supported inquiry learning.

Zacharias C. Zacharia is the director of the Research in Science and Technology Education Group at the University of Cyprus and an Associate Professor of Science Education in the Department of Educational Sciences at the University of Cyprus. His research interests focus on the implementation of computer supported inquiry learning in science education.

Ton deJong is full professor and department head of the Department of Instructional Technology at the University Twente. His research interests include technologies for learning, inquiry learning and problem solving in STEM education.

Anjo Anjewierden is a researcher in the Department of Instructional Technology at the University Twente. His research interests focus on science education, data mining, and artificial intelligence.

Siswa A. N. van Riesen is a PhD candidate in the Department of Instructional Technology at the University Twente. Her research interests focus on enhancing students' conceptual knowledge and inquiry skills within online learning environments.