Using Virtual Labs in an Inquiry Context:

The Effect of Hypothesis Formulation and Experiment Design Tools on Student Learning

Nikoletta Xenofontos, Zacharias Zacharia, Tasos Hovardas University of Cyprus, Cyprus

ABSTRACT

The present study employed two different Go-Lab tools. These tools were used by primary school students to carry out successive learning tasks during experimentation. The first tool assisted learners in formulating hypotheses, while the second guided students in experimental design. Both tools were designed to take into account the trade-offs between structuring and problematizing student inquiry. The aim of the study was to investigate the effect of each tool separately, as well as the combined effect of the tools in supporting student work. Participants were 41 fifth graders from two classes of a public primary school in Cyprus. They were randomly assigned to four conditions: Condition 1 involved the use of both tools, Condition 2 included the hypothesis tool only, Condition 3 included the experiment design tool only, and Condition 4 had no tools provided. The students in the conditions including one of the two tools outperformed the condition with no tools in the corresponding skill scaffolded by the tool. The cumulative effect of both tools seems to have been greater than the effect of each tool separately.

Keywords: Computer-supported Learning Environments; Go-Lab; Problematizing Student Inquiry; Software Scaffolds; Structuring Student Inquiry

INTRODUCTION

The aim of this chapter was to examine the effect of two scaffold tools, namely a formulating hypotheses tool and an experiment design tool, on students' learning. This work builds on the need for designing computer-supported learning platforms that promote students' knowledge and skills. Previous research has highlighted the benefits that accumulate from the use of such learning environments (e.g., de Jong, van Joolingen, Giemza, Girault, Hoppe, Kindermann, et al., 2010; de Jong, Weinberger, van Joolingen, Ludvigsen, Girault, Kluge et al., 2012). However, it has also stressed the need for providing guidance to the students when using these learning environments due to the complexity involved, not only in their architecture/organization, but also in the pedagogy accompanying the enactment of the learning activities (e.g., Hovardas, 2016). In this case, we test the effectiveness of the aforementioned scaffolding tools in an inquiry science learning context.

BACKGROUND

The optimal degree of guidance for supporting student inquiry in science education has long been debated (Arnold, Kremer, & Mayer, 2014). Previous research has highlighted the possibility that guided inquiry could be beneficial for learners, for instance, in improving science process skills (e.g., Kirschner, Sweller, & Clark, 2006; Koksal & Berberoglou, 2014). However, there is always the need to engage students as active learners in inquiry-based science instruction, capable of taking over responsibility for a range of tasks (Minner, Jurist Levy, & Century, 2010). This unresolved controversy over emphasis on guidance, at the one extreme, and openness, at the other, has been also

reflected in the design of computer-supported learning environments. In this case, guidance is taken over by software scaffolds, which aim to structure student tasks in order to decrease complexity and offload certain aspects of a variety of tasks (de Jong, 2006; Pea, 2004; Reiser, 2004; Reiser, Tabak, Sandoval, Smith, Steinmuller, & Leone, 2001; Simons & Klein, 2007; van Joolingen, 1999). If technology can narrow down the multiplicity of potential routes students might follow, then student effort can be devoted to following these more tractable trajectories. However, a rigidly structured learning activity sequence would reduce student ownership and responsibility of their inquiry (e.g., Chang, Chen, Lin, & Sung, 2008).

The challenge of configuring the optimal balance between guidance and openness in inquiry learning in computer-supported learning environments translates into tension between structuring student work, on the one hand, and "problematizing" student inquiry, on the other (Reiser, 2004). Eliminating task complexity, overall, might reduce student active engagement and lock them into unproductive pathways. After any learning gain has been accomplished, the tasks that follow should challenge students to move on at a higher level, beyond their current expertise (Kalyuga, 2007). In contrast to structuring, which reduces complexity, problematizing student inquiry introduces complexity (Reiser, 2004), at least up to a point, so that the difficulty students are confronted with always surpasses the knowledge and skills they have already acquired. By adding such challenge, learner focus is usually re-directed towards parts of the task that otherwise might not be addressed (Reiser, 2004). Despite the unsettled theoretical and methodological interplay between structuring and problematizing student work, to the best of our knowledge, the question of how to problematize inquiry has not yet received the attention it deserves in the relevant literature (see the work of Wieman on PhET simulations for a notable exception, e.g., Wieman, Adams, & Perkins, 2008).

The contrast between structuring and problematizing student inquiry is pronounced in procedures that involve a series of interrelated tasks to be completed (Reiser, 2004). Such a situation is encountered in scientific experimentation, which involves identifying variables, formulating hypotheses, designing and executing experiments, gathering, analysing, and interpreting data (e.g., Germann, Aram, & Burke, 1996; Hofstein, Navon, Kipnis, & Mamlok-Naaman, 2005; Kremer, Specht, Urhahne, & Mayer, 2014, van Joolingen & Zacharia, 2009). Students face quite a few obstacles in designing and executing valid experiments (de Jong, 2006; Reiser, 2004; Zacharia, Manoli, Xenofontos, de Jong, Pedaste, van Riesen, et al., 2015); such obstacles include, among others, classifying variables as dependent, independent and controlled, and planning experimental trials (e.g., Arnold et al., 2014; Chinn & Malhotra, 2002; De Boer, Quellmalz, Davenport, Timms, Herrmann-Abell, Buckley, Jordan, et al., 2014; Lin & Lehman, 1999; Roberts & Gott, 2003; van Joolingen & de Jong, 1991). Due to the modular and difficult nature of experimentation, obstacles have been identified even among older students (Arnold et al., 2014; Furtak, 2006; Germann et al., 1996; Kirschner et al., 2006). Therefore, it should not be surprising that experimentation is the part of the inquiry cycle that has been most often supported by software scaffolds (Zacharia et al., 2015).

Research on the impact of software scaffolds on experimentation has delivered mixed results (for a comprehensive review of support of student inquiry in computer-supported learning environments see Zacharia et al., 2015). More to the point, previous studies have been confined to examining separate tools and scaffolds, whereas experimentation involves a set of stages that need to be effectively executed.

MAIN FOCUS OF THE CHAPTER

The present study employed two different tools, which were developed within the Go-Lab project (http://www.go-lab-project.eu/), and which were used by primary school students to carry out successive learning tasks during experimentation. The first tool assisted learners in formulating hypotheses ("Hypothesis Scratchpad"; http://www.golabz.eu/app/hypothesis-tool), while the second tool guided students in designing experiments ("Experiment Design Tool"; http://www.golabz.eu/apps/experiment-design-tool). Both tools were designed to take into account the trade-offs between structuring and problematizing student inquiry, as previously discussed [for analogous considerations see Etkina, Karelina, Ruibal-Villasenor, Rosengrant, Jordan, & Hmelo-

Silver (2010) and Jordan, Ruibal-Villasenor, Hmelo-Silver, & Etkina (2011)]. The core objective of the study was to investigate the effect of each tool separately as well as the combined effect of the tools in supporting student work. In this regard, we aimed at answering the following research questions:

- 1. Does the use or no use of either of the aforementioned tools differentiate students' inquiry performance (for the inquiry performance dimensions see Table 2)?
- 2. For students, who used both tools, does this aggregate tool usage yield higher learning gains as compared to using each tool separately?
- 3. Can students apply the inquiry skills they had acquired during the study's treatment in a novel inquiry context?
- 4. Can a valid hypothesis correlate with a valid experimental design and, further, can a valid experimental design be accompanied by effective execution of the experiment in a virtual lab?
- 5. What is the relative weight of each variable studied (i.e., variables reflecting student performance and referring to software scaffolds and to learning products during the learning activity sequence) across conditions (i.e., participants given each tool separately; those given both tools; and those given no tools)?

Methods

Learning Environment

The Graasp authoring tool was used to create an online Inquiry Learning Space (ILS) (for details on ILSs, see: de Jong, Sotiriou, & Gillet, 2014; Govaerts, Cao, Vozniuk, Holzer, Zutin, Ruiz, et al. 2013; Rodríguez-Triana, Holzer, Vozniuk, & Gillet, 2015), following the inquiry cycle design framework (Pedaste et al., 2015). The Graasp authoring tool allows teachers to embed resources offered via the Go-Lab platform (e.g., software scaffolds and virtual or remote laboratories) in subsequent phases of student inquiry (see http://www.golabz.eu/video/create-ils-graasp for detailed instructions of how to create an Inquiring Learning Space from Graasp). An ILS is an online computer-supported learning environment, which is designed as a template within the Go-Lab project. The ILS is structured around a virtual or remote laboratory (http://www.golabz.eu/labs) and provides software scaffolds for students undertaking learning tasks (http://www.golabz.eu/apps). The content of the ILS referred to electrical circuits and included the Electrical Circuit Lab (Figure 1), which is available on the Go-Lab platform (http://www.golabz.eu/lab/electrical-circuit-lab). The focus of the study was on the effect of two tools, namely, the Hypothesis Scratchpad (HS) and the Experiment Design Tool (EDT), both when these tools are used separately, and when they are present together in the learning environment. In the latter case, the combined effect of both tools was examined. Therefore, four different versions of the same ILS were developed. The first version included both the HS and the EDT, the second and third versions included only the HS or the EDT, respectively, while the fourth version included neither the HS nor the EDT. Whenever a software scaffold was absent, it was replaced by an Input Box (http://www.golabz.eu/apps/input-box), which is a simple note-taking application and does not provide the specific scaffolding functionalities of either the HS or the EDT.

The Hypothesis Scratchpad

Terms needed for formulating a hypothesis were given in the upper part of the HS (Figure 2). Students could drag and drop predefined conditionals and concepts in the space provided by the tool to create a hypothesis in the form of an "if…then" statement. Students could also create their own words or phrases in order to use them in their hypotheses by typing them in the gray box in the tool. Students who used the Input Box instead of the HS formulated their hypotheses without receiving any support in the form of keywords.

The Experiment Design Tool

The EDT (Figure 3) included, first, a classification task, where students had to distinguish which variable to vary (independent variable), which variables to keep constant (control variables), and which variable to measure (dependent variable). To do so, students dragged pre-set variables from the left side of the tool's interface and dropped them in the proper column. A second task involved the arrangement of experimental trials to be undertaken. Students had to specify the values of each

variable in each experimental trial they added to their experimental design. Students who used the Input Box instead of the EDT completed their experimental designs without receiving any support in terms of classifying variables or setting values for their experimental trials.

Participants

Participants were 41 fifth graders (10-11 years old) from two classes of a public primary school in Larnaca, Cyprus. They were randomly assigned to four conditions (Table 1): Condition 1 involved both tools (HS+EDT; 11 students; 6 boys, 5 girls), Condition 2 included the HS only (12 students; 6 boys, 6 girls), Condition 3 included the EDT only (9 students; 4 boys, 5 girls), and Condition 4 involved neither the HS nor the EDT (9 students; 5 boys, 4 girls). Students in all conditions were taught about the simple electrical circuit at the beginning of the school year, while the current study was conducted at the end of the same school year. Two pre-tests showed that conditions differed in neither the skills of formulating hypotheses and designing experiments (two scales from the TIPSII inquiry skills test, focused on formulating hypotheses and designing experiments, respectively; see Burns, Okey, & Wise, 1985) nor in their prior knowledge (knowledge dimensions examined were "remember", "understand" and "apply"; for a detailed description of pre-tests see the sub-section on Data collection and analyses). All students had basic computer and processing skills.

Learning Activity Sequence

The learning activity sequence started with an Orientation phase, where students were first reminded of the simple electrical circuit, and then, they were introduced to circuits connected in series and in parallel, through videos, diagrams and text (see Figure 4 for a complete flow of inquiry phases). This preliminary set of learning activities ended with a problem presented to students, which referred to how light fixtures in a house are connected. After the Orientation phase, students proceeded to the Conceptualization phase, where they had to predict how the brightness of bulbs connected in series and in parallel would differ from the brightness of bulbs in a simple electrical circuit. Afterwards, students formulated hypotheses on how the brightness of bulbs would be impacted when adding more bulbs to circuits in series and in parallel. The next phase involved designing an experiment and executing the experiment in the Electrical Circuit Lab (Investigation phase). When the experimentation procedure had been completed, students responded to several questions in order to interpret their results (see Appendix 1). The last activity in the Investigation phase was an examination of what would happen when a bulb in a circuit, connected either in series or in parallel, burned out but was not removed from the circuit. In the Conclusion phase, students were asked to provide an answer for the initial problem stated in the ILS, which was about how the light fixtures in a house are connected. Students were prompted to provide enough evidence to justify their answer. The inquiry cycle ended with the Discussion phase, in which students responded to several reflection questions. They were asked to describe the steps that they went through in order to address the initial problem presented in the lesson. They were further asked if they had completed all the activities in the learning environment and if they could think of any activity which could have been done in a different way.

Procedure

The implementation lasted for three class meetings. In the first meeting (80 minutes), each condition completed pre-tests and undertook a familiarization activity with the tools and the virtual lab to be used, in a different context (weather) than the one encountered later on in the ILS (electrical circuits). Students were instructed explicitly about how to formulate a hypothesis in the form of an "if...then..." statement. Then they had the opportunity to create their own hypotheses regarding how weather would affect children's decision to play indoors or outdoors. Students were also explicitly taught the VOTAT strategy ("Vary One Thing At a Time"; also referred to as the "control-of-variables" strategy – CVS), in order to execute fair experiments, and they tried to set up a fair experiment for the hypotheses they had previously formulated in the weather context. VOTAT is a heuristic in designing experiments, where manipulating one independent variable at a time allows learners to attribute any change in the dependent variable to the independent variable which was manipulated (Glaser, Schauble, Raghavan, & Zeitz, 1992; Lin & Lehman, 1999; Tsirgi, 1980; Klahr, & Nigam, 2004; Veermans, van Joolingen & de Jong, 2006). A demonstration of the Electrical Circuit

Lab followed, where each student had the opportunity to create several circuits in the lab (see Figures 1a and 1b).

In the second meeting (80 minutes), students went through the ILS. Attention was paid to time-on-task effects so that participants in each condition spent about the same time to accomplish the entire learning activity sequence. The teacher mainly provided technical support to students when it was necessary; for instance, when students had accidentally exited the learning environment, the teacher would assist them with re-entering the ILS. For content-specific issues, the recommendation to students was to go through the instructions and hints included in the learning environment. In the third meeting (50 minutes), students in every condition used the HS and the EDT to formulate hypotheses and design experiments, respectively, in two new learning contexts. The first context was about rolling marbles in an inclined ramp and the second context addressed the solubility of sugar in water.

Data Collection and Analyses

Assessment of students' prior knowledge and inquiry skills involved two different instruments administered in a pre-test format. The knowledge test (see Appendix 2) consisted of four items, which corresponded to three cognitive processes termed "remember", "understand" and "apply", based on the revised Bloom's (1956) taxonomy, as presented by Anderson and Krathwohl (2001) and further elaborated in the reports edited by de Jong (2014) and Zervas (2013). For the inquiry skills test, items included in the TIPSII were selected and translated into Greek (Burns et al., 1985). The inquiry skills test consisted of 5 multiple-choice items, where 3 items referred to "identifying and stating hypothesis", and another 2 items referred to "designing investigations". The selection of TIPSII items was based on the appropriateness of wording and content in relation to student age. Both tests were scored blind to the students' assigned condition. A rubric was used for evaluation of the open-ended items on the knowledge test, and two independent coders scored 20% of the data. The inter-rater agreement between the coders was found to be high (Cohen's Kappa = 0.87).

Apart from the pre-tests, computer screen-captured data were collected for all students during the completion of the ILS (second meeting) and during the activities in the new learning contexts (third meeting) by means of RiverPast software (see http://river-past-screen-recorder-pro.soft112.com/ for a technical description of the software). Data analysis of this material focused on the learning products constructed by students as they progressed along the learning activity sequence (see in this regard Hovardas, 2016). Specifically, the activity sequence involved students' hypotheses, students' experimental designs, correspondence of students' hypotheses with their experimental designs, and correspondence of students' use of the virtual laboratory with their experimental designs. All the variables referring to software scaffolds and learning products constructed by students during the learning activity sequence are shown in Table 2. Coding schemes for scoring learning products were developed and two coders independently rated 10% of each category of learning products. Inter-rater reliability (proportion of agreement) amounted to over 85% across all categories, while divergences between raters were settled through discussion.

Non-parametric tests and analyses were used to investigate trends in data, which involved *Kruskal-Wallis* tests, *Mann-Whitney* tests, *Wilcoxon* two-related samples tests, and *Spearman's* rank correlations. In addition, a correspondence analysis was performed to examine the relative weight of each variable studied across conditions.

RESULTS

Research Question 1

Table 3 presents mean values for all variables studied, by assigned condition and learning context. An overview of the table indicates that there were significant differences among conditions in most cases (i.e., in both the learning activity sequence and the new learning contexts). More specifically, conditions providing one of the two tools outperformed the condition with no tools in the corresponding skill pertaining to that tool. With regard to student performance when working in the ILS, students who used the HS only (Condition 2; HS only) scored higher for their hypotheses (Table 3; "ScoreHypo") than students who did not use this tool (Condition 4; no tool) (*Mann-Whitney Z* = -2.73, p < 0.01). Additionally, students who used the EDT only (Condition 3; EDT only) showed

increased implementation of the VOTAT heuristic (Table 3; "EDT_VOTAT"), compared to students who did not use this tool (Condition 4; no tool) ($Mann-Whitney\ Z = -4.12,\ p < 0.001$).

Research Questions 2 and 3

The combined effect of both tools seems to have outweighed the effect of each tool separately in the new contexts provided to students for experimentation after they concluded activities in the ILS. This was especially evident in the case of formulating hypotheses (Table 3; "postScoreHypo") (*Mann-Whitney Z* = -2.05, p < 0.05, for the difference between Conditions 1 and 2; *Mann-Whitney Z* = -3.16, p < 0.01, for the difference between Conditions 1 and 3), planning experimental trials (Table 3; "postEDT_Trials") (*Mann-Whitney Z* = -3.18, p < 0.01, for the difference between Conditions 1 and 2; *Mann-Whitney Z* = -2.48, p < 0.05, for the difference between Conditions 1 and 3) and the correspondence between student hypotheses and experimental designs (Table 3; "postHS_EDT") (*Mann-Whitney Z* = -2.26, p < 0.05, for the difference between Conditions 1 and 2; *Mann-Whitney Z* = -2.81, p < 0.01, for the difference between Conditions 1 and 3).

Overall, the condition that employed both tools (Condition 1; HS+EDT) showed increased scores for most of the studied variables in the new learning contexts, namely, the new contexts given to students after they had exited the ILS, in comparison to the values recorded in the ILS (Table 3). This finding provides an indication that the beneficial effect of software scaffolding on student inquiry performance was either maintained or even increased across the learning contexts for Condition 1. The opposite can be observed for the other three conditions, where students scored lower across all variables in the new learning contexts as compared to the ILS. Students who used both tools (Condition 1) outperformed students who used only one tool or no tool across all dimensions examined in the new learning contexts. However, *Wilcoxon* two-related-samples tests revealed that there were no significant time trends for any condition from the ILS to the new learning contexts.

Research Question 4

Scores for the condition with both tools (Condition 1) showed significant correlations among various parameters studied. Hypothesis scores correlated positively with correspondence of hypotheses and experimental designs (Spearman's rho = 0.74, p < 0.05), while planning of experimental trials correlated positively with correspondence of student use of the virtual laboratory with their experimental designs (Spearman's rho = 0.98, p < 0.001). Another significant correlation was between planning experimental trials and employing the VOTAT heuristic in the new learning contexts (Spearman's rho = 0.71, p < 0.05). No such correlations were found in the other conditions. All of these results offer a strong indication that for Condition 1 (HS+EDT), student competence on a learning task could have positively catalyzed performance on forthcoming learning tasks in either the ILS or in the new learning contexts. Namely, using both tools seemed to offer an added value within the learning process, as compared to using one tool only. In other words, the use of both tools not only reinforced separate dimensions of student performance but it also initiated a co-evolvement of skills.

Research Question 5

A correspondence analysis was performed in order to investigate which variables were most pronounced according to tool usage (i.e., Condition 1: HS+EDT, Condition 2: HS, Condition 3: EDT, Condition 4: no tool). This analysis aimed at revealing aspects of the student performance (i.e., variables plotted by the analysis) across varying tool usage and acrossconditions. The biplot of the analysis is shown in Figure 5. Along the biplot, white cycles denote variables under study (i.e, variables depicting various dimensions of student performance) are positioned according to tool usage (i.e., conditions students had been assigned to with regard to tool usage). The closer a specific dimension of student performance (i.e., variable) is placed to a condition reflecting tool usage, the more expressed this dimension had been for that condition. On the positive part of the first axis we can observe that Condition 3 (EDT only) is related to the implementation of the VOTAT heuristic in the ILS as well as in the new learning contexts ("EDT_VOTAT" and "postEDT_VOTAT", respectively). On the negative part of the first axis, Condition 2 (HS only) is characterized by relatively increased scores for hypotheses in both the ILS and the new learning contexts ("ScoreHypo" and "postScoreHypo", respectively). Condition 1 (HS + EDT) features in the negative part of the second axis and it is distinguished by a marked correspondence between hypothesis formulation and experimental designs in the new learning context ("postHS EDT") as well as by

increased validity in planning experimental trials in the new learning contexts ("postEDT_Trials"). The relative position of conditions and variables on the biplot implies that each tool separately fostered performance on learning tasks related to its scaffolding properties. Namely, the HS reinforced student ability to adequately formulate hypotheses, while the EDT enhanced student competence in implementing the VOTAT heuristic. Indeed, this was the case in the learning context of the ILS as well as in the new learning contexts. The condition that combined both tools (Condition 1; HS + EDT) was marked by relatively increased inter-contextual transfer. The latter means that the learning gains in that condition were highly probable to be also detected in an upcoming learning context. That condition was also characterized by inter-task transferability of learning gains. For instance, an increased validity in student hypotheses was highly possible to be accompanied later on by a valid experimental design.

CONCLUSION AND DISCUSSION

A first major result of the study was that each software scaffold separately succeeded in supporting students in the corresponding learning tasks, which addressed the first research question. Such an encouraging finding must be attributed to the design and properties of each tool. Further, it implies that both tools may have adequately handled the two contrasting needs of structuring student inquiry on the one hand, and problematizing student work, on the other. An explanation for that delicate and balanced contribution might be that both tools were fairly effective in offering procedural information to students, namely, information on what students should do and how they should address the corresponding learning tasks (see, for instance, Arnold et al., 2014). The implication coming out of this finding, for instruction purposes, is that specific configurations of software scaffolds might prove adequate both for decreasing complexity in student inquiry, as well as for actively engaging students in subject matter. A related explanation might refer to the tools' allowance for serial processing of learning tasks (see, for instance, the VOTAT strategy that was related to the use of the EDT), where the complexity in a learning activity sequence might be addressed by partitioning tasks into smallerscale, shorter assignments that need to be dealt with one after the other (Clarke, Ayres, & Sweller, 2005; Pollock, Chandler, & Sweller, 2002). Such a serial processing of learning tasks might have kept essential cognitive load within the limits that learners could manage based on their cognitive capacity and the processing limitations of working memory (Kalyuga, 2007; Sweller, van Merrienboer, & Paas, 1998). It might also have kept students alert and motivated to encounter upcoming tasks, which could have adequately touched upon the dimension of problematizing student inquiry.

With regard to the second research question, the results presented in this paper indicate that students using both tools benefited more as compared to students who used one tool only. This finding is a clear sign of a combined effect, where the impact of both software scaffolds might have outweighed the impact of each tool separately. Concerning the third research question, the combined effect of tool usage was also detected in new inquiry contexts, which is a strong indication of inter-contextual transfer of skill gains. An additional combined effect was implied by significant correlations among studied parameters in the condition that incorporated both tools (fourth research question). These correlations indicate that using both tools might not only increase the corresponding skills separately (i.e., formulating hypotheses and designing experiments) but that it might also result in skill gains that spread across learning tasks as the learning activity sequence unfolds. Namely, skill gains in formulating hypotheses might pass on to designing experiments and the latter might pass on to handling a virtual lab. Although previous research may have indicated such a linkage (e.g., Arnold et al., 2014; Veermans, van Joolingen, & de Jong, 2006), the results of this study offer the first empirical validation of this transfer. It is also highlighted that the results imply that transfer could involve two different scales. The first addresses transition between different contexts (inter-contextual transfer), while the second is anchored within a single learning activity sequence and refers to gains being transferred from a learning task to a forthcoming learning activity (inter-task transfer). Results concerning transfer were further validated by the correspondence analysis (fifth research question).

FUTURE RESEARCH DIRECTIONS

Previous research has highlighted the benefits of storing, retrieving, and exchanging learning products in computer-supported inquiry learning environments (e.g., de Jong et al., 2010, 2012). It has also stressed the potential of using learning products for diagnosing student performance and enacting formative assessment (e.g., Hovardas, 2016). This study's findings also concur with these claims. However, this research domain is still in need of further research. There are quite a few unresolved issues. For example, there is a need to examine whether learning products, which have been constructed and stored during a learning activity sequence, might themselves serve as scaffolds for student inquiry either during upcoming activities or when students would encounter new learning contexts.

Future research could also use learning products to investigate additional functionalities, which software scaffolds could carry out. For example, the HS can include varying numbers of words (variables involved in the phenomenon under study plus conditionals necessary to interrelate variables) to support students in formulating hypotheses. Varying the number of words provided to students might be a way to vary the degree of scaffolding the HS offers to students, perhaps with fading of the scaffolding as learners gain experience. This would mean that the HS could include fewer words for more experienced learners (see in this regard Kalyuga & Sweller, 2004; Reisslein, Atkinson, Seeling, & Reisslein, 2006; Seufert & Brünken, 2006). In the case of the EDT, fading in and out of scaffolding could involve the number of variables offered to students to design their experiments. Since most analogous tools in computer-supported learning environments are delivered in a "one-for-all fashion" (Kalyuga, 2007), future research might screen optimal timing for introducing and removing scaffolding, as well as varying of scaffolding based on student experience and competence (see, for instance, de Jong, 2006). If learners proceed effectively in a learning activity sequence even after scaffolding has been removed, this would provide a substantial indication that the scaffolded skill has been acquired. Future research might further examine possible effects on cognitive load and demands on working memory imposed by different scaffolding configurations for tools.

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 and Education*, *51*, 1486-1498.

Chinn, C. A., & Malhotra, B. A. (2002). Epistemologically authentic inquiry in schools: A theoretical framework for evaluating inquiry tasks. *Science Education*, *86*, 175-218.

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 Boer, G. E., Quellmalz, E. S., Davenport, J. L., Timms, M. J., Herrmann-Abell, C. F., Buckley, B. C., Jordan, K. A., Huang, C.-W., & Flanagan, J. C. (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)*. 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., & van Joolingen, W. R. (1998). Scientific discovery learning with computer simulations of conceptual domains. *Review of Educational Research*, 68, 179-201.
- de Jong, T., van Joolingen, W. R., Giemza, A., Girault, I., Hoppe, U., Kindermann, J., Kluge, A., Lazonder, A. W., Vold, V., Weinberger, A., Weinbrenner, S., Wichmann, A., Anjewierden, A., Bodin, M., Bollen, L., d'Ham, C., Dolonen, J., Engler, J., Geraedts, C., Grosskreutz, H., Hovardas, T., Julien, R., Lechner, J., Ludvigsen, S., Matteman, Y., Meistadt, Ø., Næss, B., Ney, M., Pedaste, M., Perritano, A., Rinket, M., von Schlanbusch, H., Sarapuu, T., Schulz, F., Sikken, J., Slotta, J., Toussaint, J., Verkade, A., Wajeman, C., Wasson, B., Zacharia, Z. C., & van der Zanden, M. (2010). Learning by creating and exchanging objects: the SCY experience. *British Journal of Educational Technology, 41*, 909-921.
- de Jong, T., Weinberger, A., van Joolingen, W.R., Ludvigsen, S., Ney, M., Girault, I., Kluge, A., Lazonder, A., Pedaste, M., Wasson, B., Wichmann, A., Geraedts, C., Hovardas, T., Julien, R., Lejeune, A., Manoli, C., Matteman, Y., Sarapuu, T., Vold, V., Wanders, B., & Zacharia, Z. (2012). Designing complex and open learning environments based on scenarios. *Educational Technology Research & Development*, 60, 883-901.
- Etkina, E., Karelina, A., Ruibal-Villasenor, M., Rosengrant, D., Jordan, R., & Hmelo-Silver, C. E. (2010) Design and reflection help students develop scientific abilities: Learning in introductory physics laboratories. *Journal of the Learning Sciences*, 19, 54-98.
- Furtak, E. M. (2006). The problem with answers: An exploration of guided scientific inquiry teaching. *Science Education*, *90*, 453-467.
- Germann, P. J., Aram, A., & Burke, G. (1996). Identifying patterns and relationships among the responses of seventh-grade students to the science process skill of designing experiments. *Journal of Research in Science Teaching*, 33, 79-99.
- 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-Verlag.
- Govaerts, S., Cao, Y., Vozniuk, A., Holzer, A., Zutin, D. G., Ruiz, E. S. C., Bollen, L., Manske, S., Faltin, N. Salzmann, C., Tsourlidaki, E., & Gillet, D. (2013). Towards an online lab portal for inquiry-based STEM learning at school. In J.-F. Wang & R. Lau (Eds.), *Advances in web-based learning ICWL 2013, 12th International Conference, Kenting, Taiwan, October 6-9, 2013, Proceedings* (pp. 244-253). Berlin: Springer.
- Hofstein, A., & Lunetta, V. N. (2003). 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.
- Hovardas, T. (2016). A learning progression should address regression: Insights from developing non-linear reasoning in ecology. *Journal of Research in Science Teaching*, 53, 1447-1470.
- Jordan, R. C., Ruibal-Villasenor, M., Hmelo-Silver, C. E., & Etkina, E. (2011). Laboratory materials: Affordances or constraints? *Journal of Research in Science Teaching*, 48, 1010–1025.
- Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review*, 19, 509-539.
- Kalyuga, S., & Sweller, J. (2004). Measuring knowledge to optimize cognitive load factors during instruction. *Journal of Educational Psychology*, *96*, 558-568.
- 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: effect of direct instruction and discovery learning. *Psychological Science*, *15*, 661-667.
- Koksal, E. A., & Berberoglou, G. (2014). The effect of guided inquiry instruction on 6th grade Turkish students' achievement, science process skills, and attitudes toward science. *International Journal of Science Education*, *36*, 66-78.
- 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.
- 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.
- 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.
- 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. N., Kamp, E. T., Manoli, C. C., Zacharia, Z. C., & Tsourlidaki, E. (2015). Phases of inquiry-based learning: Definitions and the inquiry cycle. *Educational Research Review*, *14*, 47-61.
- Pollock, E., Chandler, P., & Sweller, J. (2002). Assimilating complex information. *Learning and Instruction*, 12, 61-86.
- Reiser, B. J. (2004). Scaffolding complex learning: The mechanisms of structuring and problematizing student work. *Journal of the Learning Sciences*, *13*, 273-304.
- Reiser, B. J., Tabak, I., Sandoval, W. A., Smith, B. K., Steinmuller, F., & Leone, A. J. (2001). BGuILE: Strategic and conceptual scaffolds for scientific inquiry in biology classrooms. In S. M. Carver & D. Klahr (Eds.), *Cognition and instruction: Twenty-five years of progress* (pp. 263-305). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

- Reisslein, J., Atkinson, R. K., Seeling, P., & Reisslein, M. (2006). Encountering the expertise reversal effect with a computer-based environment on electrical circuit analysis. *Learning and Instruction*, 16, 92-103.
- Roberts, R., & Gott, R. (2003). Assessment of biology investigations. *Journal of Biological Education*, 37, 114-121.
- Rodríguez-Triana, M. J., Holzer, A., Vozniuk, A., & Gillet, D. (2015). Orchestrating inquiry-based learning spaces: An analysis of teacher needs. In F. W. B. Li, R. Klamma, M. Laanpere, J. Zhang, B. F. Manjón, & R. W. H. Lau (Eds.), *Advances in web-based learning --ICWL 2015*, *14th International Conference, Guangzhou, China, November 5-8*, 2015, *Proceedings* (pp. 131-142). Cham, Switzerland: Springer.
- Seufert, T., & Brünken, R. (2006). Cognitive load and the format of instructional aids for coherence formation. *Applied Cognitive Psychology*, 20, 321-331.
- 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.
- Sweller, J., van Merrienboer, J., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, 10, 251-296.
- Tsirgi, J. E. (1980). Sensible reasoning: A hypothesis about hypotheses. *Child Development*, *51*, 1-10.
- van Joolingen, W. R. (1999). Cognitive tools for discovery learning. *International Journal of Artificial Intelligence in Education*, 10, 385-397.
- van Joolingen, W. R., & de Jong, T. (1991). Supporting hypothesis generation by learners exploring an interactive computer simulation. *Instructional Science*, 20, 389-404.
- van Joolingen, W. R., & Zacharia, Z. C. (2009). Developments in inquiry learning. In N. Balacheff, S. Ludvigsen, T. de Jong, & S. Barnes (Eds.), *Technology-enhanced learning* (pp. 21-37). Dordrecht, The Netherlands: 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.
- Wieman, C. E., Adams, W. K., & Perkins, K. K. (2008). PhET: Simulations that enhance learning. *Science*, 322, 682-683.
- Zacharia, Z. C., Manoli, C., Xenofontos, N., de Jong, T., Pedaste, M., van Riesen, S. A. N., Kamp, E. T., Mäeots, M., Siiman, L., & Tsourlidaki, E. (2015). Identifying potential types of guidance for supporting student inquiry when using virtual and remote labs in science: A literature review. *Educational Technology Research and Development*, 63, 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). http://www.go-lab-project.eu/sites/default/files/files/deliverable/file/Go-Lab-D2.1.pdf

KEY TERMS AND DEFINITIONS

Coherence in an inquiry-based learning activity sequence: Many learning tasks can be interrelated along a learning activity sequence. This is especially pronounced in scientific experimentation, where effective identification of variables is reflected in formulating hypotheses, and then, in designing and executing experiments. Coherence in an inquiry-based learning activity sequence, which entails experimentation, can be studied by analyzing learning products along that sequence, namely, artifacts constructed by students themselves as they complete the sequence of learning tasks. In that regard, a research question might be whether valid hypotheses lead to an acceptable and complete experimental design, and if the latter is followed by thorough execution of the experiment after having adequately planned experimental trials.

Combined effect of software scaffolds: Each software scaffold is expected to have an effect on learning. When more than one scaffold of this kind have been integrated in a computer-supported learning environment, then their combined effect might surpass the impact of each tool separately. Such an effect might emerge as an additional increase in student knowledge or skills, as well as an interrelation of dimensions of their knowledge or skills. The synergistic effect of software scaffolds might be even more pronounced in new learning contexts, where the effect of a combined tool usage on student performance might outperform the sum of the effects on student performance when using each tool separately.

Inter-contextuality: Inter-contextuality concerns the application of acquired knowledge and skills in a novel learning context (not previously experienced). If students apply their knowledge and skills adequately in this new context, this is a quite clear and strong indication of acquired learning gains. Inter-contextuality is of primary importance in subsequent inquiry cycles, because each cycle can be fed by the knowledge and skills already acquired. For example, when students learn how to construct models, then they do not need to receive any further formal instruction in subsequent inquiry cycles that involve model construction.

Introducing and removing scaffolding (fading in / fading out): Guidance provided to students by software scaffolds can be introduced or removed, depending upon the learning task and student experience or prior knowledge and skills. For instance, complex tasks can necessitate full-fledged scaffolding, whereas easier tasks could be attempted with lesser support from software scaffolds. Further, scaffolding can gradually fade out, as students gain experience. More experienced students or those with higher prior knowledge and skills could succeed in effectively undertaking learning tasks with lesser support. The timing of the introduction or removal of scaffolds is related to the controversy between structuring and problematizing student work. For example, more experienced students should need less structuring but more problematizing in their inquiry, so that their motivation and potential for further learning gains are maintained. This assumption would mean, once again, that scaffolding should decrease as student experience progresses.

Problematizing student inquiry: Problematizing student inquiry involves challenging students and attracting their attention to parts of an inquiry-based learning task or activity sequence that might otherwise have been missed. A simple example of how software scaffolding might handle the problematization of student inquiry is when students are notified about a misclassification of variables when distinguishing between dependent, control, and independent variables in an experimental design. In that case, students would need to revisit the task at hand before proceeding with their inquiry.

Structuring student inquiry in computer-supported learning environments: To structure student inquiry in computer-supported learning environments, computer-based guidance is suggested (e.g., scaffolds, heuristics, prompts). The idea is to remove the embedded complexity in an inquiry-based learning process. The nature of the guidance should relate to what is required in order to enact each phase of inquiry (i.e., orientation, conceptualization, investigation, conclusion and discussion) (for details, see Zacharia et al., 2015). For instance, a scaffold designed to support students in developing their experimental designs could help the students identify the dependent, control, and independent variables at hand.

Transfer of knowledge or skills: Student ability to apply knowledge or skills in a new learning situation is a strong indicator of competence acquired in previous learning experiences. This ability has been termed "transfer" exactly because it denotes a transition between two different but related learning instances, namely, between the learning context in which knowledge or skills have been acquired and the new learning context, in which the same knowledge and skills would be applied/transferred.

APPENDIX 1

Questions provided to students as prompts fior data interpretation.

The brightness of the bulbs, which are connected in series, is (1) higher, (2) lower or (3) the same as the brightness of the bulb in a simple electrical circuit?

When bulbs are added in series, their brightness (1) increases, (2) decreases or (3) remains the same?

The brightness of the bulbs, which are connected in parallel, is (1) higher, (2) lower or (3) the same as the brightness of the bulb in a simple electrical circuit?

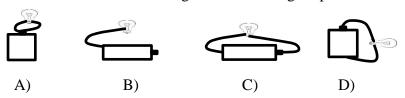
When bulbs are added in parallel, their brightness (1) increases, (2) decreases or (3) remains the same?

APPENDIX 2

Knowledge test

Note: In order to complete this test, you will need approximately 20 minutes. You need to respond to all items (1-6). The results of this test will not count to your total score in the lesson. Your input will be used anonymously for research purposes.

1. In which of the following the bulb will light up? Please choose **one** answer.



- a) A
- b) B
- c) C
- d) D
- e) C και D
- f) A, C και D
- g) None of them

Explain your reasoning:

2. Look at the following circuits:

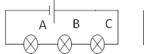


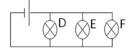




Note: The symbol $\neg \vdash$ represents the battery which is connected to the circuit. An identical battery is connected to all the circuits above.

3. What will happen if the middle bulb burns out (and if it will not be removed from the circuit)? Please choose **one** answer.





Note: The symbol represents the battery which is connected to the circuit. An identical battery is connected to all the circuits above.

- a) A and C will not light up, D and F will have the same brightness
- b) All the bulbs (A, C, D and F) will not light up
- c) A and C will have the same brightness, but less than D and F, which will have the same brightness
- d) All the bulbs (A, C, D and F) will have the same brightness
- e) D and F will not light up, while A and C will have the same brightness
- **4.** Multiple electrical sockets are used for the operation of multiple electrical appliances. How are the electrical appliances connected in multiple sockets? Explain your reasoning.

Figure legends

Figure 1. Two screenshots of the Electrical Circuit Lab (http://www.golabz.eu/lab/electrical-circuit-lab), depicting two bulbs connected in series (a) and three bulbs connected in parallel (b)

Figure 2. A screenshot of the Hypothesis Scratchpad (http://www.golabz.eu/app/hypothesis-tool)

Figure 3. A screenshot of the Experiment Design Tool (http://www.golabz.eu/apps/experiment-design-tool)

Figure 4. Flow of phases and sub-phases of student inquiry. Rectangles depict phases (in bold) or sub-phases, dark rhombuses learning products and white rhombuses any reference material offered to students by the teacher or the learning environment. Arrows show the sequence of phases (for a complete account of the inquiry-based learning framework and the succession of phases and sub-phases along the inquiry cycle see Pedaste et al. 2015).

Figure 5. Bi-plot of correspondence analysis depicting variables studied (white circles: ScoreHypo; EDT_VOTAT; EDT_Trials; HS_EDT; EDT_Trials_Lab; postScoreHypo; postEDT_VOTAT; postEDT_Trials; postHS_EDT) and conditions (black boxes: HS + EDT; HS; EDT; no tool). Overall, the closer a variable is to a condition, the more this condition is characterized by increased values of this variable, as compared to the other conditions. The first axis accounted for 88% of total variance, while the second axis added another 11%.

Table 1. Conditions, tools involved, and number of participants

Conditions	Tools involved	Number of participants
Condition 1	Hypothesis Scratchpad and	11 (6 boys and 5 girls)
	Experiment Design Tool	
Condition 2	Hypothesis Scratchpad	12 (6 boys and 6 girls)
Condition 3	Experiment Design Tool	9 (4 boys and 5 girls)
Condition 4	No tool	9 (5 boys and 4 girls)

Table 2. Variables referring to software scaffolds and learning products during the learning activity sequence

Variable code	Description of variable and values	Measure	Range (min-max)
ScoreHypo	dependent variable (i.e., one that cannot be tested in the Electrical circuit lab); "1" = valid dependent variable but		0-2
EDT MOTAT	missing or invalid independent variable; "2" = valid dependent and independent variable	C - 1 -	0.2
EDT_VOTAT	VOTAT strategy implemented in the ILS; "0" = no implementation; "1" = partial implementation; "2" = full implementation (e.g., across all hypotheses)	Scale	0-2
EDT_Trials	Experimental trials planned in the ILS; "0" = no planning; "1" = partial planning; "2" = full planning (at least two trials were planned for each hypothesis)	Scale	0-2
HS_EDT	Correspondence between hypotheses and experimental designs in the ILS; "0" = no correspondence; "1" = partial correspondence; "2" = full correspondence (e.g., across all hypotheses)	Scale	0-2
EDT_Trials_Lab	Correspondence between experimental designs and circuits in the lab in the ILS; "0" = no correspondence; "1" = partial correspondence; "2" = full correspondence (e.g., across all trials)	Scale	0-2
postScoreHypo	Maximum score across hypotheses formulated by students in the new learning contexts; "0" = no dependent variable included or invalid dependent variable (i.e., one that cannot be tested in the provided labs); "1" = valid dependent variable but missing or invalid independent variable; "2" = valid dependent and independent variable	Scale	0-2
postEDT_VOTAT	VOTAT strategy implemented in the new learning contexts; "0" = no implementation; "1" = partial implementation; "2" = full implementation (e.g., for both novel contexts)	Scale	0-2
postEDT_Trials	Experimental trials planned in the new learning contexts; "0" = no planning; "1" = partial planning; "2" = full planning (e.g., for both novel contexts)	Scale	0-2
postHS_EDT	Correspondence between hypotheses and experimental designs in the new learning contexts; "0" = no correspondence; "1" = partial correspondence; "2" = full correspondence (e.g., for both novel contexts)	Scale	0-2

Note: HS = Hypothesis Scratchpad; EDT = Experiment Design Tool; ILS = Inquiry Learning Space; VOTAT = Vary-One-Thing-At-a-Time.

Table 3. Mean values of variables studied, by condition and learning context

	HS and EDT	HS only	EDT only	No tool	2 Kruskal-Wallis χ^2
Learning Activity Space					
ScoreHypo	1.73	1.82	1.22	1.00	11.69**
EDT_VOTAT	1.45	0.00	2.00	0.00	35.32***
EDT_Trials	1.73	0.75	1.44	0.67	10.95^{*}
HS_EDT	1.36	0.83	0.56	0.67	$5.09^{\rm ns}$
EDT_Trials_Lab	1.73	1.08	1.22	0.67	10.03^{*}
New learning contexts					
postScoreHypo	2.00	1.42	0.67	0.78	12.90**
postEDT_VOTAT	1.55	0.00	1.33	0.00	30.57***
postEDT_Trials	1.82	0.67	0.89	0.44	13.90**
postHS_EDT	1.27	0.50	0.22	0.38	11.11^*

^{1:} HS = Hypothesis Scratchpad; EDT = Experiment Design Tool; all values range between 0 (min) and 2 (max) 2: ns = non-significant; * p < 0.05; ** p < 0.01; *** p < 0.001

