Identifying productive inquiry in an exploratory virtual labs using sequence mining

Anonymous[1]

Some institution, somewhere.

Abstract. Virtual labs are exploratory learning environments in which students learn by conducting scientific inquiry. While students often fail to learn efficiently in these environments, providing effective support is challenging since it is unclear how productive engagement looks like. This paper focuses on the mining and identification of student inquiry strategies during an unstructured activity with the D/C Circuit Construction Kit (https://phet.colorado.edu/). We use an information theoretic sequence mining method to identify productive and unproductive behaviors of a hundred students. Low domain knowledge students who successfully learned during the activity paused more after testing their circuits, particularly on simply structured circuits that target the activity's learning goals. Moreover, our results show that using pauses strategically so that they become opportunities for reflection and planning are highly associated with productive learning. Implication to theory, support, and assessment are discussed.

Keywords: inquiry learning, sequence mining, exploratory learning environments, virtual lab, self-regulated learning

1 Introduction

Learning science through inquiry has received increased attention over the last two decades[5, 8]. However, together with the push for incorporating more inquiry into the STEM curriculum, there has also been criticism towards students ability to manage their learning in exploratory learning environments (ELEs)[7, 9, 21]. Offering support for students in their inquiry requires a learner model that describes what inquiry strategies learners apply when engaging with an ELE and how productive these are in achieving the expected learning outcomes.

Our study aims to identify inquiry strategies used by students in open-ended virtual lab inquiry activities, and to evaluate their relationship with learning outcomes. The virtual lab studied is the D/C Circuit Construction Kit (CCK) which is part of the widely used Physics Education Technology (PhET) simulation suite[24]. This lab allows students to create D/C circuits with various elements such as batteries, wires, light bulbs, resistors and measure basic physical properties such as current and voltage by connecting voltmeters and ammeters to their circuits (Figure 1). Our study aims to answer the following research questions:

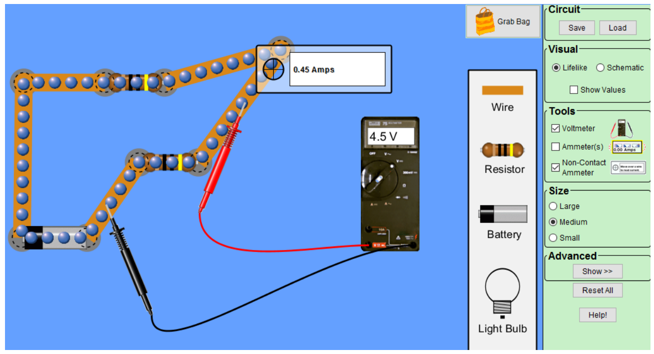


Fig. 1. A screenshot of the CCK virtual lab showing a parallel circuit with a voltmeter and ammeter.

1. What inquiry behaviors do students apply when learning with an open-ended virtual lab activity?
2. How do these behaviors change throughout the activity?
3. What productive inquiry strategies are associated with these behaviors?

We begin by situating these questions in inquiry learning theories and relevant prior work. We then describe the data collection and the information theoretic sequence-mining approach adapted from Biswas and colleagues[11]. Following the results, we discuss the identified patterns and their contribution to theory, support, and assessment of inquiry learning using virtual labs.

2 Background

Virtual labs are a class of ELEs that facilitate inquiry learning. Virtual labs and microworlds support students as they design experiments, collect data, and analyze it, with the goal of uncovering an underlying physical model[7, 9]. Virtual labs offer rich, complex, and open-ended environments that allow learners to choose from a multitude of different action paths. This makes the regulation of cognitive and metacognitive skills particularly important and challenging[6]. Studies have shown that given proper regulative support (e.g. cognitive tools), inquiry learning can have a positive impact on learning outcomes[14, 23], which gives students the unique opportunity to learn domain-specific science concepts while developing key scientific skills.

Inquiry learning is usually described in terms of the inquiry cycle, which is broadly composed of the four phases of orientation, hypothesis generation, experimentation and evaluation[13]. Njoo and de Jong specifically address two different types of learning processes that a learner uses when navigating the inquiry cycle[17]: transformative and regulative processes. Transformative processes yield knowledge directly by searching for the most appropriate hypothesis and evidence that can support or disprove it. Regulative processes encompass all metacognitive strategies that manage the execution of the transformative learning process. These include specific behaviors that are core to self-regulated inquiry learning such as planning next steps, monitoring progress and reflecting, and are key especially when only little support is offered[17].

Since student interactions with open and unstructured virtual labs are rich and complex, researchers have been leveraging automated techniques from educational data mining to more efficiently extract the key features of those interactions[2]. Many studies have focused on extracting and assessing the behaviors associated with inquiry learning processes in the context of scaffolded activities such as with cognitive tools[10] or with metacognitive support[14]. While these studies capture meaningful skills, these skills are often pre-defined and supported by the environment. Studies that have studied unstructured activities typically focus on predicting learning (with much success)[1, 4]. However, these studies often do not focus on understanding the strategies that lead to productive inquiry. Recently, more research efforts have been made towards identifying productive inquiry strategies in more open and minimally supported virtual labs [22]. For example, Bumbacher and colleagues identified three aspects of productive inquiry: focusing on target element, doing so deliberately, and seeking contrasts [3].

Despite the work presented above, our understanding of what strategies are used by learners in relatively unstructured activities, and how effective these strategies are, is lacking. Therefore, we apply an analytical framework suggested by Biswas and colleagues[11]. In their work, the authors identify common sequences and evaluate their use by different groups of learners. As this approach looks at sequences of actions, unfolded over time, in a complex environment, it seems particularity apt to our case.

As productive strategies should help learners transition from novices to experts, we look at the set of strategies used by students with low-incoming domain knowledge that eventually achieved mastery of the domain after engaging in the virtual lab activity. These "intelligent novices"[15] lack domain knowledge but are typically better at extracting meaningful feedback in open-ended learning environments and can thus achieve better learning outcomes.

3 Methods

3.1 Data collection

A hundred first year undergraduate students from a Canadian University participated in this study (one student was removed due to technical malfunctions). During the study, students completed two 25-minute activities on circuits. The first activity focused on light-bulbs and the second on resistors. Before and after the activities students completed pre- and post-tests on both topics. Analysis of identical items between pre- and post-test shows that overall students learned from the activities ($mean \pm sd$: pre: 0.47 ± 0.17; $post$: 0.62 ± 0.23; $t(96) = 6.1$, $p < 0.0005$). During the first activity on light bulbs, students were randomly assigned to a scaffolded or unstructured condition which differed in the amount of support students were given[19]. Scaffolding in the first activity had no effect on learning in the second activity[19]. Here, we focus on the second activity on the topic of resistors, as explained in the Materials section.

Pre-test score distribution is bimodal, which we use to assign students to either the low or high incoming knowledge group with 74 and 25 students respectively (low knowledge z-score: -0.5 ± 0.6; high knowledge z-score: 1.5 ± 0.5). The 25 students with high incoming knowledge had high post test scores and, being experts, they were excluded from analysis. To identify productive inquiry strategies, we applied a median split to the post-test, distinguishing between students who performed well on the post-test ("intelligent novices"; Low-High, or LH, with 38 students) as compared to those that did poorly on both tests (Low-Low , or LL, with 36 students). The pre-test did not correlate significantly with post-test scores for low pre-test students ($r_s = 0.06$, $p = 0.63$). Within these groups, an even number of students were in the scaffolded condition in the first activity: 17 out of 36 LL and 21 out of 38 LH.

3.2 Materials

Students used the CCK lab shown in Figure 1 for 25 minutes. Students received minimal guidance in the form of an activity sheet with the following learning goals: to understand what happens to the current and voltage of a circuit when 1) the resistance of a single resistor is changed, 2) when multiple resistors with different resistances are arranged 3) when multiple resistors are connected in a circuit in a variety of arrangements. The post-test assessed conceptual understanding of resistors in circuits. The post-test was a reliable measure of student's knowledge with a Cronbach alpha of 0.75.

3.3 Sequence mining

Our first step was to choose an appropriate representation when parsing the log data. We applied top-down approaches such as think-aloud protocols and expert cognitive tasks analysis, as well as bottom-up approaches such as visualizing log data of individual students. For example, retrospective think-aloud protocols showed that many students had periods of inactivity during which students either thought back to reflect (e.g, while writing down conclusions about experiments), or thought forward to plan (e.g, what circuit to build or test next). We thus defined pauses as non-activity for more than 15 seconds, based on the long tail of the distribution of time between actions. Changes to this threshold have nearly no effect on relative frequencies of actions.

We then categorized individual log actions to action families (Figure 2). Specifically, we defined two broad categories: Construct actions and Test actions. Construct actions encompass all actions that change the circuit configuration such as adding or connecting batteries and resistors. Test actions encompass all actions related to connecting a voltmeter or ammeter to measure the circuit. Other actions such as those relating to the interface (e.g, zooming) were ignored.

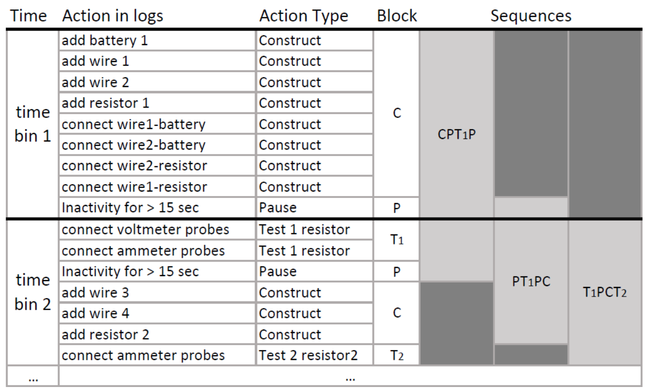


Fig. 2. Log data was parse by action type, to blocks of action, to sequences. Here, sequences of 4 blocks are parsed; two occurring in time bin 1 and one in time bin 2.

Within the lab there are many ways to achieve a certain effect. For example, students can build the same circuit in any order of actions. Given that we are mainly interested in the final configuration of the circuit to be tested, and less so in the steps that students took to construct it, we collapsed individual actions of the same type into blocks. Merging actions of the same type into a block reduces the granularity of the data and facilitates the interpretation of sequences.

Analysis of trace data often focuses on the events alone. However, in this context (as in many other ELEs'), the state of the virtual lab affects the essence of the action[16]. Given the activity's learning goals we classified the circuits that were tested as one of three kinds: circuits with only one resistor (T_1), circuits with two resistors (connected in parallel or in series; T_2), and circuits with any other configuration (T_m).

3.4 Mining interesting sequences of actions

In order to identify potential strategies employed by students, we break down their activity into sequence patterns of two to ten blocks of actions long (Figure 2). We limited the length to ten blocks in order to focus on common sequences. We then collected all the different possible sequence patterns done by students ($\sim 10,454$ different sequence patterns). Only sequences that were used at least once by a minimum of 35% of LL or LH students were examined. The results thus obtained are not sensitive to the value of this threshold within a range of $\sim 10\%$. A higher threshold yielded sequences that are typically very short (< 3 blocks) and used more homogeneously by many students while a lower threshold yields long sequences (> 6 blocks) that are highly distinctive across groups but that were used by very few students.

To associate between sequences and learning, we implemented the Differential Temporal Interestingness of Patterns in Sequences outlined by Biswas and colleagues[11]. This measure uses an adaptation of a traditional information gain measure to rank sequences in terms of their differential use over time, by groups, or both. The obtained rankings serve as a guide to reduce the number of sequence patterns to evaluate and interpret. To do so, we 1) bin student's activity into four bins of equal time length (determined on a per student basis), 2) determine in which time bin a sequence pattern was used based on the time bin (Figure 2), 3) count the number of students that used it at least once per group per time bin, and 4) score each sequence based on the added information it provides given how it was used over time and by groups and compared to being used evenly[11].

After following this sequence mining and ranking procedure we obtained three rankings of 119 sequences. When comparing the use of sequences by learners we report a Pearson χ^2 test. When examining whether specific building blocks of sequences were used more frequently by one group, we report their median (mdn) or mean frequency ($\mu \pm \sigma$) and compare them using a one-tailed Mann-Whitney (U; corrected for ties) or Student's T-test (t), respectively, depending on the normality of the distributions of frequency of use of the building block.

4 Results

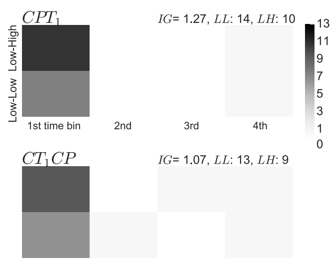
Within the activity, students performed over 400 of individual actions each, grouped into 47 ± 16 blocks of actions. We note that all students used Pause, Construct and Test actions more than once throughout the activity. Notably, both groups used Pauses with the same frequency (LL:0.29±0.08, LH:0.31±0.08)

Sequences that differentiate time in the activity Figure 3A shows two sequences that had a significant amount of information gain with respect to time. Other sequences that ranked highly were quite similar in that they included testing of a basic single resistor circuit (T_1) and were used by both groups of students in the first or second quarter of the activity.

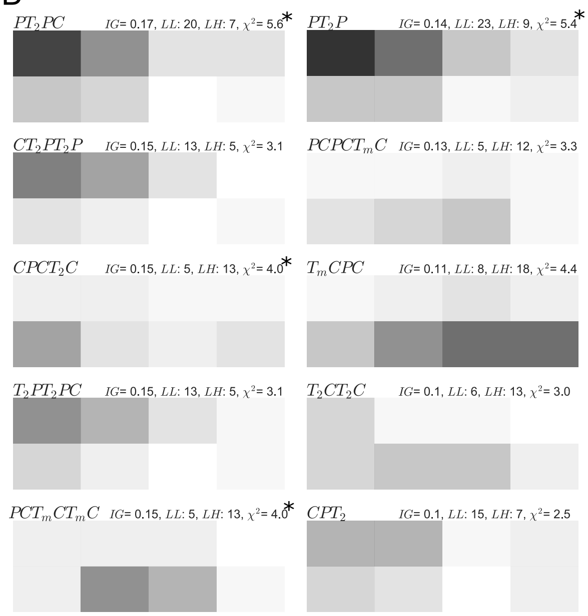
Sequences that differentiate learner groups Figure 3B shows the top ten sequences that have the most significant amount of information gain with respect to differentiating LL and LH. The number of students who used four of these sequences is statistically significantly different across the two groups: more LH do sequences of the form PT_2P, $\chi^2 = 5.4 (p = 0.02)$ and PT_2PC, $\chi^2 = 5.6 (p = 0.02)$ while more LL do sequences of the form $CPCT_2C$, $\chi^2 = 4.0 (p < 0.05)$ and PCT_mCT_mC, $\chi^2 = 4.0 (p < 0.05)$. LH use PT_2P and PT_2PC primarily early on the in the activity; LL use the sequence $CPCT_2C$ early in the activity and use the sequence PCT_mCT_mC mid activity (Figure 3B).

Sequences that differentiates by both groups and time Sequences obtained with the criteria of highest information gain with respect to the interaction of group and time were overwhelmingly dominated by the same sequences obtained with respect to time only.

A



B



C

	Previous T	Previous C
Subsequent T	TPT	CPT
C	TPC	CPC

Low-Low

	Previous T	Previous C
Subsequent T	2.5 ± 2.2	1.7 ± 1.5
C	3.8 ± 2.0	5.5 ± 3.7

Low-High

	Previous T	Previous C
Subsequent T	2.9 ± 2.3	1.8 ± 1.5
C	4.5 ± 2.1	3.5 ± 2.5

Fig. 3. A) The temporal use of sequences which ranked highly in their information gain (IG) by time. B) The temporal use of sequences which ranked highly in their IG by group. A * indicates which sequence had a Pearson χ^2 test of homogeneity between groups with a significance level of at least $p < 0.05$. C) Mean and standard deviation of counts of Pauses given the previous and subsequent action for LL and LH groups.

Building blocks All sequences with high information gain across groups are composed of common building blocks that are unique to each group. To better understand the nature of the inquiry strategies these sequences allude to, we first examine their building blocks. Of particular interest, are the T_2P building block from LH sequences and the T_mC and T_2C building block from LL sequences.

We test whether these building blocks were used with a higher frequency by one of the two groups. LH perform the unit T_2P more often than LL (LL: mdn= 0.03; LH: mdn= 0.06; $U = 848.5$, $p = 0.04$). Controlling for the use of T_2, LH also followed a T_2 action with a Pause more often than LL. (LL: mdn= 0.20; LH: mdn= 0.56; $U = 919.0$, $p < 0.01$). This also applies more generally to any testing action T. LH use TP more overall: LL: 0.40 ± 0.20; LH: 0.53 ± 0.19; $t(35) = 2.86$, $p < 0.01$. Controlling for T, LH Pause more after any Test block: LL: 0.25 ± 0.14; LH: 0.34 ± 0.12; $t(35) = 2.60$, $p = 0.01$).

LL do not perform T_2C more (LL: mdn= 0.06; LH: mdn= 0.05) nor do they Construct more after T_2 (LL: mdn= 0.33; LH: mdn= 0.25). However, LL perform the T_mC unit more frequently than LH (LL: mdn= 0.17; LH: mdn= 0.11; $U = 508.5$, $p = 0.03$). LL Construct more after T_m also when controlling for frequency of T_m (LL: mdn= 0.57; LH: mdn= 0.39; $U = 440.5$, $p < 0.01$).

However, LL do not use TC more overall (LL: 0.35 ± 0.16; LH: 0.28 ± 0.15) however, controlling for frequency of T, LL do Construct more after a Test block (LL: 0.56 ± 0.19; LH: 0.42 ± 0.20; $t(35) = 2.60$, $p < 0.01$).

Evaluating pauses Pausing after testing was found to be very common among LH, yet not among LL. Evaluating whether Pauses are used strategically before or after certain actions may explain whether students are using pauses as opportunities to reflect or plan and predict. Figure 3C shows the frequency with which each group uses Pauses as a function of the previous and subsequent action (Test or Construct). We evaluate the dependency of pauses on their context by running a repeated measures ANOVA with frequency of pauses as a dependent variable, and three factors as independent variables: the previous action, the subsequent action and their interaction.

For the LH group, there was a strong main effect for subsequent action, where LH Pause more before Construct than before Test: $F(1, 37) = 22.10$, $p < 0.01$, $\eta^2 = 0.37$. There was also a significant main effect for the previous action, where LH students Pause significantly more after Test than after Construct: $F(1, 37) = 5.91$, $p = 0.02$, $\eta^2 = 0.14$. There was no significant interaction. Taken together, LH pause both before constructing and after testing, a strategic use of pauses to plan the subsequent circuit and reflect on the previous test.

For LL, there was a main effect on the subsequent action, where LL Pause more before Construct than before Test, $F(1, 35) = 37.80$, $p < 0.001$, $\eta^2 = 0.51$. However, there was no significant effect for the previous action, $F(1, 35) = 0.90$, $p = 0.35$, $\eta^2 = 0.25$. There was also a significant interaction, $F(1, 35) = 9.43$, $p < 0.01$, $\eta^2 = 0.21$; given that Pause before Construct are always more frequent than Pause before Test, interpreting the main effects is possible. The interaction shows that LL pause before constructing mainly if their previous action was also Construct, as opposed to Test (CPC sequence as opposed to CPT). That is, overall LL do not pause much in relation to testing but mainly do so before a Construct action, and most frequently in the process of building (i.e. between successive Construct blocks).

5 Discussion

Results presented above highlight several differences in the way in which LL and LH learners engaged with the activity. First, each novice learner groups applied different sequences. LH students engaged in repeated testing-pausing cycles (e.g., T_2PT_2PC and CT_2PT_2P). This strategy is evocative of an efficient transformative inquiry process whereby the students iteratively test and reflect. Importantly, this process was performed by LH students on simpler series and/or parallel two resistors circuit, which are amongst the most useful configurations to test to acquire domain knowledge on basic D/C circuits. Thus, intelligent novices were able to infer more meaningfully by pursuing simpler experiments, ones that match their zone of proximal development. The fact that they did so iteratively suggests that these students applied a compare-and-contrast strategy, where they compared two elements of the same circuit or two elements across different circuits (since they can build more than 1 circuit at a time). This is an effective inductive reasoning strategy in building a mental framework necessary to explain an unknown model[20]. In contrast, LL students' sequences involve building circuits and subsequently measuring the properties of these circuits without pausing. They also do so iteratively, and often on complex circuits, as suggested by the sequences: T_2CT_2C and PCT_mCT_mC. Their greater focus on complex structures that are likely beyond their zone of proximal development likely had a negative impact on their ability to infer meaningful relationships from the learning environment.

Second, learner groups behave differently in the way they pause. Notably, both groups pause the same number of times given the number of action blocks in their activity - thus, it is not the frequency of pausing, but rather their context that is associated productive inquiry. Intelligent novices apply pauses in two main scenarios: after testing and before constructing. Pausing after testing is done possibly to reflect on the result, evaluate a new observation, and validate/refine an existing hypothesis, or, having inferred new knowledge from a test, create a new hypothesis. Pausing before constructing is likely done to plan for the next experiment. Novices that had a less productive inquiry did not pause much in relation to testing actions and mostly did so before Construction. The strategic use of pauses during the inquiry process of novices is thus likely to play an important role in learning form the activity.

Interestingly, lab manuals often instruct students to build simple experimental setups, test iteratively, take notes, reflect on past experiments, and plan new ones. These are the same patterns that were found productive in the analysis presented above. Yet, to the best of our knowledge, this is the first time that these strategies are associated with intelligent novice behaviors when no such support is given. Moreover, given that support was provided randomly to half the students in the first activity on light bulbs and that learning gains from this scaffold did not transfer to learning gains after the second activity, it seems that supporting these strategies in a worksheet form is ineffective for teaching these strategies and supporting transfer. Adaptive scaffolding that provides feedback in context is likely to have a greater impact on learning [10].

The analysis also found a simple pattern that changes in time: students test the simplest circuits first. Testing such circuits makes sense for these low domain knowledge students, particularly early on in the activity (Figure 3A). Doing so may be a combination of metacognitive awareness (recognizing one's limitations) and being able to recognize informative configurations in a new domain.

This result suggests that effective strategy use has time dependency. Why was this not observed in our analysis beyond the first five minutes? It may be that the first five minutes of problem set up, are the most distinct in the learning process. Also, it may be that afterwards, learners disperse in their behaviors in a manner that does not converge to common patterns that emerge in the sequence mining method used here. As the first time bin was vastly different from the other, the most common patterns with time dependency occur during this bin.

An analysis that begins after this bin may reveal more subtle patterns with time dependency.

Finally, it is important to emphasize that the sum of the identified patterns is larger than its parts. While each of the building blocks makes sense on its own, it is their combination that constitutes powerful strategies. For instance, it is not the individual act of pausing and testing on simple circuits that creates a powerful learning strategy, but the recurring process of pausing after the test action block. These results show a striking resemblance to the findings of Bumbacher and colleagues, which associated learning with seeking contrasts (similar to the repetitions in the current study), addressing the target knowledge (shown in the circuit configuration here), and doing so deliberately (shown as pauses here) [3].

This work leaves three important questions unanswered. First, how do high incoming knowledge students engage with the environment? That is, are experts similar to intelligent novices, or do they show different patterns? In fact, it may be that their expertise allows them to benefit from patterns that are used by LL, such as testing complex circuits. If that is the case, then perhaps LL apply generally good strategies, but fail to adapt their strategies to their low incoming knowledge, in a form of the Kruger-Dunning effect (unskilled and unaware)[12].

A second question is about the generality of these patterns. How specific are they to this virtual lab, activity, and population? We kept our knowledge representation rather domain-general, to support comparison to other simulations, and look to replicate this work in other contexts.

Lastly, this work shows association, not causation. It would of interest to evaluate the impact of supporting the identified productive patterns. What are the best ways to do so? Can students be taught these skills? Prior evidence shows that following suggestions does not necessary carry the benefits with it, as students merely do the motions[18]. For example, it is much easier to make someone pause then reflect.

6 Summary

We looked at two groups of learners who had low domain knowledge when entering an activity on DC circuits using a virtual lab: those who learned well in the activity (intelligent novices) and those who failed to learn. We then applied sequence mining to identify what differentiates between the two groups, showing that students who learned well tested simpler circuits, used pauses strategically, and did so iteratively. This work has implications to theories of learning, by revealing how learning in virtual labs unfolds. It also has implications to instructional design, suggesting strategies that should be the focus of instruction on inquiry skills. Most intriguingly, its granularity allows us to evaluate learning as it happens which is an important step towards offering learners adaptive feedback based on their strategy and creating smarter simulations that help learners become better scientists.

Bibliography

[1] R. S. Baker, J. Clarke-Midura, and J. Ocumpaugh. Towards general models of effective science inquiry in virtual performance assessments. *Journal of Computer Assisted Learning*, 32(3):267–280, 2016.

[2] M. Berland, R. S. Baker, and P. Blikstein. Educational data mining and learning analytics: Applications to constructionist research. *Technology, Knowledge and Learning*, 19(1-2):205–220, 2014.

[3] E. Bumbacher, S. Salehi, M. Wierzchula, and P. Blikstein. Learning Environments and Inquiry Behaviors in Science Inquiry Learning: How their Interplay Affects the Development of Conceptual Understanding in Physics. *International Educational Data Mining Society.*, pages 61–68, 2015.

[4] C. Conati, L. Fratamico, S. Kardan, and I. Roll. Comparing representations for learner models in interactive simulations. *Lecture Notes in Computer Science*, 9112:74–83, 2015.

[5] N. R. Council. *Inquiry and the national science education standards: A guide for teaching and learning.* National Academies Press, 2000.

[6] T. De Jong. Technological advances in inquiry learning. *Science*, 312:532–533, 2006.

[7] T. De Jong and W. R. Van Joolingen. Scientific Discovery Learning with Computer Simulations of Conceptual Domains. *Review of Educational Research*, 68(2):179–201, 1998.

[8] European Commission. A Renewed Pedagogy for the Future of Europe. pages 1–29, 2007.

[9] X. Fan and D. Geelan. Enhancing Students' Scientific Literacy In Science Education Using Interactive Simulations: A Critical Literature Review. *Journal of Computers in Mathematics and Science Teaching*, 32(2): 125–171, 2013.

[10] J. D. Gobert, M. Sao Pedro, J. Raziuddin, and R. S. Baker. From Log Files to Assessment Metrics: Measuring Students' Science Inquiry Skills Using Educational Data Mining. *Journal of the Learning Sciences*, (4):521–563.

[11] D. M. John Kinnebrew and G. Biswas. Mining Temporally-Interesting Learning Behavior Patterns. *Proceedings of the 6th International Conference on Educational Data Mining*, pages 252–25, 2013.

[12] M. Krajc and A. Ortmann. Are the unskilled really that unaware? an alternative explanation. *Journal of Economic Psychology*, 29(5):724–738, 2008.

[13] S. Löhner, W. R. Van Joolingen, E. R. Savelsbergh, and B. Van Hout-Wolters. Students' reasoning during modeling in an inquiry learning environment. *Computers in Human Behavior*, 21(3 SPEC. ISS.):441–461, 2005.

[14] S. Manlove, A. W. Lazonder, and T. de Jong. Trends and issues of regulative support use during inquiry learning: Patterns from three studies. *Computers in Human Behavior*, 25(4):795–803, 2009.

[15] S. A. Mathan and K. R. Koedinger. Fostering the Intelligent Novice: Learning from errors with metacognitive tutoring. *Educational Psychologist*, (4): 257–265.

[16] A. Mitrovic and P. Suraweera. Teaching database design with constraint-based tutors. *International Journal of Artificial Intelligence in Education*, 26(1):448–456, 2016. ISSN 1560-4306. doi: 10.1007/s40593-015-0084-6. URL http://dx.doi.org/10.1007/s40593-015-0084-6.

[17] M. Njoo and T. De Jong. Exploratory Learning with a Computer Simulation for Control Theory: Learning Processes and Instructional Support. *Journal of Research in Science Teaching*, 30(5):821–844, 1993.

[18] I. Roll, R. Baker, V. Aleven, and K. R. Koedinger. On the benefits of seeking (and avoiding) help in online problem solving environment. journal of the learning sciences. 23:4:537–560, 2014.

[19] I. Roll, N. Yee, and A. Cervantes. Not a magic bullet: the effect of scaffolding on knowledge and attitudes in online simulations. In *International Conference of the Learning Sciences*, pages 879–886, 2014.

[20] J. T. Shemwell, C. C. Chase, and D. L. Schwartz. Seeking the general explanation: A test of inductive activities for learning and transfer. *Journal of Research in Science Teaching*, 52(1):58–83, 2015.

[21] S. Tobias and T. M. Duffy. *Constructivist instruction: Success or failure?* Routledge, 2009.

[22] O. Uzan, R. Dekel, O. Seri, et al. Plan recognition for exploratory learning environments using interleaved temporal search. *AI Magazine*, 36(2):10–21, 2015.

[23] W. van Joolingen and T. de Jong. *Model-Based Diagnosis for Regulative Support in Inquiry Learning*, pages 589–600. Springer New York, New York, NY, 2013.

[24] C. E. Wieman, W. K. Adams, and K. K. Perkins. PhET: Simulations that enhance learning. *Science*, (October):682–683.