Exploring Middle School Students' Science Learning and Discourse in Physical and Virtual Labs

Dana Gnesdilow, Nafsaniath Fathema, Feng Lin, Seokmin Kang, Catherine Dornfeld, and Sadhana Puntambekar
gnesdilow@wisc.edu, fathema@wisc.edu, feng.lin@wisc.edu, skang79@wisc.edu, cldornfeld@wisc.edu, puntambekar@education.wisc.edu
University of Wisconsin—Madison

Abstract: This study compared middle school students' science learning and discourse patterns in physical and virtual labs. Students in virtual labs had significantly higher learning gains than students in the physical labs on a pre and post-test content knowledge test. To understand the difference in learning gains, we analyzed the discourse of the students and teacher as they worked in groups during the labs. We found three main categories of discourse: conceptual talk, task-based talk, and teacher scaffolding. Our analysis of discourse showed that students in the virtual group engaged in more conceptual talk than the physical group, while the physical group had more task-based talk. We also found the teacher provided more conceptual scaffolding during virtual labs than physical labs.

Introduction

Despite the increased interest in researching the differential benefits and limitations of physical and virtual labs, there have been conflicting results about which mode of experimentation better supports students' science learning. Some studies have found that virtual labs better supported students' learning (e.g., Pyatt & Sims, 2012; Zacharia, Olympiou, & Papaevripidou, 2008), while others have found that physical labs helped students to learn more (e.g., Marshall & Young, 2006; Smith & Puntambekar, 2010). Other studies have indicated that there were no significant differences in students' learning from physical versus virtual labs (e.g., Pyatt & Sims, 2012; Zacharia & Olympiou, 2011). In this study, we explored how physical and virtual labs might influence students' science learning and students' and teacher's discourse in middle school science classes.

There is evidence that physical and virtual labs support students' learning in different ways. For example, virtual labs can combine multiple representations, possibly facilitating a deeper understanding of underlying concepts and providing visualizations of processes that cannot be seen during a physical experiment (Ainsworth, 2006). Virtual labs may support learning by offloading error-prone calculations and eliminating measurement error, leading to "cleaner" data, allowing students to explore conditions, such as zero friction or gravity (Zacharia & Anderson, 2003) and to run multiple trials under different conditions and explore variety of ideas in a shorter period of time. Alternatively, even though setting up and conducting physical labs can be time consuming, many researchers assert that manipulating real materials and receiving haptic feedback is essential for learning. According to theories of embodied cognition, conceptual processing is influenced by our movements, bodily states, and use of physical manipulatives cognition (e.g., Glenberg, Brown, & Levin, 2007). Yet, there are many unanswered questions about the learning benefits of physical and virtual experimentation in fostering deep learning of science concepts. To date, most studies have compared students' science conceptual learning via pre and post-tests, and only a few studies have examined classroom discourse to understand how students and teachers interacted when engaging in physical or virtual experiments (e.g, Zacharia & de Jong, 2014; Marshall & Young, 2006). We propose that a full examination of discourse patterns generated by students and teachers during physical and virtual labs will help us better understand the unique affordance of each type of lab. Our research questions for the study were: 1) Does participation in physical versus virtual experimentation result in different learning gains for middle school science students? and 2) Does engagement in physical versus virtual experimentation promote different kinds of discourse? If so, what are the differences?

Methods

Participants and instructional context

Forty-one 6th graders from two science classes, taught by the same teacher, at a Mid-Western middle school in 2015 participated in this study. Students in both classes were engaged in a design-based science curriculum to learn about inclined planes and pulleys (Puntambekar, Stylianou, & Goldstein, 2007). Each class was randomly assigned into one of two conditions: 1) Physical Labs (PL) (N = 19), 2) Virtual Labs (VL) (N = 22).

Sequence of the learning

In both the PL and VL conditions, students: 1) took the pre-test, 2) engaged in research and experiments about pulley, 3) engaged in research and experiments about the inclined plane, and then 4) took a post-test. The pulley and inclined plane units were taught back-to-back. Both units focused on how simple machines can help reduce the force needed to lift an object by increasing mechanical advantage. Students also learned about other concepts, such as potential energy and friction. In PLs students learned about simple machines by setting up and using real pulleys and inclined planes. Students also needed to do calculations to determine values of their dependent variables such as potential energy, ideal mechanical advantage, and actual mechanical advantage. In the VLs students did the same experiments virtually. VL students also explored additional pulley configurations, such as movable and quadruple compound pulleys and carried out idealized experiments with no friction.

Data sources and analysis

Pre and post test measures

The pre test and post test were identical. The test was composed of 21 multiple choice questions and assessed students' knowledge of physics concepts related to pulleys and inclined planes. Student earned one point for each correct answer. The maximum score was 21 points.

Analysis of group discourse

To answer if there were differences between the students' talk while performing physical versus virtual labs, we analyzed the discourse of a subset of five student groups: three from PL conditions and two from VL conditions. There were three to four students in each group. This subset was chosen because we had audio recordings for two pulley and one inclined plane experiment/s for all groups. We segmented the transcripts by turns of talk as the unit of analysis which was multiply coded. There were 2,819 total turns of talk when combining all transcripts. To see if engagement in the PL versus VL condition promoted different discourse patterns, we developed a coding scheme. We inductively identified four macro categories of discourse that the students and the teacher were engaged in during the labs (see Table 1). Our code development was influenced also by Zacharia and de Jong's (2014) work, who found that students doing physical labs often engaged in more procedural versus conceptual talk. Two coders independently coded 25% of the transcripts. An inter-rater reliability of approximately 96% was established after resolving disagreements through discussion. The remainder of the transcripts were coded separately by the same coders.

Table 1: Discourse codes and examples

Macro Category	Micro Code and Definition
Conceptual	**Science Questions:** questions about science relationships or connections to the real world
	Science Explanation: explaining science concepts and relationships or evaluating / making sense of data
	Science Predictions / Patterns: predicting how changing variables affects others / seeing patterns in data
Task-Based	**Reporting and Calculating:** talk about calculating or writing data into charts
	Setting up experiments: talk about manipulating tools, running trials, and the progress of the lab
	Fact Based Questions: questions about how to proceed with lab
Not Applicable	**Not Applicable**: Off task or inaudible
Teacher Scaffolding Talk	**Conceptual Scaffolding**: encouraging students to think about the science concepts and relationships
	Task-Based Scaffolding: helping student to complete tasks, not the science

Findings

Comparison of students' learning gains in physical and virtual conditions

To answer our first research question, we conducted a Mann-Whitney U test to see if there were differences in the learning gains of students in the PL and VL conditions. To calculate learning gains, we divided the actual gain each student made by the total possible gain he or she could have made from pre to post-test. We found that there was a statistically significant difference in the learning gains made by the students in the PL versus VL conditions; Students in the VL condition made significantly greater learning gains than in the PL condition, $z = -2.395$, $p < 0.05$, VL mean rank = 25.82 (55% average learning gain), PL mean rank = 16.84 (36% average learning gain).

Examination of the differences in discourse between physical and virtual conditions

To better understand what may have influenced VL students' significantly higher learning gains, we compared the discourse that occurred in groups by condition. Out of the total 2,819 total turns of talk in all transcripts, we removed 1,164 turns of talk that had been coded as '*not applicable'*. Of the remaining 1,655 turns of talk, 901 were students' and 57 were the teacher's. We then calculated the proportion of students' talk in each macro and micro coding category. To do so, we divided the frequency of each type of talk students were engaged in by the total number of students' turns of talk by condition. For the analysis of teacher's scaffolding, we divided the frequency of the teacher's turns of talk in each category by the total number of teacher's turn of talk during each condition. Due to the number of statistical comparisons conducted, we used a Bonferroni correction for multiple comparisons to minimize Type I error. This means that all significant results (*) were based on a $p < .0005$ significance level.

Results from the chi-squared test on our macro codes showed that students in the VL condition had a significantly higher proportion of conceptual talk (0.12) than the students in PL conditions (0.04) (z = 5.0847*). In terms of task-based talk, we found that students in the PL conditions had a significantly higher proportion of task-based talk (0.96) than those in the VL condition (0.88) (z = -5.0659*). These results indicated that even though students in both conditions were engaged in more task-based talk, the students in the VL condition were engaged in significantly more conceptual talk than the students in the PL condition.

For our chi-squared analysis of the conceptual talk micro codes, we found that students in VL condition had a significantly higher proportion of science explanation talk (0.06) and science predictions and patterns talk (0.05) than the students in PL condition (0.03 (z = 2.8226*) and 0.01 (z = 3.9527*), respectively). No statistically significant difference was found between the proportions of science questions asked. For the task-based micro-codes, we found statistically significant differences between the proportions of talk by condition. Students in the PL condition had a significantly higher proportion of talk related to reporting and calculating (0.41) than students in the VL condition (0.20) (z = -9.1231*). On the other hand, we found that students in the VL condition had a significantly higher proportion of talk related to setting-up experiments (0.41) as compared to those in the PL condition (0.34) (z = 2.9384*). No significant difference in the proportion of fact-based questions asked was found between conditions. These results indicated that students in the VL condition had a significantly higher proportion of talk discussing science ideas, making predictions and looking at patterns in data, and discussing setting up experiments than the PL students. Alternatively, the students in the PL condition spent a greater proportion of their talk on reporting and calculating.

Finally, we ran a chi-squared test of homogeneity of proportions to analyze the teachers' scaffolding talk and found the teacher was engaged in a greater proportion of conceptual scaffolding when working with the VL students (0.33) than PL students (0.04) (z = 4.6458*). In terms of task-based scaffolding, the teacher was engaged in a greater proportion of task-based scaffolding talk with the PL students (0.96) than with VL students (0.67) (z = -4.6458*).

Conclusions and implications

The goals of our study were: 1) to examine and compare the effects of virtual and physical labs on students' learning gains and 2) to examine students' and teacher's discourse to shed light on how the context of learning in the different conditions may have influenced science learning. In term of our first question, we found that students in the VL condition had significantly higher learning gains than the students in the PL condition. This finding is in line with prior studies showing that students who participated in virtual labs did better than those who participated in physical ones (such as, Pyatt & Sims, 2012; Zacharia, et.al., 2008). For our second research question, we identified two major categories of talk: conceptual and task-based talk. Our discourse analysis showed that VL students had a greater proportion of conceptual talk than those in PLs. Students in the PL condition had a significantly higher proportion of task-based talk. These findings align with Zacharia and de Jong's (2014) analysis that found that students in physical conditions talked more about procedural aspects of labs, rather than building conceptual understanding. It is possible that the process-oriented problems in physical labs restricted students from developing better conceptual understanding. These findings helped us understand how discourse factors may have caused different learning gains of students in the two conditions.

To further tease out the differences between the types of talk *students* in each condition were engaged in, we created micro-coding categories. For conceptual talk, students in the VLs had a significantly higher proportion of science explanations and talk about science predictions and patterns, than PL students. In terms of task-based talk, we found that students in the VL group had a significantly higher proportion of discourse related to setting up experiments, while students in PLs had a significantly higher proportion of talk about reporting and calculating. Even though VL students were engaged in more talk about setting up experiments, the simulation provided outputs for dependent variables and students did not need to take so much time for doing calculations.

Thus, they had more time to engage in science talk such as making sense of data, identifying patterns, explaining, and making predictions.

In terms of the teacher's talk, we identified that the *teacher*'s scaffolding discourse differed between the two conditions. The teacher had a significantly higher proportion of conceptual scaffolding during the VL condition. It may be the case that in the VL condition the teacher could address students' conceptual learning needs as he did not need to help students set up physical materials. This may also be why the students in the virtual labs had a higher proportion of conceptual talk than the students in the physical labs. Thus, not only were the students in the PLs less able to explore conceptual ideas due to setting up materials and doing calculations, when compared to VL students, they also received almost no conceptual scaffolding to help them to think about the science concepts and relationships. So in addition to all of the benefits of using virtual labs, students in the VL condition also had more time to process and think about the data they were collecting, and received more conceptual scaffolding from the teacher. These findings are in contrast with Marshall and Young's (2006) study, which reported that undergraduate students had difficulty processing information from a simulation. However, these differences might be explained by the fact that undergraduate students may have an easier time negotiating and setting up physical materials than sixth graders.

Our study contributes to the literature by identifying and elaborating on students' learning processes and discourse patterns. We developed codes to characterize students' and the teacher's discourse and systematically examined classroom discourse with a mixed-methods approach, which provided a richer understanding of the learning process. Our results indicated that teacher's scaffolding seems to be influenced by the context of learning and the needs of the students. Our findings suggest that teachers should be reflective about their practice to ensure that conceptual scaffolding is always provided to students, no matter the type of lab. Since our results are based on a small sample of students from only two classes and one teacher, replication of this study in a larger setting with more teachers would help improve the power and strengthen the validity of our claims. We suggest that studies in this area should move beyond only examining students' learning on content-based tests. A more detailed account of how different forms of experimentation influence students' learning and discourse is necessary to extract commonalities across different studies.

References

Ainsworth, S. (2006). DeFT: A conceptual framework for considering learning with multiple representations. *Learning and Instruction*, *16*(3), 183-198.

Glenberg, A. M., Brown, M., & Levin, J. R. (2007). Enhancing comprehension in small reading groups using a manipulation strategy. *Contemporary Educational Psychology*, *32*(3), 389-399.

Marshall, J. A., & Young, E. S. (2006). Preservice teachers' theory development in physical and simulated environments. *Journal of Research in Science Teaching*, *43*(9), 907-937.

Pyatt, K., & Sims, R. (2012). Virtual and physical experimentation in inquiry-based science labs: Attitudes, performance and access. *Journal of Science Education and Technology*, *21*(1), 133-147.

Puntambekar, S., Stylianou, A., & Goldstein, J. (2007). Comparing classroom enactments of an inquiry curriculum: Lessons learned from two teachers. *The Journal of the Learning Sciences, 16*(1), 81-130.

Smith, G. W., & Puntambekar, S. (2010). Examining the combination of physical and virtual experiments in an inquiry science classroom. In *Proceedings of the Conference on Computer Based Learning in Science, Warsaw, Poland*.

Zacharia, Z., & Anderson, O. R. (2003). The effects of an interactive computer-based simulation prior to performing a laboratory inquiry-based experiment on students' conceptual understanding of physics. *American Journal of Physics*, *71*(6), 618-629.

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., & Olympiou, G. (2011). Physical versus virtual manipulative experimentation in physics learning. *Learning and Instruction*, *21*(3), 317-331.

Zacharia, Z. C., Olympiou, G., & Papaevripidou, M. (2008). Effects of experimenting with physical and virtual manipulatives on students' conceptual understanding in heat and temperature. *Journal of Research in Science Teaching*, *45*(9), 1021-1035.

Acknowledgments

We thank the participating students and the teacher. This research has been supported by an ECR grant from the National Science Foundation (Grant # 1431904).