Scaffolding during Science Inquiry

Haiying Li
Rutgers University
New Brunswick, NJ, USA
haiying.li@gse.rutgers.edu

Janice Gobert
Rutgers University
New Brunswick, NJ, USA
Janice.gobert@gse.rutgers.edu

Rachel Dickler
Rutgers University
New Brunswick, NJ, USA
rachel.dickler@gse.rutgers.edu

ABSTRACT

Prior studies on scaffolding for investigative inquiry practices (i.e. forming a question/hypothesis, collecting data, and analyzing and interpreting data [21]) revealed that students who received scaffolding were better able to both learn practices and transfer these competencies to new topics than were students who did not receive scaffolding. Prior studies have also shown that after removing scaffolding, students continued to demonstrate improved inquiry performance on a variety of practices across new driving questions over time. However, studies have not examined the relationship between the *amount of scaffolding received* and transfer of inquiry performance; this is the focus of the present study. 107 middle school students completed four virtual lab activities (i.e. driving questions) in Inq-ITS. Students received scaffolding when needed from an animated pedagogical computer agent for the first three driving questions for the Animal Cell virtual lab. Then they completed the fourth driving question without access to scaffolding in a different topic, Plant Cell. Results showed that students' performances increased even with fewer scaffolds for the inquiry practices of hypothesizing, collecting data, interpreting data, and warranting claims; furthermore, these results were robust as evidenced by the finding that students required less scaffolding as they completed subsequent inquiry activities. These data provide evidence of near and far transfer as a result of adaptive scaffolding of science inquiry practices.

Author Keywords

Scalability; science inquiry; inquiry practices; scaffolding.

INTRODUCTION

The Next Generation Science Standards (NGSS; [21])were designed with three foci, namely, disciplinary core ideas, crosscutting concepts, and inquiry practices. In terms of the inquiry practices, some practices can be categorized as the "doing" science inquiry (e.g. forming a question/hypothesis, collecting data, and analyzing and interpreting data; we refer to these as investigating). These practices can be challenging for students to engage in without scaffolded supports [9, 10, 18].

Scaffolding and Inquiry Practices

Scaffolds are supports provided to students in order to assist them in carrying out a task that is too difficult for them to complete independently [29]. Researchers have integrated scaffolds into inquiry environments for the practices involved in investigating [5, 16, 27]. Specifically, researchers have integrated scaffolds to support students on hypothesizing and planning investigations [16, 24, 25, 28], carrying out investigations/collecting data [16, 24, 25], and analyzing and interpreting data [19, 20, 27]. The types of scaffolds provided to students for these inquiry practices included visual adaptations to a task [27], explicit guidance provided to students [28], and individualized hints in the form of pop-ups in online environments [5].

The various types of scaffolds provided to support investigative inquiry practices have generally been shown to improve student performance on the practices on which they were helped [5, 27]. The visual scaffolds provided in the Biology Guided Inquiry Learning Environment (BGUILE; [27]) were found to benefit students' performance on interpreting data. In the intelligent tutoring Inq-ITS [5], adaptive scaffolds in the form of hints that popped-up on students' screens were shown to benefit students' performance on several investigative inquiry practices [12, 13, 19, 20, 24, 25]. In addition, scaffolds have been shown to support both near and far transfer of inquiry practices.

Near Learning Transfer and Far Learning Transfer

Transfer of inquiry practices may occur within contexts that are similar to the initial context in which the practices were learned and are experienced shortly after the initial learning occurred (i.e. near transfer; [2]). Additionally, transfer of inquiry practices can occur in contexts that are different from the initial context in which the practices were learned and are experienced long after the initial learning occurred (i.e. far transfer; [2]). Central to the transfer of inquiry practices, however, is the timing and type of scaffolds that are provided to students [22].

Fixed scaffolds are supports that are presented to all students completing an inquiry task and the timing, frequency, and specificity of these scaffolds are not individualized based on students' performance [18, 27]. Faded scaffolds are supports that are reduced with students' increasing use of the system, but the reduction of supports is not based on students' performance or needs [16, 18, 28].

Adaptive scaffolds, on the other hand, are supports that are provided based on the real-time assessment of students' performance and corresponding needs [5, 22]. Adaptive scaffolds, therefore, provide support to students when they need it most and when it is most effective for learning [11], therefore resulting in transfer of learning in the future [22]. For instance, prior studies have demonstrated that after receiving adaptive scaffolds in the intelligent tutoring system Inq-ITS, students were able to transfer inquiry practices to new contexts after only a short time period (near transfer; i.e. just over one month) and between inquiry investigations of similar content (near transfer; i.e. learning the practices in animal cell inquiry activities and applying the practices to plant cell activities [13]). Studies on the adaptive scaffolding in Inq-ITS have also shown that students' learning of inquiry practices transfers to inquiry tasks several months after initial scaffolding when applied to inquiry in new topics (far transfer; i.e. learning practices in animal cell activities and applying the practices in activities on natural selection; [13]).

Inq-ITS

Inq-ITS is, to our knowledge, the only online learning and assessment system that provides adaptive scaffolding for multiple investigative inquiry practices [5]. The adaptive scaffolding in Inq-ITS is possible as a result of automated scoring algorithms that are based on educational data mining (EDM) and knowledge engineering (KE) techniques [4, 6].

The EDM and KE algorithms score students' performance on the investigative practices of asking questions/forming hypotheses, carrying out investigations/collecting data, and interpreting and analyzing data at the sub-component level, allowing for real-time, fine-grained assessment of inquiry competencies at scale [6]. Automated scoring has also been implemented for students' construction of written claim, evidence, and reasoning statements at the end of the inquiry investigations using natural language processing techniques (researchers are currently in the process of developing and implementing adaptive scaffolding for students' explanations based on this automated scoring [14]).

The fine-grained assessment of students' competencies in Inq-ITS allows for providing real-time adaptive support based on the specific type and extent of students' difficulties. In Inq-ITS, students can receive one of several types of scaffolds from the pedagogical agent, Rex, including: *orienting scaffolds* (i.e. Rex directs the student's attention to a particular component of a step/task), *conceptual scaffolds* (i.e. Rex provides an explanation of an inquiry practice needed for a particular step), *procedural scaffolds* (i.e. Rex provides the student with information about the procedure to use on a particular step), *instrumental scaffolds* (i.e. Rex tells the student exactly how to move forward on a particular step). As students demonstrate increased difficulty with particular sub-components of practices, they will receive increased support from orienting, to conceptual, to procedural, and finally to instrumental scaffolds designed to address the inquiry practice sub-component.

Prior studies on Inq-ITS have demonstrated the transfer of practices (near and far; [5, 12, 13, 19, 20, 24, 25]) in terms of student learning gains across activities over time. Researchers, however, have yet to investigate the pattern in *the number of scaffolds* provided to students over time and across inquiry topics, and how this pattern could be used to predict students' inquiry performance. As a result, it is unclear whether the *number of scaffolds* students receive has any relationship to their later performance on inquiry. Additionally, studies have yet to investigate whether the amount of help students require (i.e. number of scaffolds) decreases with increased use of the system as a result of initial adaptive scaffolds.

Automated adaptive scaffolding is scalable in terms of how, regardless of the number of students completing an activity, students have the opportunity to receive individualized attention and support. It is important to first examine the effectiveness of this scaffolding at a smaller scale before implementing at a large scale.

RESEARCH QUESTIONS

In this study we explored whether the amount of scaffolding that students needed and received led to transfer of practices over time and across topics. We conducted two studies: (1) we investigated the relationship between *the amount of scaffolding that students received* and the *number of driving questions they completed;* the efficacy of the scaffolds would be confirmed if fewer scaffolds were required over time and across driving questions; and (2) we examined the relationship between *the amount of scaffolding students received and their inquiry performance over time and across driving questions*. Specifically, we were interested to test whether students' inquiry performance would improve over time, even with fewer scaffolds. If borne out, these data have important implications for scaling-up online inquiry learning and assessment environments to best support students' inquiry practice competencies.

METHOD

Participants

107 middle school students in 6th grade participated in the present study. The middle school is located in the northeastern United States and the demographics of the student population are: 39.2% white, 20.6% Hispanic, 23.5% Asian, 11% black, and the remaining students are two or more races.

Materials and Scaffolding

The students in the present study completed the Inq-ITS [6] Animal Cell and Plant Cell virtual labs during their regular science class period. In the Animal Cell labs, the students investigated three driving questions including: 1) how to help the golgi body receive more protein, 2) how to decrease the production of ribosomes, and 3) how to decrease the amount of protein being produced by the cell. About 40 days later, students completed one driving question in the Plant Cell virtual lab where they had to investigate: how to fix the problem that the cell was not capturing enough energy from sunlight.

Each of the virtual lab activities that the students completed in Inq-ITS contained four stages. In the first stage, students were forming questions/hypothesizing. They then were carrying out investigations/collecting data followed by analyzing and interpreting data. Finally, students were communicating findings. Adaptive scaffolding is currently available for inquiry sub-practices in the first three stages of the Inq-ITS lab [6, 15, 19, 20, 24, 25] based on the automated scoring (described in the measures section) of students' performance on fine-grained components of inquiry practices.

In the present study, all students were assigned into the scaffolding condition for the Animal Cell Labs. Therefore, *if students demonstrated low performance* on an inquiry practice (i.e., hypothesizing, collecting data, and analyzing data including interpretation and warranting) in the Animal Cell labs, then the pedagogical agent (Rex) popped up on their screen with individualized feedback presented in a speech bubble. The particular feedback provided from Rex was iteratively developed based actual effective feedback provided by teachers to students on these practices. Therefore, pre-developed feedback for practices at varying levels of specificity is triggered based on student performance and delivered by Rex (see [24] and [25] for more information on the development of scaffolds).

The feedback starts off by orienting the student toward the particular practice that the student is having difficulty with (see Figure 1). The feedback next involves a procedural scaffold with hints on how to engage in the inquiry practice correctly (see Figure 2). Rex will eventually provide the student with more detailed information on the inquiry practice in the form of a conceptual scaffold (see Figure 3). Finally, Rex provides an instrumental scaffold with more explicit instruction on how to move forward in the activity and explaining the instructions (see Figure 4). The student must address Rex's feedback in order to move forward in the activity. If the student demonstrates perfect performance on their first attempt on each inquiry practice, then it is possible that the student may not receive any scaffolding from Rex. In the present study, Rex's adaptive scaffolding was only available in the three Anima Cell virtual lab activities.

In the Plant Cell virtual lab, all the students were assigned into the no-scaffolded condition. The reason why we added the fourth driving question in a different topic (Plant Cell versus Animal Cell) and had students complete the activity 40 days later was so that we could investigate whether inquiry learning from Rex's scaffolding could be transferred to a different topic after a long period of time. While both Animal and Plant Cell activities occur within the domain of life science, these activities contain different content and goals that students must investigate regarding different organelles of cells (some of which exist within plant cells but not within animal cells; i.e., investigating Chloroplast in the plant cells). Therefore, students are challenged to transfer their inquiry practice competencies from one topic to the other in these activities that require different understandings.

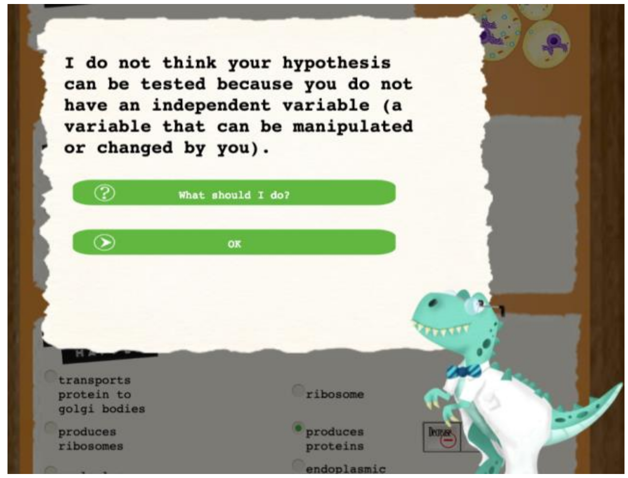

Figure 1. Example of an orienting scaffold.

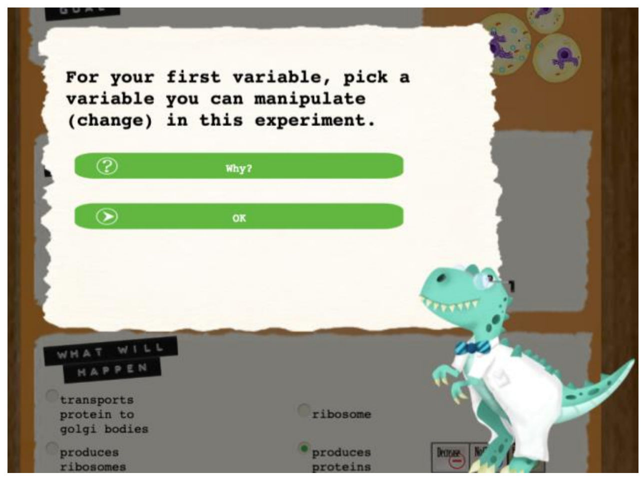

Figure 2. Example of a procedural scaffold.

Figure 3. Example of a conceptual scaffold.

Figure 4. Example of an instrumental scaffold.

STUDY 1: SCAFFOLDING AND PRACTICE

Study 1 investigated whether the scaffolding for a specific inquiry practice decreased with an increased number of inquiry practice experiences. A negative relationship is evidence of successful near transfer, i.e., from driving question 1 to 2 and 3 in the Animal Cell virtual labs.

Measures in Study 1

The dependent variable was the amount of Rex's scaffolding within practices within driving questions. Specifically, it counted the number of scaffolds that students received for the same practice that they completed in the prior driving question activity. For example, a student could have completed four driving question activities in total and received a different number of scaffolds per practice across the driving questions. For example, on the first driving question, they could have received 5 scaffolds for Hypothesizing, 7 scaffolds for Collecting Data, and 6 scaffolds for Analyzing Data; on the second driving question, they could have received 0 scaffolds for Hypothesizing, 4 scaffolds for Collecting Data, and 5 scaffolds for Analyzing Data; on the third driving question activity, they could have received 0 scaffolds for Hypothesizing, 2 scaffolds for Collecting Data, and 1 scaffold for Analyzing Data (see the "Number of Scaffolds" column in Table 1 for a visualization of this example). For the fourth driving question, students were assigned into the no scaffolding condition and therefore did not have access to help from Rex (i.e. scaffolding was not available in the fourth activity).

The independent variable in Study 1 was the number of driving questions. In the fourth driving question, students were assigned into the no-scaffolded condition; therefore, only the first three driving questions were included in the present analyses.

Results: Scaffolding and Practice

Table 2 displays the means and standard deviations (SD) of the amount of scaffolding within practices within driving questions. For the practice of generating a hypothesis, the number of scaffolds that students received on average decreased from 2.20 to 1.79 and to 1.84 from the first driving question to the second, and to the third, but increased from 1.79 to 1.84 from the second to the third. For the practice of collecting data, the average number of scaffolds that students received decreased from 2.84, to 0.85, and then to 0.66 from the first driving question, to the second and then to the third driving question. A similar pattern was found for the practice of analyzing data: the number of scaffolds that students received dramatically decreased from 11.78 in the first driving question, to 6.50 in the second driving question, and to 3.07 to the third driving question.

Practice	Driving Question	Number of Scaffolds	Number of Prior Scaffolds
Generating Hypotheses	1	5	0
	2	0	5
	3	0	0
	4	0	0
Collecting Data	1	7	0
	2	4	7
	3	2	4
	4	0	2
Analyzing Data	1	6	0
	2	5	6
	3	1	5
	4	0	1

Table 1. Example of scaffolding that a student received.

Driving Question	Generating Hypotheses	Collecting Data	Analyzing Data
1	2.20(3.56)	2.48(3.79)	11.78(19.52)
2	1.79(3.47)	0.85(1.66)	6.50(11.94)
3	1.84(3.75)	0.66(1.57)	3.97(6.85)

Table 2. Means and SD of the Number of Scaffolds.

We computed Pearson correlations between the number of scaffolds and the number of driving questions for practices of hypothesizing, collecting data, and analyzing data, respectively. Results of Pearson correlations for the practice of generating a hypothesis, the least complex of the practices studied, showed that the number of scaffolds was not significantly correlated with the number of driving questions. The regression model was also not significant.

However, results for the practice of collecting data yielded significant negative correlations between the number of scaffolds and the number of driving questions activities that students completed (r = -0.28, $p < .001$). The regression model showed a significant constant ($\beta = 3.14$, $p < .001$) and a significant, negative coefficient for the number of driving questions (β = -0.91, $p < .001$), which indicates that as students completed more driving questions, the average number of Rex's scaffolds decreased by 0.91 scaffolds. These findings imply that when students received Rex's scaffolding for the practice of collecting data, this scaffolding helped students learn and improve their performance on collecting data in the next driving question as evidenced by students needing less scaffolding.

Practice	*r*	*β*
Generating Hypotheses	-0.04	
Constant		2.30^{***}
Number of Driving Question		-0.18
Collecting Data	-0.28^{***}	
Constant		3.14^{***}
Number of Driving Question		-0.91^{***}
Analyzing Data	-0.23^{**}	
Constant		15.22^{***}
Number of Driving Question		-3.90^{***}

Note. $^{***}p < .001$. $^{**}p < .01$. $^{*}p < .05$. $df = 1, 319$. $N = 321$.

Table 3. Correlations between the Number of Scaffolds and the Number of Driving Questions and Coefficients.

A similar pattern was found for the practice of analyzing data. Results showed a significant negative correlation between the number of scaffolds and the number of driving question activities that students completed ($r = -0.23$, $p < .01$). The regression model showed a significant constant ($\beta = 15.22$, $p < .001$) and a significant, negative coefficient for the number of driving questions ($\beta = -3.90$, $p < .001$), which indicates that with more driving questions, the number of Rex scaffolds decreased by 3.90 scaffolds on average. These findings suggest that when students received Rex's scaffolding for the practice of analyzing data, it dramatically helped students learn and improve their performance on this practice on the next driving question because students' need for scaffolding was considerably reduced.

The marked need for Rex's scaffolding (more than 11 scaffolds on average with a standard deviation of 19.52; see Table 2) on the first driving question showed that the practice of analyzing data was the most challenging inquiry practice. Specifically, students required a great deal of support in order to successfully engage in this practice. However, students' learning during this first driving question was successfully transferred in later driving questions. With each additional driving question that students completed, the need for Rex's scaffolding was reduced. These findings showed that even for this challenging inquiry practice, greater inquiry learning gains and transfer could be successfully achieved as a result of scaffolding.

The need for Rex's scaffolding in the first driving question for the practice of collecting data indicates that it was slightly less challenging relative to the practice of analyzing data: students only needed approximately 3 scaffolds on average. Evidence of the benefits of initial scaffolding for this practice, however, could be seen as students required significantly fewer scaffolds over time (see Table 2).

Students needed very few scaffolds for the practice of hypothesizing (around 2 scaffolds on average; see Table 2) in their first driving question activity, indicating that the practice of hypothesizing was relatively simple for students. Students required even fewer scaffolds over time (less than 2 on average) for the practice of hypothesizing, but there was no significant difference between the number of scaffolds needed from the first to the third driving question. Therefore the transfer of hypothesizing performance was not as substantial as for the practices of collecting data and analyzing data, even though students required very minimal support for hypothesizing across each driving question activity.

Overall, these findings show promise regarding the benefits of scaffolds for the practices of analyzing data and collecting data in that students require less scaffolding on these practices over time (see Figure 5).

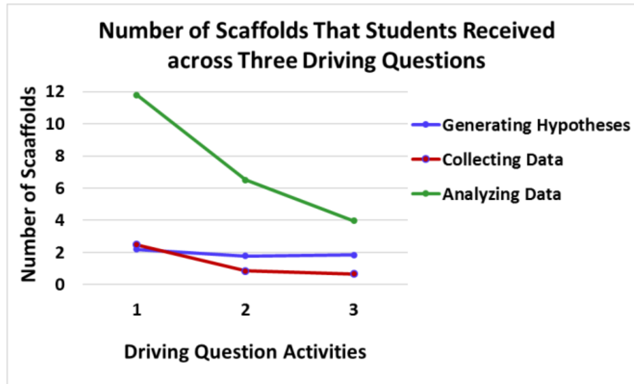

Figure 5. Number of scaffolds that students received within each practice across three driving questions in Animal Cell virtual labs.

STUDY 2: SCAFFOLDING AND PERFORMANCE

Study 1 revealed a significant, negative relationship between the number of scaffolds students received for several inquiry practices and the number of driving questions they completed. Therefore, students required less support over time to successfully engage in inquiry practices. However, Study 1 did not explicitly provide evidence of whether less help over time was positively related to students' inquiry performance. If students' performance continued to increase as their need for help decreased, then this finding would provide clear evidence of transfer. Study 2, therefore, further investigated the relationship between the number of prior scaffolds received for a certain inquiry practice and students' corresponding inquiry performance. Similar to Study 1, a negative relationship would demonstrate both successful near transfer (from driving question 1 to 2 and 3 in the Animal Cell virtual labs) and far transfer (from the topic of Animal Cell to the topic of Plant Cell completed 40 days later).

Measures in Study 2

In study 2, the four dependent variables in each analysis were the students' scores on inquiry practices as automatically assessed within Inq-ITS [4, 6]. Specifically, each practice is scored based on the binary scoring (1 if correct or 0 if incorrect) of its sub-practices. The first practice is hypothesizing, which is based on two sub-practices: identifying an IV (independent variable) and DV (dependent variable). The second practice is the practice of collecting data, which is based on two sub-practices: testing the articulated hypothesis and conducting a controlled experiment. The third practice is interpreting data, which is based on four sub-practices: selecting the correct IV and DV for a claim, interpreting the relationship between the IV and DV, and interpreting the hypothesis/claim relationship. The final practice is warranting claims, which is based on four sub-practices: warranting the claim with more than one trial, warranting with controlled trials, correctly warranting the relationship between the IV and DV, and correctly warranting the hypothesis/claim relationship. The analyses used students' performance on their first attempts for each inquiry practice, prior to receiving Rex scaffolding (when applicable). The scores used for analyses for each practice were the averages across the sub-practices.

The independent variable in this study was the number of prior Rex scaffolds within practices across driving questions. Therefore, we examined the number of scaffolds students received for the practices of hypothesizing, collecting data, interpreting data, and warranting data (note that the practices of interpreting data and warranting data occur during the Analyzing Data stage of the Inq-ITS lab and therefore were combined for the purposes of Study 1; we examine performance on these two separately for the purposes of Study 2 to get greater insight on how our scaffolding helps students' inquiry). Specifically, we counted the number of scaffolds that students received for the same practice that they completed in the prior driving question activity. Take the same example as in Study 1. A student completed four driving questions in total. He/she received Rex's help 5 times for the practice of hypothesizing in the first driving question activity, 0 times in both second and third driving question activities, and no scaffolding in the fourth driving question activity. Thus, the number of prior scaffolds in the first driving question was 0, 5 times in the second driving question, and 0 times in the third and fourth driving questions (see the "Number of Prior Scaffolds" column in Table 1 for a visualization of this example). This feature will allow us to examine how the amount of scaffolding received in the previous activity for a practice is related to performance on that practice in the next driving question activity, thereby identifying whether there is near transfer (across the driving questions in the Animal Cell virtual lab). However, far transfer will be tested through examining the number of scaffolds in the third driving question relative to the fourth driving question because the students switched topics (from Animal Cell to Plant Cell) and there was a 40-day gap between the completion of the third and fourth driving question activities.

Prior studies [12, 13] showed that the number of driving questions and time (day 1 and day 40) were both significant predictors for competencies on science inquiry practices. In Study 2, therefore, we performed hierarchical regression analyses for each inquiry practice. Model 1 for each practice included the number of driving questions and time as predictor variables. Model 2 included all of the features in Model 1 as well as the number of prior scaffolds to examine whether adding the number of prior scaffolds could significantly improve the model. If the model is significantly improved, the number of prior scaffolds as a robust predictor would be confirmed.

Analyses: Scaffolding and Performance

Table 4 displays the means and standard deviations (*SD*) of the number of prior scaffolds within practices across driving questions and inquiry scores across four driving questions in terms of each inquiry practice: hypothesizing, collecting data, interpreting data, and warranting claims. The number of prior scaffolds for interpreting data and warranting claims practices was the same because Rex provided support for analyzing data, which included these two practices. With the comparison of the number of scaffolds in Table 2 and the number of prior scaffolds in Table 4 for a particular practice, the value of the number of scaffolds was shifted below to the next driving question and then feature of the number of prior scaffolds was obtained.

Practice	**Time**	**Driving Questions**	**Mean (Standard Deviations)**	
			Score	**Number of Prior scaffolds**
Generating Hypotheses	1	1	0.74(0.34)	0.00(0.00)
	1	2	0.78(0.28)	2.20(3.56)
	1	3	0.84(0.28)	1.79(3.47)
	2	4	0.88(0.26)	1.84(3.75)
Collecting Data	1	1	0.52(0.48)	0.00(0.00)
	1	2	0.80(0.37)	2.48(3.79)
	1	3	0.89(0.27)	0.85(1.66)
	2	4	0.88(0.27)	0.64(1.55)
Interpreting Data	1	1	0.79(0.27)	0.00(0.00)
	1	2	0.84(0.22)	11.78(19.52)
	1	3	0.82(0.25)	6.26(11.82)
	2	4	0.87(0.28)	4.09(7.11)
Warranting Claims	1	1	0.55(0.39)	0.00(0.00)
	1	2	0.68(0.35)	11.78(19.52)
	1	3	0.75(0.31)	6.26(11.82)
	2	4	0.81(0.32)	4.09(7.11)

Note. Time 1 = day 1, Animal Cell. Time 2 = day 40, Plant Cell.

Table 4. Means and SD of Inquiry Scores and the Number of Prior scaffolds within Practices across Driving Questions.

A series of relevant assumptions were tested before conducting analyses. The criteria for interpreting magnitude of correlations was: small (r = 0.1), medium (r = 0.3), and large (r = 0.5) [3]. Table 5 displays the correlations of all the variables. First, an examination of the correlations revealed that all variables were significantly correlated for at least one inquiry practice and that the highest correlations did not exceed the limit of the assumption that correlations between each pair of independent variables should be less than .80. Therefore, we kept all the variables. The sample size was deemed adequate, given three independent variables (N = 428) within each inquiry practice [26]. The collinearity statistics were all within acceptable limits, as tolerance was greater than 0.10 and the variance inflation factor was below 10; thus, the assumption of multicollinearity was satisfied [1, 8, 23]. A value of Cook's distance less than 1 met the assumption of outliers. Residual and scatterplots indicated that the assumptions of normality, linearity, and homoscedasticity were all satisfied [7, 23].

We conducted a 2-step hierarchical regression analysis for hypothesizing, collecting data, interpreting data, and warranting claims practices, respectively. For each analysis, time and the number of driving questions were entered at Step 1. This was to control for the repeated measures of inquiry practices (across four driving questions) and time from day 1(first three driving questions) to day 40 (fourth driving question; Model 1). The number of prior scaffolds that students received was entered at Step 2 (Model 2). The order of these three variables could answer our question of whether the effects of Rex's scaffolding could be successfully transferred when time and the number of driving questions were controlled.

Variable	**Scores**	**1**	**2**
	Hypothesizing		
1. Time	0.13^{**}		
2. Number of Driving Questions	0.17^{***}	0.78^{***}	
3. Number of Prior Scaffolds	-0.16^{**}	$0.07^{\dagger}$	0.18^{***}
	Collecting Data		
1. Time	0.16^{***}		
2. Number of Driving Questions	0.34^{***}	0.78^{***}	
3. Number of Prior Scaffolds	-0.17^{***}	-0.09^{*}	0.01
	Interpreting Data		
1. Time	0.09^{*}		
2. Number of Driving Questions	0.09^{*}	0.78^{***}	
3. Number of Prior Scaffolds	-0.18^{***}	$-0.07^{\dagger}$	0.06
	Warranting Claims		
1. Time	0.18^{**}		
2. Number of Driving Questions	0.27^{***}	0.78^{***}	
3. Number of Prior Scaffolds	-0.13^{**}	$-0.07^{\dagger}$	0.06

Note. *** $p < .001$. ** $p < .01$. * $p < .05$. † $p < .10$.

Table 5. Pearson correlations between variables (N = 428).

Results: Scaffolding and Performance

Table 6 displays the statistics related to the change in R^2 at each step in terms of four inquiry practices, including hypothesizing, collecting data, interpreting data, and warranting claims. Table 7 shows the coefficients of each variable in the best model, which ended up occurring at Step 2 for the practices of hypothesizing, collecting data, interpreting data, and warranting claims, respectively.

To answer the second research question, whether Rex's scaffolding could lead to successful transfer (near and far), the changes in variance explained by the models (R^2) were compared across the two models for each inquiry practice. Specifically, we examined whether adding the number of prior scaffolds to the regression model of time and the number of driving questions would significantly improve the model.

Hypothesizing

Results of the hierarchical regression analysis revealed that at Step 1, time and the number of driving questions significantly contributed to the regression model, accounting for 3% of the variance in hypothesizing performance, $F(2,425) = 6.65$, $p = .001$, $R^2 = 0.03$. At Step 2, adding the number of prior scaffolds that students received explained an additional 4% of the variance in hypothesizing performance, and this change was significant, $p < .001$ (see Table 6). The variables of time, the number of driving questions, and the number of prior scaffolds together significantly explained 7% of the total variance in hypothesizing performance, $F(3,424) = 10.16$, $p < .001$, $R^2 = 0.07$.

The full regression model (Model 2) showed a significant constant (β = 0.71, $p < .001$), a significant, positive coefficient for the number of driving questions ($\beta = 0.06$, $p < .01$), and a significant, negative coefficient for the number of prior scaffolding ($\beta = -0.02$, $p < .001$) (see Table 7). These findings indicate that as students completed more driving questions, their hypothesizing performance increased by 0.06 and that as students received less scaffolds, their hypothesizing performance increased by 0.02.

Practice: Model	R^2	R^2 ***Change***	***df***	***F Change***
Hypothesizing				
Model 1	0.03	0.03	2, 425	6.65^{**}
Model 2	0.07	0.04	1, 424	16.70^{***}
Collecting Data				
Model 1	0.14	0.14	2, 425	34.46^{***}
Model 2	0.18	0.04	1, 424	21.43^{***}
Interpreting Data				
Model 1	0.01	0.01	2, 425	1.93
Model 2	0.05	0.04	1, 424	15.85^{***}
Warranting Claims				
Model 1	0.08	0.08	2, 425	17.14^{***}
Model 2	0.10	0.03	1, 424	12.21^{**}

Note. Model 1: Predictors: Time + Number of Driving Questions. Model 2: Predictors: Time + Number of Driving Questions + Number of Prior Scaffolds within Practices across Driving Questions.

Table 6. Unique contribution of the number of prior scaffolds within practice across driving questions to inquiry scores.

These findings indicated that the number of driving questions and the number of prior scaffolds were both related to performance on the practice of hypothesizing, with the number of driving questions having more predictability than the number of prior scaffolds. The larger predictive weight of the number of driving questions implies that students may acquire more information on the practice of hypothesizing when they obtained help from Rex during the prior driving question activity (i.e. if students received Rex's help in the second driving question activity, then they would require less help from Rex in the third driving question activity). This is consistent with our findings that the repeated use of Inq-ITS supports improvement on the practice of hypothesizing [12]. Moreover, students' hypothesizing performance increased even when they received fewer scaffolds, indicating that as students mastered the practice of hypothesizing, they required less support. Time was not a significant predictor for the performance on the practice of hypothesizing, which indicated that hypothesizing performance was not substantially influenced by the 40 days between the activities. This was likely because the hypothesizing practice is a simpler inquiry practice relative to the other inquiry practices. Thus, findings demonstrated successful transfer of the practice of hypothesizing as students required fewer scaffolds over time but continued to demonstrate increased performance.

Variable	***B***	***SE B***	***β***	***T***	***R^2***	***F***
Hypothesizing					0.07	10.16^{***}
(Constant)	0.71	0.04		16.61^{***}		
Time	-0.03	0.05	-0.04	-0.48		
Driving Questions	0.06	0.02	0.24	3.13^{**}		
Prior Scaffolds	-0.02	0.004	-0.20	-4.09^{***}		
Collecting Data					0.18	31.22^{***}
(Constant)	0.65	0.05		12.12^{***}		
Time	-0.27	0.06	-0.30	-4.30^{***}		
Driving Questions	0.20	0.02	0.58	8.20^{***}		
Prior Scaffolds	-0.03	0.01	-0.21	-4.63^{***}		
Interpreting Data					0.05	6.62^{***}
(Constant)	0.80	0.04		21.03^{***}		
Time	-0.01	0.05	-0.02	-0.26		
Driving Questions	0.03	0.02	0.12	1.57		
Prior Scaffolds	-0.004	0.001	-0.19	-3.98^{***}		
Warranting Claims					0.10	15.79^{***}
(Constant)	0.55	0.05		10.71^{***}		
Time	-0.10	0.06	-0.12	-1.62		
Driving Questions	0.12	0.02	0.37	5.03^{***}		
Prior Scaffolds	-0.01	0.001	-0.16	-3.49^{**}		

Note. *** $p < .001$. ** $p < .01$. * $p < .05$. † $p < .10$

Table 7. Coefficients in the full model (*df* (2, 425)).

Collecting Data

Results of the hierarchical regression analysis for the practice of collecting data revealed that at Step 1, time and the number of driving questions significantly contributed to the regression model, accounting for 14% of the variance in collecting data performance, $F(2,425) = 34.46$, $p = .001$, $R^2 = 0.14$. At Step 2, adding the number of prior scaffolds that students received explained an additional 4% of the variance in collecting data performance, and this change was significant, $p < .001$ (see Table 6). The variables of time, the number of driving questions, and the number of prior scaffolds together significantly explained 18% of the total variance in collecting data performance, $F(3,424) = 3.86$, $p < .001$, $R^2 = 0.18$.

The full regression model (Model 2) showed a significant constant ($\beta = 0.65$, $p < .001$), a significant, positive coefficient for the number of driving questions ($\beta = 0.20$, $p < .01$), a significant, negative coefficient for time ($\beta = -0.27$, $p < .001$), and a significant, negative coefficient for the number of prior scaffolding ($\beta = -0.03$, $p < .001$) (see Table 7). These findings indicate that as students completed more driving questions, their collecting data performance increased by 0.20, but decreased 0.27 over time. As students received less scaffolds, collecting data performance increased by 0.03.

These findings indicated that time, the number of driving questions, and the number of prior scaffolds were all related to performance on collecting data, with time having more predictive power than the number of driving questions and the number of prior scaffolds. The larger predictive weight of time implies that students may perform slightly lower on collecting data due to the long time gap from the first topic to the second topic, about 40 days. The number of driving questions had the second greatest predictive power, which indicates that students improved on the inquiry practice of collecting data when they obtained help from Rex in the prior driving question activity. The patterns of time and the number of driving questions are consistent with our findings that the repeated use of Inq-ITS facilitates improvement on the practice of collecting data, but with a slight decrease in performance over long periods of time; however, overall performance improved from the initial attempt [12]. Moreover, students' performance on the practice of collecting data increased even when they received fewer scaffolds. This indicates that as students began to master the practice of collecting data, they required fewer supports from Rex. Therefore, students were able to transfer their learning for the practice of collecting data with increased independence over time and across activities.

Interpreting Data

Results of the hierarchical regression analysis for the practice of interpreting data revealed that at Step 1, time and the number of driving questions did not significantly contribute to the regression model for the practice of interpreting data. However, at Step 2, adding the number of prior scaffolds that students received explained an additional 4% of the variance in interpreting data performance, and this change was significant, $p < .001$ (see Table 6). The variables of time, the number of driving questions, and the number of prior scaffolds together significantly explained 5% of the total variance in interpreting data performance, $F(3,424) = 6.62$, $p < .001$, $R^2 = 0.45$.

The full regression model (Model 2) showed a significant constant (β = 0.80, p < .001) and a significant, negative coefficient for the number of prior scaffolding (β = -0.004, p < .001) (see Table 7). These findings indicate that as students received less scaffolds, their interpreting data performance increased by 0.004.

These findings indicated that only the number of prior scaffolds was significantly related to performance on interpreting data. The small predictive weight of the number of prior scaffolds implies that the amount of scaffolding students received decreased, but this decrease did not result in poorer performance on interpreting data. Thus, findings demonstrate the efficacy of scaffolding on the practice of data interpretation.

Warranting Claims

Results of the hierarchical regression analysis revealed that at Step 1, time and the number of driving questions significantly contributed to the regression model, accounting for 8% of the variance in warranting claims performance, $F(2,425) = 17.14$, p < .001, $R^2 = 0.08$. At Step 2, adding the number of prior scaffolds that students received explained an additional 3% of the variance in warranting claims performance, and this change was significant, p = .001 (see Table 6). The variables of time, the number of driving questions, and the number of prior scaffolds together significantly explained 10% of the total variance in warranting claims performance, $F(3,424) = 15.79$, p < .001, $R^2 = 0.10$.

The full regression model (Model 2) showed a significant constant (β = 0.55, p < .001), a significant, positive coefficient for the number of driving questions (β = 0.12, p < .01), and a significant, negative coefficient for the number of prior scaffolding (β = -0.01, p < .01) (see Table 7). These findings indicate that that as students received less scaffolds, their warranting claims performance increased by 0.01.

The findings of the estimates of coefficients indicated that the number of driving questions and the number of prior scaffolds were both related to the performance on the practice of warranting claims, with the number of driving questions having more predictability than the number of prior scaffolds. The larger predictive weight of the number of driving questions implies that students may acquire more information on warranting claims with additional help from Rex in prior activities. This is consistent with our findings that the repeated use of Inq-ITS facilitates improvement on warranting claims [12]. This finding indicates that the amount of scaffolding that students needed on this practice decreased, but their performance on this practice continued to increase. Time was not a significant predictor for the performance of warranting claims practice. Thus, the present findings demonstrated successful transfer of the practice of warranting claims in terms of how students required less support over time, but still improved their performance.

GENERAL DISCUSSION AND CONCLUSIONS

In this study, we explored whether the scaffolding that students needed and received during science inquiry could be successfully transferred to the next practice over time and across topics, that is, we tested near and far transfer. We conducted two studies to investigate this question. First, we examined the relationship between the number of scaffolds that students received and the number of driving questions. We found that the number of scaffolds needed significantly decreased with an increasing number of driving questions for the practices of collecting data, interpreting data, and warranting claims. These findings demonstrate that students improved their inquiry practice competencies (for collecting data, interpreting data, and warranting claims) after using an online, scalable system, as indicated by their need for less scaffolding with increased use of the system over time.

Second, we examined the relationship between the amount of prior scaffolding that students received and their performance on inquiry practices in relation to time and topic. We found that the number of prior scaffolds that students needed negatively predicted their performance on all four inquiry practices: hypothesizing, collecting data, interpreting data, and warranting claims. This finding indicates that students required less assistance from the pedagogical agent Rex as they improved their inquiry practice competencies. As the goal of scalable online environments such as Inq-ITS is to promote students' competencies at science inquiry practices as well as their ability to conduct inquiry independently, the results of the present studies are extremely promising. In the future, it would be valuable to more closely monitor student activities between implementations of Inq-ITS in order to understand how classroom learning and effects may have influenced student performance. Additionally, it would be valuable to use different analytic techniques to examine near and far transfer effects, respectively.

Real-time, adaptive support enables successful inquiry learning and transfer of practices. This work exemplifies how assessment and scaffolding of science inquiry practices in virtual labs can be scaled by using automated scoring and educational data mining techniques to capture the quality of student inquiry practices. In particular, a large-scale implementation of this adaptive scaffolding would provide individual students with the opportunity to receive individualized, effective support. Automated evaluation that takes into account students' science inquiry proficiencies at a fine-grained level combined with the scaffolding in virtual environments is an important step towards assessing and supporting the full complement of inquiry practices at scale.

REFERENCES

1. R. S. Baker, J. Clarke-Midura, J. Ocumpaugh. 2016. Towards general models of effective science inquiry in virtual performance assessments. *J Comp Assist Learn* 32: 267-280.

2. Z. Chen, D. Klahr. 2008. Remote transfer of scientific-reasoning and problem-solving strategies in children. In *Advances in child development and behavior*. JAI, 419-470.
3. J. Cohen. 1992. A power primer. *Psychological Bulletin* 112: 155-159.
4. J. D. Gobert, R. S. Baker, M. A. Sao Pedro. 2014. Inquiry skills tutoring system, U.S. Patent 9,373,082, Filed February 1, 2013, issued January 29, 2014.
5. J. Gobert, R. Moussavi, H. Li, M. Sao Pedro, R. Dickler. 2018. Scaffolding students' on-line data interpretation during inquiry with Inq-ITS. In *Cyber-Physical Laboratories in Engineering and Science Education.* Springer.
6. J. D. Gobert, M. Sao Pedro, J. Raziuddin, R. S. Baker. 2013. From log files to assessment metrics: measuring students' science inquiry skills using educational data mining. *J Learn Sci* 22: 521–563.
7. A. C. Graesser, D. S. McNamara, J. Kulikowich. 2011. Coh-Metrix: providing multilevel analyses of text characteristics. *Educ Research* 40: 223–234.
8. J. F. Hair Jr., R. E. Anderson, R. C. Tatham, W. C. Black. 1998. *Multivariate data analysis.* Prentice-Hall.
9. C. E. Hmelo-Silver, R. G. Duncan, C. A. Chinn. 2006. Scaffolding and achievement in problem-based and inquiry learning: a response to Kirschner, Sweller, and Clark. *Ed Psych* 42: 99-107.
10. H. Kang, J. Thompson, M. Windschitl. 2014. Creating opportunities for students to show what they know: The role of scaffolding in assessment tasks. *Science Ed* 98: 674-704.
11. K. R. Koedinger, J. R. Anderson. 1998. Illustrating principled design: the early evolution of a cognitive tutor for algebra symbolization. *Interactive Learning Environments* 5: 161-180.
12. H. Li, J. Gobert, R. Dickler. submitted. Evaluating the transfer of scaffolded inquiry: What sticks and does it last?. Submitted to *Conference on Artificial Intelligence in Education.*
13. H. Li, J. Gobert, R. Dickler. submitted. Testing the robustness of inquiry practices once scaffolding is removed. Submitted to *Conference on Intelligent Tutoring Systems*.
14. H. Li, J. Gobert, R. Dickler. 2017. Automated assessment for scientific explanations in on-line science inquiry. In *Proceedings of the 10th International Conference on Educational Data Mining*. EDM Society, Wuhan, 214-219.
15. H. Li, J. Gobert, R. Dickler, R. Moussavi. 2018. The impact of multiple real-time scaffolding experiences on science inquiry practices. In *Lecture Notes in Computer Science.* Springer, 99-109.
16. N. D. Martin, C. D. Tissenbaum, D. Gnesdilow, S. Puntambekar. 2018. Fading distributed scaffolds: the importance of complementarity between teacher and material scaffolds. *Instructional Science*: 1-30.
17. K. L. McNeill, J. S. Krajcik. 2011. *Supporting grade 5-8 students in constructing explanations in science: the claim, evidence, and reasoning framework for talk and writing.* Pearson.
18. K. McNeill, D. J. Lizotte, J. Krajcik, R. W. Marx. 2006. Supporting students' construction of scientific explanations by fading scaffolds in instructional materials. *J Learn Sci* 15: 153-191.
19. R. Moussavi. 2018. *Design, development, and evaluation of scaffolds for data interpretation practices during inquiry.* Worcester Polytechnic Institute, Worcester.
20. R. Moussavi, J. Gobert, M. Sao Pedro. 2016. The effect of scaffolding on the immediate transfer of students' data interpretation skills within science topics. In *Proceedings of the 12th International Conference of the Learning Sciences*. Scopus, Ipswich, 1002-1005.
21. Next Generation Science Standards Lead States. 2013. *Next generation science standards: for states, by states*. National Academies Press.
22. O. Noroozi, P. A. Kirschner, H. J. Biemans, M. Mulder. 2017. Promoting argumentation competence: extending from first-to second-order scaffolding through adaptive fading. *Ed Psych Review:* 1-24.
23. J. Pallant. 2013. *SPSS survival manual.* McGraw-Hill Education.
24. M. Sao Pedro. 2013. *Real-time assessment, prediction, and scaffolding of middle school students' data collection skills within physical science simulations*. Worcester Polytechnic Institute, Worcester.
25. M. Sao Pedro, R. Baker, J. Gobert. 2013. Incorporating scaffolding and tutor context into bayesian knowledge tracing to predict inquiry skill acquisition. In *Proceedings of the 6th International Conference on Educational Data Mining*. EDM Society, 185-192.
26. B. G. Tabachnick, L. S. Fidell. 1996. *Using multivariate statistics* (3rd. ed.). HarperCollins.
27. I. Tabak, B. J. Reiser, B. J. 2008. Software-realized inquiry support for cultivating a disciplinary stance. *Pragmatics & Cognition* 16: 307-355.
28. W. R. van Joolingen, T. de Jong, A. W. Lazonder, E. R. Savelsbergh, S. Manlove, S. 2005. Co-Lab: research and development of an online learning environment for collaborative scientific discovery learning. *Computers in Human Behavior* 21: 671-688.
29. V. S. Vygotsky. 1978. *Mind in society: the development of higher psychological processes*. Harvard University Press, Cambridge.

Acknowledgements
This research is funded by the Department of Education (R305A120778). Any opinions expressed are those of the authors and do not necessarily reflect those of the funding agencies.