A Web-based System to Support Inquiry Learning
Towards Determining How Much Assistance Students Need

Bruce M. McLaren[1], Michael Timms[2], Doug Weihnacht[3], Daniel Brenner[4], Kim Luttgen[4], Andrew Grillo-Hill[4] and David H. Brown[5]

[1]*Carnegie Mellon University, Pittsburgh, PA, U.S.A.*
[2]*Australian Council for Educational Research, Melbourne, Australia*
[3]*MW Productions, San Francisco, California, U.S.A.*
[4]*WestEd, San Francisco, California, U.S.A.*
[5]*D. H. Brown & Associates, Torrance, U.S.A.*

Keywords: Hints, Feedback, Bayesian Network, the Assistance Dilemma, Inquiry Learning, Science Learning.

Abstract: How much assistance should be provided to students as they learn with educational technology? Providing help allows students to proceed when struggling, yet can depress their motivation to learn independently. Assistance withholding encourages students to learn for themselves, yet can also lead to frustration. The web-based inquiry-learning program, *Voyage to Galapagos (VTG),* helps students “follow” the steps of Darwin through a simulation of the Galapagos Island and his discovery of evolution. Students explore the islands, take pictures of animals, evaluate their characteristics and behavior, and use scientific methodology to discover evolution. A preliminary study with 48 middle school students examined three levels of assistance: (1) no support, (2) error flagging, text feedback on errors, and hints, and (3) pre-emptive hints with error flagging, error feedback, and hints. The results indicate that higher performing students gainfully use the program’s support more frequently than lower performing students, those who arguably have a greater need for it. We conjecture that this could be a product of the current *VTG* program only supporting an early phase of the learning process and also that higher performers have better metacognition, particularly in knowing when (and when not) to ask for help. Lower performers may benefit at later phases of the program, which we will test in a future study.

1 INTRODUCTION

A key problem in the Learning Sciences is the *assistance dilemma:* How much assistance is the right amount to provide to students as they learn with educational technology? (Koedinger and Aleven, 2007). While past research with, for instance, inquiry-learning environments clearly points toward *some* guidance being necessary (Geier et al., 2008), it doesn’t fully answer the assistance dilemma (which has also been investigated under the guise of “productive failure” (Kapur, 2009)). Essentially the issue is to find the right balance between, on the one hand, full support and, on the other hand, allowing students to make their own decisions and, at times, mistakes.

There are benefits and costs associated with both ends of this spectrum. *Assistance giving* allows students to experience success and move forward when they are struggling, yet can lead to shallow learning and the lack of motivation to learn on their own. On the other hand, *assistance withholding* encourages students to think and learn for themselves, yet can lead to frustration and wasted time when they are unsure of what to do. Advocates of direct instruction point to the many studies that show the advantages of assistance giving (Kirschner et al., 2006; Mayer, 2004), but this still does not address the subtlety of exactly when and how instruction should be made available, particularly in light of differences in domains and learners (Klahr, 2009).

Research in the area of scientific inquiry learning, where students tackle non-trivial scientific problems by investigating, experimenting, and exploring in relatively wide-open problem spaces, has provided various results about how different types of guidance support students. Researchers

M. McLaren B., Timms M., Weihnacht D., Brenner D., Luttgen K., Grillo-Hill A. and H. Brown D..
A Web-based System to Support Inquiry Learning - Towards Determining How Much Assistance Students Need.
DOI: 10.5220/0004810100430052
In *Proceedings of the 6th International Conference on Computer Supported Education* (CSEDU-2014), pages 43-52
ISBN: 978-989-758-020-8
Copyright © 2014 SCITEPRESS (Science and Technology Publications, Lda.)

have built on inquiry learning theory (Edelson, 2001; Quintana et al., 2004) and have developed and experimented with simulations, cognitive tools, and microworlds to support inquiry learning in science. Systems of this kind that have demonstrated learning benefits include BGUILE (Sandoval and Reiser, 2004), the Co-Lab collaborative learning system (van Joolingen et al., 2005), a chemistry virtual laboratory (Borek et al, 2009; Tsovaltzi et al., 2010), WISE (Linn and Hsi, 2000; Slotta, 2004), and Metafora (Dragon et al., 2013). A large study by Geier et al. (2008) with over 1,800 middle-school students in the experimental condition versus more than 17,000 students in the control, showed that students who were given scaffolded tools for performing inquiry learning exercises (in earth, physical and life science) did significantly better on standardized exams than students who did not use the tools.

Thus, there is evidence that supporting and guiding students in inquiry learning is beneficial. Yet questions still remain: How much support is the right amount? How should assistance vary according to different levels of prior knowledge? To explore these questions we have developed (and continue to develop) a web-based inquiry learning system called *Voyage to Galapagos (VTG)* and will experiment with the software in a systematic manner intended to uncover how much help is necessary for students to learn about the theories of natural selection and evolution. At this stage of our research, we are testing various types of feedback across a spectrum of learners; we are not focused on varying / adapting the general type of feedback based on performance and prior knowledge. Eventually, we will adapt feedback to suit specific learners, but we first intend to answer the question of how much and what kind of feedback is appropriate for different learners with different levels of understanding. By fixing the feedback according to condition at this stage, we will learn, for instance, whether high assistance helps low prior knowledge learners and low assistance is better for high prior knowledge learners.

Voyage to Galapagos is software that guides students through a simulation of Darwin's journey through the Galapagos Islands, where he collected data and made observations that helped him develop his theories. The program provides students with the opportunity to do simulated science field work, including data collection and data analysis during investigation of the key biological principles of variation, function, and adaptation.

In typical inquiry learning fashion, the *VTG* program also provides a wide range of actions that a student can take. For instance, as they travel on the virtual paths of individual islands, students can take pictures of a variety of animals, some of which are relevant to understanding evolution, and some of which are not. This variety of action implies that there are also many possibilities to guide – or not guide – students as they learn and work through the program. Such variety also means that *VTG* is a rich instructional environment to experiment with the assistance dilemma and different amounts and types of guidance.

It is possible to provide assistance at different frequencies (e.g., never, when a student is struggling, or always) and different levels (e.g., flagging errors only, flagging errors and providing textual feedback, providing hints) in *VTG*. In the preliminary study presented in this paper, we evaluate assistance given according to these two dimensions and have conducted a classroom study with 48 middle school students. The study has provided us with initial insights and ideas about modifying and extending *VTG*, and we present and discuss our plans for a larger, more comprehensive study.

2 MISCONCEPTIONS ABOUT THE THEORIES OF NATURAL SELECTION AND EVOLUTION

Evolution is a fascinating academic topic to investigate because few students come to the subject without preconceptions and misconceptions. More often than not, students have misconceptions that must be overcome in order for them to learn and attain a correct understanding. Misconceptions that students have about evolution originate from multiple sources, all of which are related to prior knowledge, beliefs, and conceptions about the world (Alters and Nelson, 2002):

1. *From-experience Misconceptions* – Misconceptions that arise from everyday experience. For example, students may think "mutations" are always detrimental to the fitness and quality of an organism, since the word "mutation" in everyday use typically implies an unwanted outcome.
2. *Self-constructed Misconceptions* – Misconceptions from trying to incorporate new knowledge into an already incorrect concept. For example, students who think that evolution is somehow "progressive", always moving

toward more "positive" variations.

3. *Taught-and-learned Misconceptions* – Misconceptions that arise from informally learned and unscientific "facts." For example, watching movies with dinosaurs and humans can lead students to the mistaken idea that these species lived at the same time (and, of course, they did not).
4. *Vernacular Misconceptions* – Misconceptions that arise from the everyday use versus scientific use of words. For example, "theory" in everyday use means an unsubstantiated idea; the scientific use of "theory" means an idea with substantial supporting evidence.
5. *Religious and myth-based Misconceptions* – Ideas that come from religious or mythical teaching that, when transferred to science education, become factually inaccurate. For example, the belief that the Earth is too young for evolution, given the Bible's dating of the Earth at 10,000 years.

To start the project, in June 2011, we met with a focus group of seven experienced middle and high school teachers from diverse institutions to determine which misconceptions they observe most frequently in their students. The teachers ranked how frequently they encountered a set of 11 common evolution misconceptions in their classrooms. The set of misconceptions was derived from a literature review (e.g., AAAS, 2011; Alters, 2005; Anderson et al., 2002; Bishop and Anderson, 1990; Lane, 2011) and identification of the misconceptions that are relevant to *VTG*. The rankings ranged from, at the top, "Natural selection involves organisms 'trying' to adapt" to the bottom ranking of "Sudden environmental change is required for evolution to occur."

We conducted this focus group in order to better understand the learning of evolution and to develop educational technology that engages students' prior knowledge and misconceptions. Research has shown that if prior knowledge is not directly engaged, students may have trouble grasping new concepts (Bransford et al., 2000). Inquiry learning is one way to engage prior knowledge and overcome misconceptions. Prior work has shown that good scientific inquiry involves systematic steps such as formulating questions, developing hypotheses, designing experiments, analyzing data, drawing conclusions, and reflecting on acquired knowledge. Essentially, students who imitate (or are guided towards) the cognitive processes of scientific experts are most likely to benefit from inquiry (De Jong and van Joolingen, 1998; Klahr and Dunbar, 1988).

In addition, while undertaking these steps, students are likely to reveal and/or act upon their misconceptions, which, in turn, can be directly addressed by the feedback and guidance provided by an educational technology system. In a study with a science inquiry learning environment, Mulder, Lazander & de Jong (2010) concluded that there were two ways of assisting students: by providing domain support in order to increase the effectiveness of their natural inquiry behavior or by supporting their inquiry behavior at the level of domain knowledge. Quintana et al. (2004) called these content support and process support respectively. In this work, we focus on process support, helping students become better inquiry learners.

3 DESCRIPTION OF VTG

Our approach to overcoming misconceptions about evolution is to have students work with *VTG,* a web-based, inquiry program that mirrors Darwin's pathway to the development of the theories of natural selection and evolution. The program, which is largely implemented but still under development, encourages the student to follow the steps of good scientific inquiry, e.g., developing hypotheses, collecting and analyzing data, drawing conclusions. The program also reveals the basic principles of evolution theory to the students. Darwin's early ideas were initially inspired by his observations in the Galapagos Islands, where he noted the patterns of species distribution on the archipelagos. Darwin's observations in Galapagos (and other islands, during his long journey) spurred him to begin formulating his revolutionary ideas (Sulloway, 1982).

Students working with *VTG* have the opportunity to "follow" Darwin's steps and observe and analyze differences among island fauna. This occurs through a virtual exploration of six Galapagos Islands where students take photographs of different animals, watch videos of animal functions, conduct various analyses in a virtual laboratory, and come to conclusions based on assessments of the data.

The VTG program involves three main phases, or "levels":

- *Level 1: Variation* – photograph a sample of iguanas; measure the variation; analyze geographic distribution of variants
- *Level 2: Function* – watch videos on animal functions (e.g., eating, swimming, foraging for food); test animals for relative performance
- *Level 3: Adaptation* – see where animals with

specific biological functions live; hypothesize about selective pressures; draw conclusions

The student is presented with a video before working on each level, which explains key points about the inquiry process of that phase, relevant evolution theory, and prompts the student to begin work. Once the student starts working, they are given assistance according to their experimental condition. There are five possible conditions, varying from no to high assistance. (Only three of the conditions were studied in the preliminary study. These will be explained later in the paper.)

The study presented in this paper focuses only on Level 1, as this has been the focus of our initial feasibility and usability studies. Figure 1 shows a screen shot of *VTG* – Level 1 in which the student is located on the island of Fernandina in the Galapagos and has the viewfinder of her camera focused on an iguana. An overall view of the Galapagos Islands is shown in the upper right, and a close-up view of a portion of a selected island, in this case Fernandina, is shown in the lower right. The student can follow or skip around the virtual path on the selected island, by selecting individual steps that are in the close-up view of that island. When a step is selected, a picture of the view from that point on the island is shown (note: the pictures in the program were actually taken in the Galapagos).

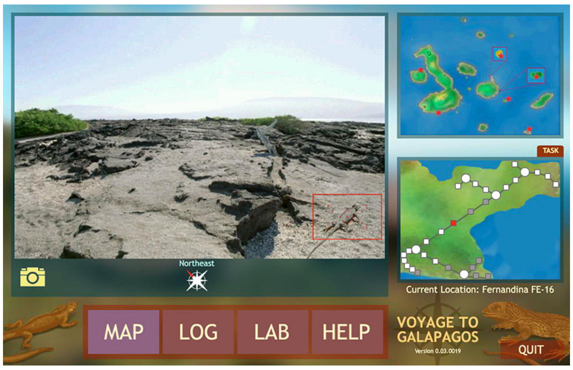

Figure 1: The *VTG* inquiry-learning program, Level 1. Here the student is about to take a picture of an iguana (centered in the red viewfinder box in the photograph).

As the student takes pictures of animals, they are stored in her Logbook, the central repository and organizing tool for the student's inquiry (see Figure 2). Students are instructed to collect iguanas that have as much variation between them as possible. They can take up to 12 photographs in an attempt to cover as wide a variety of iguana traits as possible.

The Lab is the place where students perform various analyses on the data they have collected. It

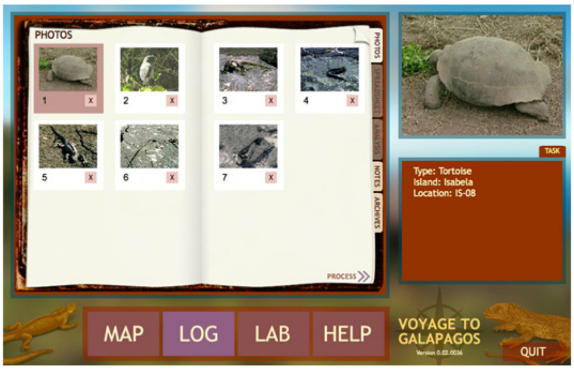

Figure 2: The Logbook of *VTG*. Here the student has taken 7 pictures – 5 iguanas 1 tortoise, and 1 finch. The student should be evaluating iguanas, so it is a mistake – but still permitted by the program – for the student to photograph the tortoise and finch.

provides a link between the three levels of the inquiry tasks that the student is asked to undertake. The Lab contains virtual software tools that the student can use in her analyses. The Schemat-o-meter is a tool that allows the student to examine, measure, and classify traits (e.g., length, tail, shape, color) of the collected animals (see Figure 3). The Trait Tester is a tool, in Level 2, that allows the student to test a hypothesis about the function of a trait variation. The Distribution Chart is a data analysis tool that allows the student in Level 3 to plot the various classified traits of the animals across the islands and habitats where they were collected.

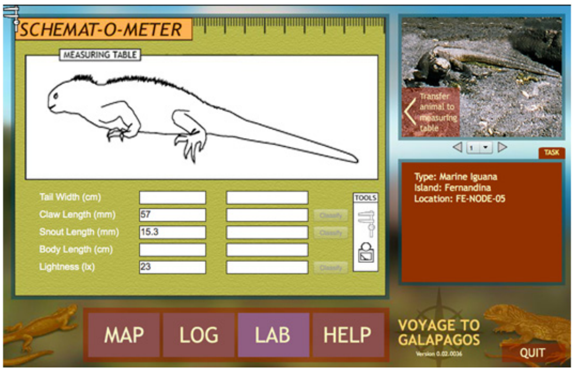

Figure 3: The Schemat-o-meter of *VTG* used to measure and classify traits of the collected animals.

There is considerable "student action" variability within *VTG*; that is, there are many degrees of freedom and opportunities for students to make mistakes. For instance, as shown in Figure 2, the student can take pictures of irrelevant species when they are supposed to focus on iguanas. The student might take pictures of a single animal, say iguanas,

SCITEPRESS
SCIENCE AND TECHNOLOGY PUBLICATIONS

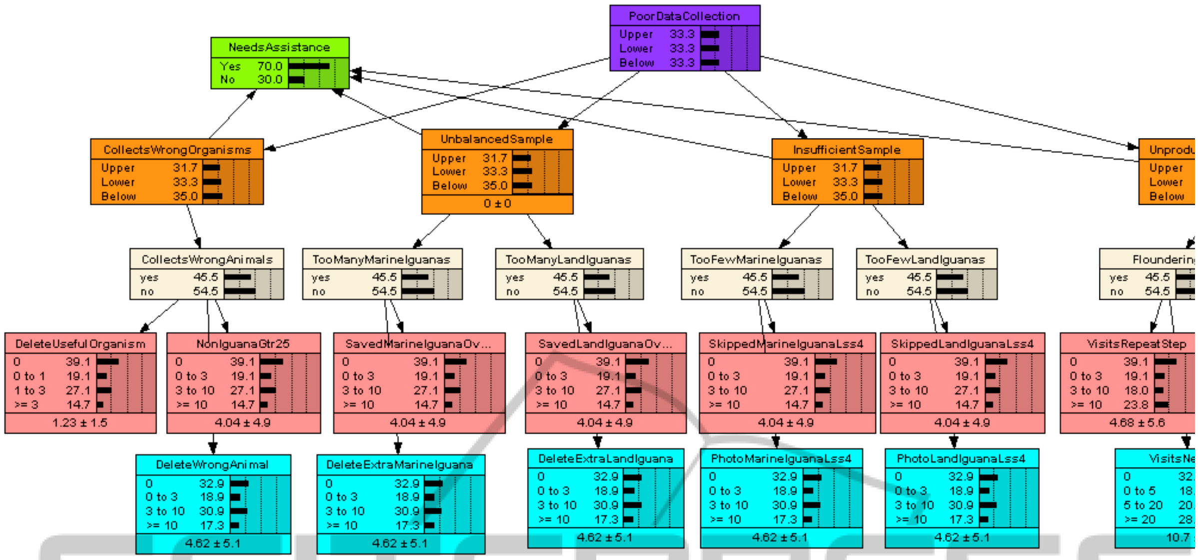

Figure 4: A fragment of the Bayesian Network for Level 1 of *VTG*.

but not take enough pictures to capture trait variation. The student can visit islands that have little useful data to collect or compare traits that will not be useful in learning about variation. This potential variability of student actions – and student errors – allows for a wide variety of assistance, and the ability to intervene with help after those actions are taken – or not. This provides the foundation for our experimental test bed.

A Bayesian Network is used to collect data about student actions, assign probabilities of students having made certain errors, and make decisions about when to turn assistance on so that students receive error feedback and hints, provided students are in conditions to receive such assistance.

The Bayesian Network has four layers (see Figure 4), which are (from the bottom upwards)

1. Supportable events
2. Error diagnosis
3. Error evaluation

Knowledge, Skills and Abilities (KSA) evaluation

The supportable events layer is the first layer of the Bayes Net, and it is where the system captures student actions—both positive and negative (see the bottom two rows of Figure 4)—capturing a combination of things such as the student's location, data collection and state of their logbook. As a student performs an action that has been designated as a supportable event (i.e. one that contributes to our understanding of whether or not a student needs assistance and the nature of their need) the action is recorded. For example, if a student has just stepped away from a place on a path on an island where there is a land iguana (location) without photographing it (action) and she has fewer than four iguana photos saved in her logbook (state of logbook), this is recorded and evaluated by the Bayes Net.

The error diagnosis layer is the second layer (see the middle row of nodes in Figure 4) of the Bayes Net, and its purpose is to determine the type and magnitude of the support a student needs. Each type of assistance a student might need is represented by a node in this layer. The starting state of a node is zero and, as data are received from the supportable events, the probability value of the node will increase or decrease.

The error evaluation layer is the third layer (see the second row of nodes from the top in Figure 4) of the Bayes Net, and its nodes monitor the kinds of errors that the student is making so that, when assistance is turned on, the error feedback can be targeted at the students' need.

The Knowledge, Skills and Abilities (KSA) evaluation layer (see the top row of nodes in Figure 4) is the fourth and highest level of the Bayes Net, which monitors the student's progress toward the instructional targets of the *VTG* module in the sub areas of science practices and science knowledge. The probability values in the nodes of this level are used to report to the student and the teacher progress in their development of knowledge and skills as a result of working through the *VTG* tasks. Nodes at this level are linked to other relevant nodes at the same level in other tasks so that information about a

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS

students' science inquiry skills and knowledge is passed from one task to another, by establishing a beginning value for the node in the new task. This allows the assistance system to weight the level of assistance according to past performance on preceding tasks.

In addition to the four layers of the Bayes Net described above, there is an assistance node in the network, which can be seen as the node in the upper left of Figure 4. As can be seen from Figure 4, the assistance node has edges (the directional lines with arrows in the diagram) that connect to it from nodes in the error evaluation layer. The assistance node is essentially a switch that turns the assistance to the student on or off, depending on its level of activation. The node monitors the probability that the student needs assistance based upon their actions to date. When the probability exceeds a threshold value, the assistance will turn on. Whether a student receives the feedback is configurable according to what experimental condition they are in. By allowing the assistance to be configured in this way, we are able to create the conditions of assistance that are the focus of our experimental design, which is discussed next. As a student receives assistance and s/he corrects the errors made, the probability values of the nodes will drop. Assistance turns off when the probability that the student still needs support falls below the threshold value.

4 STUDY TO EXPLORE THE ASSISTANCE DILEMMA IN VTG

We have two research questions to answer with our quasi-experimental design:

1. How much assistance do students who learn with *VTG* require to achieve the highest learning gains and maximize their inquiry-learning skills?
2. Which mode of assistance is optimal for students with high, medium, and low levels of prior science knowledge and practices?

Our goal is to find the right balance between, on the one hand, full support (i.e., keeping students focused on the learning goals of the program and avoiding mistakes) and, on the other hand, minimum support (i.e., allowing students to make their own decisions and, at times, mistakes).

4.1 General Experimental Design

Our approach and implemented conditions within *VTG* conceives of this as a spectrum of assistance, driven by two orthogonal variables, *Frequency of Intervention* and *Level of Support*. *Frequency of Intervention* is characterized as being in one of three possible states:

- "Never" is when the system does not intervene;
- "When Struggling" is when the system detects that the student is making repeated errors or off-task actions (in this condition students might make several errors before the system decides that they are struggling); and
- "Always" is when the system intervenes at every error or off-task action.

Level of Support is characterized as having three levels of increasing support:

- "Error Flagging" is when an error is flagged by a red outline and exclamation mark only;
- "Error Flagging + Error Feedback" is flagging of the error plus an explanation of the error made; and
- "Error Flagging + Error Feedback + Hints" is flagging of the error, plus explanation, with a series of three levels of available hints.

Table 1 shows the research design that results from crossing these two variables. This 3 x 3 matrix provides a maximum of 9 assistance conditions, but we have combined some of the cells and will not be testing two others, resulting in five conditions (highlighted in yellow in the table).

First, a *Frequency of Intervention* of "Never" essentially means that no assistance will ever be provided, so the *Level of Support* variable is not applicable in that case. Thus, we combine all three cells of the first column of Table 1 to create a single condition, Condition 1 - *No Support*.

Second, we wanted to have a relatively wide mid-range of assistance, achieved by having variations of "When Struggling": Condition 2 – *Flagging, When Struggling*; Condition 3 – *Flagging & Feedback, When Struggling*; and Condition 4 – *Flagging & Feedback & Hints, When Struggling*. In all three of the "when struggling" conditions the provision of assistance is predicated on the current value of nodes in the Bayesian Network, as discussed in the previous section. In particular, when the probability of the student needing assistance exceeds a given threshold for a particular task, the student is assumed to be "struggling" and support is provided, as appropriate to that condition. For instance, in Level 1 students are required to photograph a balanced sample of land and marine iguanas. If a student photographs only land iguanas and no marine iguanas, each time they add a land iguana to their sample, it will increment the probability of "Unbalanced Sample" in the Bayes Net. That node, with others, is connected to an assistance node that switches assistance on at a certain threshold value. When assistance is turned on, the next time a student photographs a land iguana or passes a marine iguana without photographing it, the *VTG* software will provide assistance.

Third, we wanted to include the most extreme level of assistance (i.e., always providing all three levels of support, and also providing pre-emptive assistance): Condition 5 – *Full Support.*

Finally, we wanted to limit the total number of conditions in the experiment, so we would have statistical power in our analysis. Thus, we exclude the somewhat less extreme forms of full support (those in the upper right of Table 1).

Table 1: The Experimental Design, crossing two variables of assistance.

Level of Support \ Frequency of Intervention	Never	When Struggling	Always
Error Flagging		Condition 2 *Flagging, When Struggling*	~~Skipped Condition (Would be Flagging always)~~
Error Flagging + Error Feedback	Condition 1 *No Support*	Condition 3 *Flagging & Feedback, When Struggling*	~~Skipped Condition (Would be Flagging & Feedback always)~~
Error Flagging + Error Feedback + Hints		Condition 4 *Flagging & Feedback & Hints, When Struggling*	Condition 5 Full Support

Ultimately, we will randomly assign approximately 500 students to these five conditions and run an experiment in which we will compare conditions and determine which level of assistance leads to the best learning outcomes, both overall and per different levels of prior knowledge.

With respect to our first research question (i.e., "How much assistance do students who learn with *VTG* require to achieve the highest learning gains and maximize their inquiry-learning skills?"), our hypothesis is that one of the middle conditions – *Flagging, When Struggling; Flagging & Feedback, When Struggling* or *Flagging & Feedback & Hints, When Struggling* – will lead to the best domain and inquiry learning outcomes for the overall student population. These conditions all trade off between assistance giving (such as what is provided by Condition 5) and assistance withholding (such as what is provided by Condition 1). With respect to our second research question (i.e., "Which mode of assistance is optimal for students with high, medium, and low levels of prior science knowledge and practices?"), we hypothesize that Condition 1 (no assistance) will be most beneficial to higher prior knowledge learners and Condition 5 (high assistance) will be most beneficial to lower prior knowledge learners. Our theory is that higher prior knowledge students are more likely to benefit by struggling a bit and exploring without guidance, while lower prior knowledge students, those who are more likely to experience cognitive load (Paas, Renkl, & Sweller, 2003) if left on their own, are more likely to benefit by being strongly supported.

4.2 Experimental Design for the Preliminary Study

4.2.1 Design and Participants

For the purposes of the preliminary study reported in this paper, we reduced the five conditions to three: Condition 1 - *No Support*; Condition 4 – *Flagging & Feedback & Hints, When Struggling*; Condition 5 – *Full Support.* We reduced the conditions as part of our iterative design and development plan. At this stage, we are hopeful of getting a general indication that we are moving in the right direction before conducting the much larger study with all five conditions. In addition, since we had a limited number of participating middle school students in the preliminary study (48), we wanted to have enough students per condition to analyze and report reasonable results. Two classes of a 7th Grade life science course from a suburban San Diego school participated in the study. Of the 48 participating students, 24 were male and 24 were female, with ages ranging from 11 to 13. All students were assigned to one of three conditions as follows:, 13 students were assigned to Condition 1, 25 students were assigned to Condition 4, and 10 students were assigned to Condition 5.

4.2.2 Materials

For this preliminary study, only "Level 1: Variation" was used in order to complete the study in a single 50-minute period. In addition, the teacher provided a rating of each student's science content understanding (Low, Medium, High) and inquiry

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS

skills (Low, Medium, High). We had the teacher provide this information, as opposed to having the students take a pretest, because we had limited class time available to us and wanted to focus on student use of the instructional software. The classes had been previously exposed to the evolution curriculum. Ultimately, *VTG* will be embedded within this curriculum.

4.2.3 Procedure

After a brief introduction to *VTG* by the teacher, Students were given the rest of the 50-minute class period to work with *VTG* (Level 1). This time included viewing an introductory video that provides a brief introduction to the theory of evolution and some general instructions on use of the program. While all learners were presented with the same task – to learn about evolution from the VTG program – they were free to take different pathways through the software in tackling the task and not all students completed the task. This is the very essence of inquiry learning: To explore and to "inquire" in different, perhaps idiosyncratic and incomplete ways.

4.3 Results

A total of 19 of the 48 students were able to complete Level 1 during their single class period of work. An additional student completed Level 1 after school, resulting in a total of 20 students who completed the work. We evaluated how productively the 48 students worked with *VTG* by collecting and calculating the following data:

- *Productive Events*: Actions taken by the student within the VTG software that help to achieve the goals of a particular level (e.g., For Level 1: Photograph a balanced sample of iguanas: 4 marine, 4 iguana; Correctly measure and classify variation).
- *Unproductive Events:* Actions taken by the student within the VTG software that do not help to achieve the goals of a particular level (e.g., For Level 1: Photograph animals other than iguanas; Photograph more iguanas than needed, etc.). These events are effectively errors; steps the student takes that are unlikely to advance his or her understanding of evolution.
- *Ratio of Productive / Unproductive Events*: This is a rough indicator of how productively students work towards solving the Level 1 task, with larger values being better.

To categorize students as high, medium, or low achievers, we took the teacher assessed scores (i.e., content understanding, inquiry skills, with a range of High=3; Medium=2; Low=1 for each), added the two scores together, giving a score between 2 and 6. Students with a score of 6 were labeled "High Achievers", students with a score of 3, 4, or 5 were labeled as "Medium Achievers" and students with a score of 2 were labeled as "Low Achievers." The high, medium, and low achievers in each of the conditions (1, 4, and 5) are shown in Table 2, along with an average number of productive events, average number of unproductive events, and ratio of productive to unproductive events for each category.

Table 2: The results of student use of *VTG – Level 1.*

Category	#	Avg. Prod. Events	Avg. UnprodEvents	Ratio Prod / Unprod. Events
High Achievers Condition 1 - No Support	6	144.5	25.7	5.6
High Achievers Condition 4 - Support	4	125.3	27.3	4.6
High Achievers Condition 5 - Full Support	5	107.0	15.8	6.8
Medium Achievers in Condition 1 - No Support	4	142.3	19.8	7.2
Medium Achievers in Condition 4 - Support	9	111.3	33.9	3.3
Medium Achievers in Condition 5 - Full Support	2	98.5	6.5	15.2
Low Achievers in Condition 1 - No Support	3	99.3	25.3	3.9
Low Achievers in Condition 4 - Support	12	117.3	33.1	3.5
Low Achievers in Condition 5 - Full Support	3	72.3	24.7	2.9

5 DISCUSSION

We emphasize once again that the study and analyses reported here are preliminary; they will be soon be followed by a more extensive experiment with a larger population of students, where we will do more extensive analyses. Thus, this should be considered a preliminary study with only suggestive results.

That said, the ratio of productive to unproductive events shows an interesting pattern, at least with respect to high achievers versus low achievers. Notice that the high achievers appeared to become *more* productive when they received more support (productive to unproductive ratio from 5.6 to 4.6 to 6.8), whereas low achievers appeared to become *less* productive when they received more support (productive to unproductive ratio from 3.9 to 3.5 to 2.9). Medium achievers generally followed the high achievers pattern of improving with support (yet with only 2 students in the medium, full support condition these results are more suspect).

Although the numbers are small and certainly not generalizable, as well as the results pointing more or less in the opposite direction of our general hypothesis (i.e., that higher achievers will do better with less support, lower achievers will do better with more support), we believe there is an underlying rationale to what we've uncovered thus far. *VTG* and this activity was novel to all students, low and high achievers alike, thus all students may have needed support to tackle the task, especially during this early phase of the work (i.e., Level 1). However, the high achievers, as better students are wont to do, seemed to more productively use the provided help (see e.g. Aleven et al., 2006). We believe this could very well change over time, after the higher achievers better understand the process and lower achievers realize the benefits that could come from using the *VTG* support. In any case, the data appears to show that support *can* make a difference, as long as students productively use it.

6 CONCLUSIONS

The assistance dilemma is a fundamental challenge to learning scientists and educational technologists. Until we better understand how much guidance students need as they learn – and how to cater guidance to the prior knowledge level of students – we won't be able to appropriately design instructional software to best support student learning. This is especially so in domains and with software that are open ended, i.e., those that encourage exploration and inquiry.

The *VTG* software, a web-based inquiry-learning environment for learning about the theory of evolution, will allow us to experiment with different types of instructional support and provide an important data point in answering the assistance dilemma. We are in the process of finishing implementation of *VTG* and will soon conduct the full experiment described in section 4.1 with a fully implemented version of the program. The results of the study described in this paper encourage us that we will soon be able to more fully address our research questions.

ACKNOWLEDGEMENTS

This research is supported by a U.S. Department of Education, Institute Of Education Science Grant (Number R305A110021). We would like to thank other past and present members of the *VTG* project team–Donna Winston, Nick Matzke, Russell Almond, and Jerry Richardson–for their contributions to the development of *VTG*. *Voyage To Galapagos* was originally developed as a non-web-based program by the third author of this paper, Weihnacht, under National Science Foundation Award # 9618014.

REFERENCES

AAAS (2011). Project 2061: AAAS Science Assessment. *AAAS: Advancing Science, Serving Society.* http://assessment.aaas.org/

Aleven, V., McLaren, B., Roll, I., & Koedinger, K. (2006). Toward meta-cognitive tutoring: A model of help seeking with a cognitive tutor. *International Journal of Artificial Intelligence in Education (IJAIED), 16*(2), 101-128.

Alters, B. (2005). *Teaching Biological Evolution in Higher Education: Methodological, Religious, and Nonreligious Issues.* Sudbury, Mass.: Jones and Bartlett Publishers.

Alters, B. and Nelson, C. E. (2002). Teaching evolution in higher education. *Evolution: International Journal of Organic Evolution*, *56*(10), 1891-1901.

Anderson, D. L., Fisher, K. M., & Norman, G. J. (2002). Development and evaluation of the conceptual inventory of natural selection. *Journal of Research in Science Teaching*, *39*(10), 952–978.

Bishop, B. A., & Anderson, C. W. (1990). Student conceptions of natural selection and its role in evolution. *Journal of Research in Science Teaching*, *27*(5), 415–427.

Borek, A., McLaren, B.M., Karabinos, M., & Yaron, D. (2009). How much assistance is helpful to students in discovery learning? In U. Cress, V. Dimitrova, & M. Specht (Eds.), *Proceedings of the Fourth European Conference on Technology Enhanced Learning, Learning in the Synergy of Multiple Disciplines* (EC-TEL 2009), LNCS 5794, September/October 2009, Nice, France. (pp. 391-404). Springer-Verlag Berlin Heidelberg.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.) (2000). *How People Learn*. Washington, D.C.: National Academy Press.

DeJong, T. & van Joolingen W. R. (1998). Scientific Discovery Learning with Computer Simulations of Conceptual Domains. *Review of Educational Research*, *68*(2), 179-201.

Dragon, T., Mavrikis, M., McLaren, B. M., Harrer, A., Kynigos, C., Wegerif, R., & Yang, Y.

(2013). Metafora: A web-based platform for learning to learn together in science and mathematics. *IEEE Transactions on Learning Technologies, 6*(3)*,* 197-207 July-Sept. 2013.

Edelson, D. (2001). Learning-for-use: A framework for the design of technology-supported inquiry activities. *Journal of Research in Science Teaching 38*(3), 355-385.

Geier, R., Blumenfeld, P. C., Marx, R. W., Krajcik, J. S., Fishman, B., Soloway, E., et al. (2008). Standardized test outcomes for students engaged in inquiry-based science curricula in the context of urban reform. *Journal of Research in Science Teaching, 45*(8) 922–939.

Kapur, M. (2009). Productive failure in mathematical problem solving. *Instructional Science*. doi: 10.1007/s11251-009-9093-x.

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

Klahr, D. & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science, 12*, 1–48.

Klahr, D. (2009). “To every thing there is a season, and a time to every purpose under the heavens”; What about direct instruction? In S. Tobias and T. M. Duffy (Eds.) Constructivist Theory Applied to Instruction: Success or Failure? Taylor and Francis.

Koedinger, K. R. and Aleven, V. (2007). Exploring the assistance dilemma in experiments with cognitive tutors. *Educational Psychology Review, 19*, 239–264.

Lane, E. (2011). U.S. Students and Science: AAAS Testing Gives New Insight on What They Know and Their Misconceptions. *AAAS: Advancing Science, Serving Society* http://www.aaas.org/news/releases/2011/0407p2061_assessment.shtml.

Linn, M. C. and Hsi, S. (2000). *Computers, Teachers, Peers: Science Learning Partners*. Hillsdale, NJ: Erlbaum.

Mayer, R. (2004). Should there be a three-strikes rule against pure discovery learning? The case for guided methods of instruction. *American Psychologist*, 59, 14–19.

Mulder, Y. G., Lazonder, A. W. & de Jong, T. (2010). Finding Out How They Find It Out: An empirical analysis of inquiry learners’ need for support. *International Journal of Science Education. 32*(15), 1971-1988.

Paas, F.G., Renkl, A., and Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist*, *38*, 1-4.

Quintana, C., Reiser, B. J., Davis, E. A., Krajcik, J., Fretz, E., Duncan, R.G. et al. (2004). A scaffolding design framework for software to support science inquiry. *The Journal of the Learning Sciences*, *13*, 337-387.

Sandoval, W. A. and Reiser, B. J. (2004). Explanation-driven inquiry: Integrating conceptual and epistemic scaffolds for scientific inquiry. *Science Education*, *88*, 345-372.

Slotta, J. (2004). The web-based inquiry science environment (WISE): Scaffolding knowledge integration in the science classroom. In: M. Linn, E.A. Davis, & P. Bell (Eds), *Internet environments for science education* (pp. 203-233). Mahwah, NJ: Erlbaum.

Sulloway, F. J. (1982). Darwin's conversion: The Beagle voyage and its aftermath. *Journal of the History of Biology, 15*(3), 325–396.

Tsovaltzi, D., Rummel, N., McLaren, B. M., Pinkwart, N., Scheuer, O., Harrer, A. and Braun, I. (2010). Extending a virtual chemistry laboratory with a collaboration script to promote conceptual learning. *International Journal of Technology Enhanced Learning*, 2(1/2), 91–110.

van Joolingen, W. R., de Jong, T., Lazonder, A. W., Savelsbergh, E., and Manlove, S. (2005). Co-Lab: Research and development of an on-line learning environment for collaborative scientific discovery learning. Computers in Human Behavior*,* 21, 671-688.

SCIENCE AND TECHNOLOGY PUBLICATIONS