Scaffolding Children's Scientific Data Collection in a Virtual Field

Yongjoo Cho, Thomas Moher, Andrew Johnson
Electronic Visualization Laboratory
Department of Computer Science (M/C 152), University of Illinois at Chicago
851 S. Morgan St. Room 1120 SEO, Chicago, IL 60607-7053, USA
ycho@evl.uic.edu

Abstract. The Field is a simulated virtual environment designed to help elementary school students learn the basic skills of observational science in the context of mathematics and science. Over the last four years, we have iteratively developed technical aids for navigation and data collection based on observations we have conducted in an elementary school. This paper discusses the problems students encountered in the Field, the augmenting tools developed to address the issues, and their impacts to students' behavior.

1. Introduction

The capability to plan and conduct scientific investigations is a core theme among science and mathematics learning standards [1,2,3]. Virtual environments are attractive complements to natural environments as venues for investigation because they can extend access to a broader range of phenomena and allow controlling over complexity. Learning scientists have begun to use virtual environments rich in simulated phenomena as loci for scientific investigations. Some examples are Newton World to teach Newtonian mechanics [4], Virtual Gorilla teaching the habitat and behaviors of Gorilla [5], and Round Earth to teach children the concept of round earth [6].

One of the component skills of scientific investigation is the ability to make observations and collect data related to those observations [1,3]. For the past few years, we have worked with elementary school children as they conducted scientific investigations that required them to undertake exhaustive searches for plants spread over relatively large virtual field. Our motivation for using exhaustive search was that it was among the conceptually simplest sampling regimens. Sampling in large geographic areas is a kind of data collection associated with many scientific, engineering, and agricultural disciplines, but not often practiced in schools. Moreover, this approach allowed students to use a well-practiced skill—simple two-dimensional navigation—as the principal activity within the virtual environments.

From the outset, we recognized the need to provide the students with *scaffolds* [7] to support their investigations. In our initial design of the Field, we provided specific features in the environment to support user orientation and location awareness, and to avoid duplicate observation. We engaged students in classroom discussions of surveying strategies designed to ensure accurate exhaustive sampling. In whole-class discussions, students and teachers collaboratively designed data collection forms for use in the Field. We believed that this

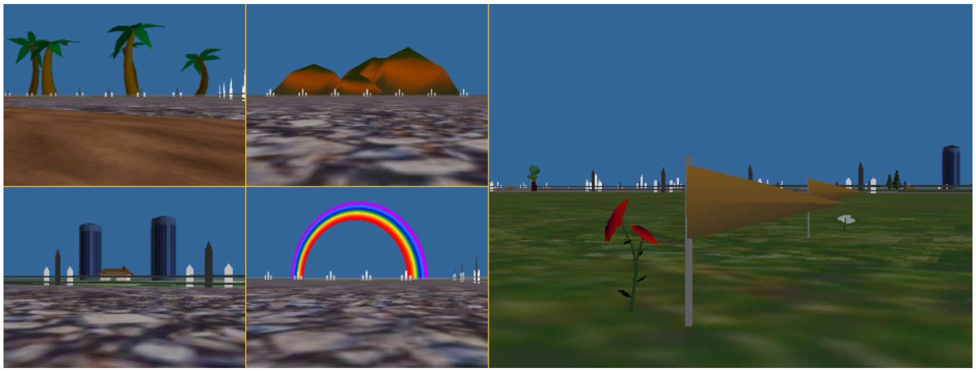

Figure 1. Visual landmarks and flags. The picket fences that separate the sectors are also shown in the pictures. Palm trees, a mountain, a farmhouse, and silos provide directional landmarks.

scaffolding would be adequate to ensure data collection sufficiently accurate to support the larger investigation.

We were wrong. The students ignored the systematic navigation strategies that they had previously discussed, and moved in seemingly random patterns. They had significant problems with orientation and position, resulting in incomplete surveys. They duplicated observations. They lost their survey data while transferring them to the data collection sheets and analysis tools. In our first classroom pilot study, this combination of problems was so bad that only about *half* of the available data were reflected in the final representations of the surveys. The loss of data skewed the results to the point that the modeled mathematical relationships were no longer recognizable, effectively losing the point of the lesson.

This paper describes our attempts to improve upon our initial design and provide more effective data collection scaffolds for learners. We describe the virtual environment—the Field—and the instructional context of its use, outline the classroom-based units that we have conducted, present empirical results of students' data collection, and describe the technological and pedagogical changes resulting from those experiences. Because the various units spanned grade levels and learning goals, it is problematic to attempt to isolate the impact of the extended scaffolding. However, taken as a whole, the results reflect improvements in performance that we believe can be attributed in significant part to those design revisions, and illustrate the complexity of "fine-tuning" scaffolds to support learning in virtual environments.

2. The Field – A Virtual Ambient Environment

The Field is a simulation of a static 3000 foot square natural outdoor world composed of patches of gravel, sand, and grass. Large visual landmarks are located on the four sides of the Field to help navigators maintain their orientation (Figure 1). In order to support users' awareness of their locations, picket fences divide the Field into nine sections in a three-by-three grid. Within this basic environment, we have the ability to insert models of up to eight different types of plants, typically on the order of 20-100 plants per section. By configuring the Field in different ways, we can support a variety of learning activities.

Users are offered limited affordances within the environment; they can navigate the world using a joystick, and they can place "biodegradable" flags at arbitrary locations. The constraints are intentional, designed to simplify the task for young learners, and consistent with the emphasis on observational science articulated in the science standards for elementary students [1,2,3].
In these studies described here, the Field ran on an ImmersaDesk®, a single-screen projection-based active stereoscopic VR system driven by an SGI Deskside Onyx. Head-tracking was disabled to "dampen" the abrupt, exaggerated movements of children; however, hand-tracking was retained to interpret the orientation of the Wanda™, the 3-D joystick device used to effect navigation within the Field.

3. Activity Structures in the Field

Each learning unit that utilizes the Field has the primary purpose of improving students' abilities to conduct empirical investigations. In order to support this goal, we have developed a three-phase activity structure that was used in each of the classroom units.

In the *initial problem introduction and planning* phase, students are introduced to the problem and the goals of the activity, and sometimes given a brief tour of the Field. In a whole-class setting, students and the teacher discuss systematic approaches for conducting exhaustive surveys, with reference to familiar activities such as mowing the lawn. Students are guided to discuss methods for preventing redundant observations of the same plants, leading to the realization of the need for a marking mechanism.

In the *exploration and data collection phase*, students are divided into groups of two to four students. Exploration groups are assigned to one section of the Field and asked to do a complete population survey in the section. Students navigate throughout the section and collect data such as the type or location of plants or the state of environmental variables (e.g., soil salinity or moisture). Within a group, students alternately navigate, take notes, collect data and monitor information available from other scaffolds.

In the *data aggregation and analysis phase*, students reassemble in their classroom and aggregate their collected data to create a single data set for the entire Field. Students manipulate representations of the data manually and with software tools in order to address the problem identified at the outset of the unit.

4. The Use of the Field in Classrooms

We have undertaken several studies that require students to perform exhaustive sampling within a section of the Field (see Table 1). In each case, students were allowed to explore their section of the field until they felt that they had found all of the plants. In 1999, we conducted the first study (*Correlations*) to teach the mathematical concept of co-occurrence to sixth graders [8]. About fifty patches of seven to eight plants in clusters were positioned in the Field. Students used simple scratch paper and pencil for data collection.

The *Snookerpuss* study was used to support a unit on graphing and linear interpolation and extrapolation [9]. Similar to the first study, no special navigation or orientation tools were offered to students. However, the Snookerpuss was different from Correlations study in some important respects. First, only a single kind of plant was used: mushrooms. Second, the terrain was specially configured and played a role in data collection and analysis (students had to find

Table 1. Organized studies with grade levels and special aids for navigation and data collection

Study	Grade	Navigation and Coverage scaffolds	Data Collection
Correlations	6th	Landmarks, Flag	Paper and pencil
Snookerpuss	4th	as above	Paper and pencil
Distributions	6th	as above + Trace-view	Data collection sheets
MyField	6th	as above + Simulated GPS, History-list, Real-time Visualization of found plants	Data collection tool, History-list on a PDA

the terrain type where mushroom grew best). Third, only one of the sections was used for the entire class, which was configured to represent a simulated time series of eight months. As in the Correlations study, simple paper and pencil were used for data collection.

In the *Distribution* study, sixth graders studied the concepts of simple sampling and statistical distribution. In this study, students were asked to perform two phases of exploration. First, students explored a sector configured with only a few instances of each plant type and took snapshots of the plants. They then created a "Field guidebook" that contained pictures and names of all plants available in the Field. In the second exploration phase, students had to conduct the exhaustive survey. Unlike the earlier units when only landmarks provided cues to position, in Distribution study students could see a trace-view on a separate laptop computer. The trace-view displayed all visited areas that students had explored in their sections.

The *MyField* study was designed to help the sixth grades discover causal relationship among salinity, bee, plants, moisture, by exploring the world and finding patterns [10]. The world was populated with only white and red flowers. Students were expected to find fewer and fewer flowers as they traveled from south to north, and to find more red flowers in a certain area. Students were then asked to explore potential reasons of the phenomena observed within the environment. In addition to the trace-view, students had access to a simulated GPS and orientation tool that showed their current location and orientation in real time.

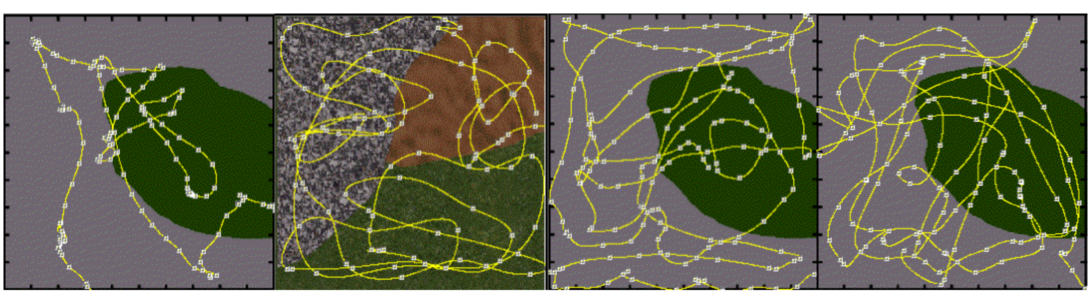

Figure 2. Trace view from Correlations, Snookerpuss, Distribution, MyField, respectively. From the Snookerpuss study, which only used the northwest corner of the Field, the trace-view with highest coverage rate is shown here. The same northwest's trace-view is chosen from all other studies.

5. Scaffolding scientific data collection with supporting technologies

5.1 The Trace-view

Students in the Correlations study showed fairly poor coverage (Figure 2) of the Field, navigating only very small areas of their sections without demonstrating any systematic surveying scheme. The need for feedback on position led us to develop the *trace-view* (Figure 3). The trace-view was designed to help students reflect on their own navigation paths and behaviors within the context of a traversal strategy. In practice, this tool was rather used by students to identify unexplored areas in their sections.

5.2 The Simulated Global Positioning System and Orientation Tool

Although the availability of trace-view improved coverage in later studies, students still suffered from navigation and way-finding problems [11]. For instance, students frequently requested their current locations and orientation in spite of the presence of large visual landmarks and picket fences in the Field. Although the trace-view helped students identify the unexplored areas of a section, locating position became more problematic as the number of trace marks increased. To help students remedy this problem, a *simulated global positioning system tool* was added to the environment (Figure 3).

5.3 Custom Data Collection Forms and Electronic Tools

In earlier studies, the use of paper and pencil (and insufficient attention to planning data collection forms) contributed to the loss of data [8]. Sometimes students forgot to record plants that they had found. They often got confused between two red flowers—roses and tulips—and made mistakes while recording. The Field provided a way to mark the found plants—placing flags—to prevent duplicated counting. However, students simply forgot to place flags or placed them too far from the plant they were intended to mark, which caused duplicated data. In the

Figure 3. The MyField study. Left image showing two children navigating the Field, while another child looks at the simulated GPS and trace view. Right image shows the simulated GPS on a PDA, trace view, and the MyWorld screen showing the found plants.

analysis phase, students spent a great deal of time agreeing upon nomenclature for the plants.

The data loss problem was addressed iteratively. The Field guidebook developed in the Distribution study contained pictures and descriptions of plants available in the Field. The custom-made data collection sheet made out of the Field guidebook also contained colored pictures of plants to help students distinguish each type of flower more easily during exploration. In MyField, a PDA-based electronic data collection tool was provided to further simplify the process (Figure 3); instead of recording on paper, students collected flower data by pressing the appropriate button on the PDA. To reduce duplicate counting, the entire list of collected flowers and their positions were listed on the PDA.

6. Activity patterns during exploration

As an estimate of the degree of coverage during student exploration, each section of the Field was divided into 100 foot square *cells* (100 cells in a section). (The choice of cell size was based on estimates of visual acuity; the presence of plants, if not their type, could be determined from 50 feet away.) Then, each was examined to see if students had passed that bounded area. The overall coverage rate was calculated by determining the percentage of visited cells. The overall coverage rate shown in Table 2 reveals how much space of a section is explored with the collected trace pattern (Figure 3).

From Table 2, the overall coverage rate seemed to improve as more scaffolds such as the trace-view or simulated GPS were added to the environment. At least for this sample, exploration without scaffolds resulted in a coverage rate only about half that of students who used scaffolds. While it is not possible to separate the relative contributions of trace-view and simulated GPS, the small difference between the Distribution (trace-view) and MyField (trace-view + simulated GPS) performances invites further investigation.

The Snookerpuss unit yielded an interesting result. Without using any scaffolding technology, fourth graders were able to achieve almost the same coverage rate as students in the Distribution study. Distinguishing Snookerpuss from the other units was that every group explored the same physical section, which had large, clearly separated areas of grass, sand, and gravel areas (Figure 3). The unique terrain in Snookerpuss allowed students to investigate traversal strategies that had not been previously discussed in class, including a radial strategy in which students placed themselves roughly in the middle of a terrain patch and walked in spoke-like patterns. . While students in the other units also had the opportunity to use terrain as a navigational scaffold, they

Table 2. Organized analyses of each activity.

Study	Overall Coverage rate (%)	Explored Path Length (feet)	Exploration Time (seconds)	Avg. Number of Plants (clusters) in a sector	Average Speed (ft/sec)	Average time to find a plant (Seconds)
Correlations	40.5	4383.14	1335.50	5	3.21	296.78
Snookerpuss	83.83	14292.14	1714.50	24	8.38	72.96
Distribution	85.35	13410.59	1986.65	39	6.75	54.74
MyField	91.44	17115.70	4169.78	60	4.10	75.06

were not required to do so to perform the task, and they did not. .

Another interesting pattern was that students moved considerably slower as more scaffolding was added. The average traversal speed of students in MyField was the lowest except for Correlations. (Students in Correlations had to draw pictures of plants found in the Field so that they could compare them with other groups' collections in a subsequent analysis phase.)

7. **Conclusion**

Navigation in the Field is not like navigation in the real world; the real world has a far richer set of cues for orientation and position. In these studies, elementary school children were able to demonstrate reasonably good performance on a sampling task requiring them to traverse a large simulated area. *A priori* expectations on our part that explicit instruction would lead to the use of orthogonally based traversal patterns (e.g., lawn-mowing patterns) proved incorrect. Instead, they relied on *ad hoc* methods. Two factors gave tentative evidence of scaffolding more mature performance. Designing the task to draw attention to the terrain allowed the students to effectively subdivide the section into smaller, more easily explored regions. The addition of real-time trace and position feedback gave preliminary evidence of improving coverage.

References

[1] American Association for the Advancement of Science (AAAS), *Benchmarks for Science Literacy: Project 2061*. New York, NY: oxford University Press.

[2] National Council of Teachers of Mathematics (NCTM), *Principles and Standards for School Mathematics Electronic Version:* (http://www.standards-e.nctm.org/document/index.htm).

[3] National Science Education Standards, http://www.nap.edu/readingroom/books/nses/html

[4] Dede, C., Salzman, M., and Loftin, B. The Development of a Virtual World for Learning Newtonian Mechanics. In "Multimedia, Hypermedia, and Virtual Reality. Models, Systems, and Applications," P. Brusilovsky, P. Kommers, and N Streitz, Eds., Springer, 1996, pp. 87-106

[5] Bowman, D., Wineman, J., Hodges, L., and Allison, D. Designing Animal Habitats Within an Immersive VE. *IEEE Computer Graphics & Applications*, September/October 1998, vol. 18, no. 5, pp. 9-13

[6] Johnson, A., Moher, T., Ohlsson, S., Gillingham, M., The Round Earth Project: Deep Learning in a Collaborative Virtual World. *Proceedings of IEEE VR. 1999,* pp. 164-171.

[7] Wood, D., Bruner, J. S., Ross, G., The role of tutoring in problem-solving, *Journal of child psychology and psychiatry, Vol. 17, pp. 89-100.*

[8] Moher, T., Johnson, A., Cho, Y., Lin, Y., Observation-based Inquiry in a Virtual Ambient Environment, *Proceedings of the Fourth ICLS (International Conference of the Learning Sciences)* 2000, pp. 238-245.

[9] Moher, T., Kim, J., and Haas, D. (2002). A Two-tiered Collaborative Design for Observational Science Activities in Simulated Environments. *Proceedings of CSCL* 2002 (Computer Support for Collaborative Learning), 2002, pp. 361-370.

[10] Johnson, A., Moher T., Cho, Y., Edelson, D., Russel, E., 'Field' work, ACM SIGGRAPH Educators Forum, 2003.

[11] Vinson, Norman, Design Guidelines for Landmarks to Support Navigation in Virtual Environments, *Proceedings of ACM Human Factors in Computing Systems*, 1999, pp. 278-285