

Design-Based Research Strategies for Developing a Scientific Inquiry Curriculum in a Multi-User Virtual Environment

**Brian Nelson
Diane Jass Ketelhut
Jody Clarke
Cassie Bowman
Chris Dede**
Harvard Graduate School of Education

This National Science Foundation funded project is studying graphical multi-user virtual environments (MUVEs) to investigate whether using this interactive medium in classroom settings can simulate real-world experimentation and can provide students with engaging, meaningful learning experiences that enhance scientific literacy. In the project's River City curriculum, teams of middle school students are asked to collaboratively solve a digital 19th century city's problems with illness, through interaction with digital artifacts, tacit clues, and computer-based 'agents' acting as mentors and colleagues in a virtual community of practice. This article describes the design-based research strategy by which we are currently extending an educational MUVE environment and curriculum. Through several iterations of design-based research, we have refined our curriculum, the MUVE environment, and the theories underlying our design.

Introduction

Scientific literacy—the capabilities (1) to understand the interrelationships among the natural world, technology, and science, and (2) to apply scientific knowledge and skills to personal decision-making and the analysis of societal issues—is a major goal for education in the 21st century (AAAS, 1993; NRC, 1996). Research suggests that, if all students are to become scientifically literate citizens, science

instruction must convey greater engagement and meaning to them. To achieve this, we believe that science instruction in secondary schools should provide students with opportunities to explore the world; to apply scientific principles; to sample and analyze data; and to make connections among these explorations, their personal lives, and their communities. However, given the constraints of classroom settings, real-world data collection and laboratory experiments are often difficult to conduct. It is no surprise, therefore, that science teachers report that teaching higher order inquiry skills, such as

Brian Nelson is a doctoral student at the Harvard Graduate School of Education. He currently studies the theory and design of adaptive guidance systems for collaborative virtual environments. Over the past 15 years, Brian has worked as an assistant professor and researcher in Japan and the USA. Most recently, he taught a course on learning theory and instructional design in the Technology in Education program at Harvard.

Diane Jass Ketelhut is a fifth-year doctoral student in Learning and Teaching at the Harvard Graduate School of Education. Her research interests are in academic self-efficacy and how technology can improve science education. Diane is the Project Director of the NSF-funded MUEVES project. Currently, she is the chair-elect of AERA's Graduate Student Council.

Jody Clarke is a doctoral student in Learning and Teaching at the Harvard Graduate School of Education. She is part of the MUEVES Research Group headed by Dr. Chris Dede. Her research interests include learning with emerging technologies.

Cassie Bowman is a doctoral student at the Harvard Graduate School of Education. For the past six years she has worked on NASA programs pairing high school students with scientists and engineers and is interested in the motivational benefits of such partnerships.

Chris Dede, a Contributing Editor and Guest Editor of this special issue, is the Timothy E. Wirth Professor of Learning Technologies at Harvard's Graduate School of Education. His funded research currently includes a grant from the National Science Foundation to aid middle school students learning science via shared virtual environments with digitized museum artifacts, a grant from the Joyce Foundation to aid the Milwaukee Public Schools in implementing a knowledge portal for teacher professional development, and a grant from Harvard to explore applications of wireless handheld devices in higher education. Chris has completed a co-edited volume on *Scaling Up Success: Lessons Learned from Technology-based Educational Innovation*, to be published by Jossey-Bass this month (e-mail: Chris_Dede@Harvard.edu).

hypothesis formation and experimental design, are among the most difficult challenges they face.

As part of the NSF-funded MUEVES (Multi-User Virtual Environment Experiential Simulator) project, we have created graphical multi-user virtual environments (MUEVs) to enhance middle school students' motivation and learning about science and society (<http://muve.gse.harvard.edu/muvees2003/>). MUEVs are similar to some online multi-player games in that they enable multiple participants to access virtual worlds simultaneously and to interact with digital artifacts. Participants negotiate the worlds through their computerized representations—avatars, interacting with other students and with computer-based agents to facilitate collaborative learning activities of various types. Unlike many online multi-player games, our "River City" MUEV is centered on skills of hypothesis formation and experimental design, as well as on content related to national standards and assessments in biology and ecology.

We are conducting a series of studies to determine if virtual environments can simulate real-world experimentation and can provide students with engaging, meaningful learning experiences that enhance scientific literacy. We are employing a design-based research (DBR) approach to the iterative, formative development of River City and to resolving the scalability issues involved in moving to large-scale implementations. Chris Dede and Kurt Squire describe the theory behind design-based research elsewhere in this issue. In this article, we reflect on our design-based methodology and discuss what we have learned using DBR in several cycles of implementations. By offering a 'glass-box' view into our research strategy, we hope to provide a guide to others interested in design-based research.

The Design

Our goal for River City is to promote learning for all students, particularly those unengaged or low performing. Using an open-ended design, students learn to behave as scientists through collaboratively identifying problems via observation and inference, forming and testing hypotheses, and deducing evidence-based conclusions about underlying causes.

The River City virtual "world" consists of a city with a river running through it; different forms of terrain that influence water runoff; and various neighborhoods and institutions, such as a hospital and a university. The learners themselves populate the city in teams of three, along with computer-based agents, digital objects that can include video clips, and the avatars of instructors (Figure 1). Content in the right-hand interface-window shifts based on what participants interact with in the virtual environment (Figure 2). Chat text and computer agent dialogues are shown in the text box below these two windows; members of each team can

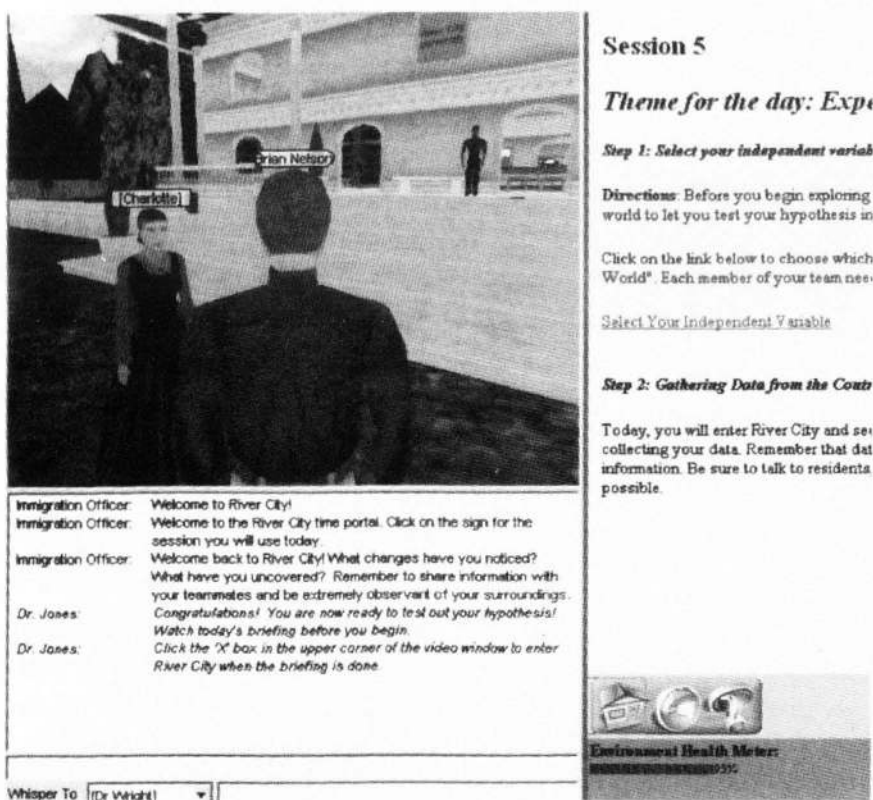


Figure 1. Talking with an agent.

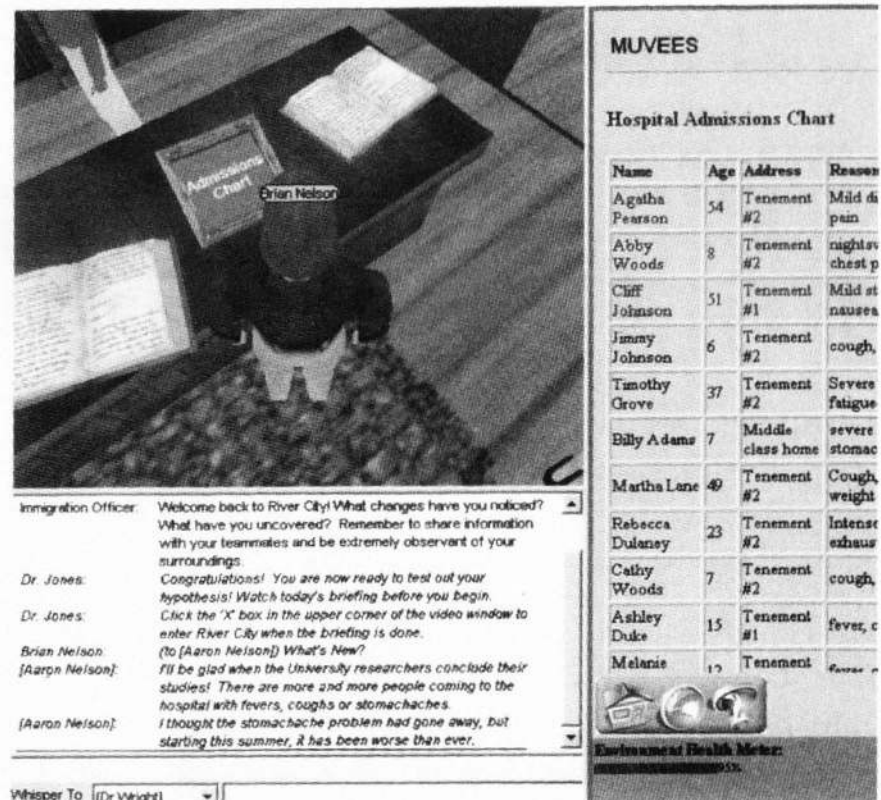


Figure 2. Hospital admissions chart.

communicate regardless of distance, but intra-team chat is displayed only to members of that team.

Students work in teams to develop hypotheses regarding one of three strands of illness in River City (water-borne, air-borne, and insect-borne). These three disease strands are integrated with historical, social, and geographical content to allow students to experience the realities of disentangling multi-causal problems embedded within a complex environment. At the end of the project, students compare their research with other teams of students to discover the many potential hypotheses and avenues of investigation to explore.

Theories Underlying the Design

River City was originally designed as a guided social constructivist environment that allowed students to explore and focus on what intrigued them. The open-ended nature served as a self-motivator. Guidance and support was supplied by the accompanying lab book, team members, and the teacher. Observations of implementations provided evidence as to whether this was effective.

We were also interested in designing a curriculum that would appeal to both boys and girls. Research suggests that middle school is the developmental level where girls tend to lose interest in science (AAUW, 1999; Butler, 2000). Therefore, in our design of River City, we intentionally created a lead female role model for girls. This model, Ellen Swallow Richards, was the first woman to earn a chemistry degree at MIT; she potentially combats stereotypes internalized by young women. Further, research on gender and technology

suggests that girls prefer environments that are collaborative in nature (Clark, 1999). This was an additional support to our decision to design the curriculum around teams such that students work collaboratively to solve the health problems in River City.

Cycles of Implementation, Findings, and Implications

First Cycle

Implementation. In our pilot implementations of River City, using two public school classrooms in urban Massachusetts, we examined usability, student motivation, student learning, and classroom implementation issues (Dede & Ketelhut, 2003). One sixth- and one seventh-grade classroom in different schools with high percentages of ESL students implemented the MUVE-based River City curriculum; two matching control classrooms used a curriculum similar in content and pedagogy, but delivered via paper-based materials.

Using design-based research methods, we collected both qualitative and quantitative data from students and teachers over a three-week implementation period. Both the Patterns for Adaptive Learning Survey (Midgley *et al.*, 2000) and a content test were administered to students, pre- and post-intervention. In addition, demographic data and teachers' expectations of students' success were collected. All teachers responded to a pre- and post-questionnaire regarding their methods and comfort with technology. The

experimental intervention classroom teachers also wrote a narrative at the end of the project about their perceptions of the MUVE. We used this data, plus our own observations, to analyze the learning outcomes for students and to inform our understanding of how the MUVE worked, in order to refine the design of our next iteration and to reflect on the theoretical foundations underlying our design.

Findings. Results indicated that the MUVE was motivating for all students, including students who had been characterized as "low ability," based on grades. For example, six out of seven experimental students who scored in the bottom-third on the content pre-test improved to average or above; however, only two out of five control students who scored in the bottom-third on the pretest moved out of this category. We also found that the ability to engage in inquiry in an authentic setting was powerful for students. They discovered multiple intriguing health problems in the MUVE to investigate. In the seventh grade classroom, five different hypotheses about the health problems emerged, with posited causes ranging from population density to immigration to water pollution. The MUVE seemed to have the most positive effects for students with high perceptions of their own thoughtfulness of inquiry (Dede & Ketelhut, 2003).

In our analysis, we found that gender was consistently not a significant predictor of success in the River City MUVE. However, we did find that on our pretest, six out of our eleven lowest performing students were female. Focused analysis on these six students led to an interesting discovery of the effect of the MUVE. The science self-efficacy of these females, the belief that they could successfully do science, increased 7% over the two-week implementation. Similarly, but with a smaller effect, their motivation also increased.

Implications. Overall, these findings encouraged further refinement and experimentation with curricular MUVES to help teachers reach students struggling with motivation and lack of content knowledge. By examining recorded and observed student interactions with the pilot curriculum, we saw ways to strengthen our content and pedagogy. Although students found the MUVE readily usable and the learning experiences motivating, we found weaknesses in this design, both from a graphical and curricular perspective.

Based on our analysis of the first River City pilot study, we refined the MUVE environment. In the initial environment, computer-based River City 'citizens' recited lines of text repeatedly as students approached. However, students could not interact with the citizens in any way. Based on student feedback, this was changed so that students could ask basic questions such as "What's new?" to the citizens to gather clues about events in the city. Students also gained the ability

to teleport (move instantly) to different locations within the virtual city. The virtual area of River City is quite large, and students expressed the wish to be able to cover more area quickly. Finally, students were given the ability to choose their avatar, to enable more self-expression in the world. From a theoretical perspective, all these were ways of increasing students' psychological immersion in the MUVE, through adding new types of actions, social situations, and participation in the learning environment.

Second Cycle

Implementation. Our first implementation of the revised MUVE curriculum was held in a small focus group in December, 2003; we concentrated our evaluation on the student responses to the new changes: interactive residents, teleporting map, and ability to choose and change their avatar. We observed student interactions and conducted exit interviews with them; we also actively solicited focus group suggestions of changes students would like to see.

Findings. From our observation of focus group participants, we noticed that our three changes seemed to elicit 'ah-ha' moments for students. From observations and these interviews, we further learned:

- Students needed time to experience the world before beginning the formal curriculum. This experience helps them to become immersed in the context.
- Students were confused by the connection and relevance of the digitized Smithsonian artifacts in the world.
- Some students became easily lost in the world.
- Students sought to access the books in the virtual library of River City when they were confused.
- Students wondered why their avatars were not also getting sick.

Implications. Based on this implementation, we concluded that our changes had been positive and should be kept, but additional modifications were needed:

- A reorganization of our lab book to allow students time to learn how to maneuver, and to explore the world.
- A new section to our lab book that guides students in understanding the digital images and artifacts embedded in the world.
- A permanent link on the interface to the interactive map.
- Clickable volumes in the library to allow students to locate background information on disease and on the scientific method to strengthen their learning outcomes.
- A health meter that would rise and fall as students

wandered close to polluted waters or stepped on manure.

Theoretically, many of these changes increase the guidance provided to students as they experience social constructivist learning.

Third Cycle

Implementation. We made these design changes and then conducted two full-scale pilot studies in January and February 2004. These implementations included similar pre- and post-assessments to those used in the first cycle of implementations.

The January implementation was conducted in an informal after-school program, and the February implementation was conducted in a west coast university laboratory school. Both of these represented different populations to our public school populations used in the previous cycle, so we focused our attention to see if our changes worked even as our participants changed.

Findings. As we implemented this new version, we determined that our alterations made significant improvements to the curriculum, resulting in improvements in student engagement and learning outcomes.

- We found that providing time for initial exploration of the environment resulted in students being immediately engaged. They used this time both to become comfortable with the MUVE interface and to start understanding what problems existed in River City. When we handed out the lab book on the second day, the students used this to guide their investigation more readily than they had previously.
- Prior to creating the new lab book section on artifacts, we found that students were likely to primarily rely on computer-based agents to understand the problems in River City. Since much of the curriculum is attached to embedded artifacts, students were limiting their understanding. After creating this section, students increased their interactions with the digital pictures, thus increasing their involvement with the curriculum.
- Students found that the teleporting map facilitated finding where they were or where they wanted to go; this increased their mobility and allowed students to access more of the curriculum than previously. However, students still complained that it was difficult to locate themselves on the map.
- Once the students discovered that they could "find answers" in the library using the new dictionary, encyclopedia, primers on microbes

and scientific method and algebraic concepts (which to many middle-school students was another ah-ha moment!), it became a popular place to get more information.

- Students used the health meter as an additional source for data in their experimentation. As they walked through the world, the movement of the meter intrigued students enough that they explored ways to make it change. As a result, students realized where the pollution and contamination in River City was at an earlier stage.
- Teachers commented that it would be great if students could actually conduct experiments in the world.

Implications. Actual experimentation in the world had previously been technologically impossible. Improvements to the technology now made that a possibility. Given our emphasis on authentic learning, this modification was made in the next cycle of implementations.

In our initial implementations, we constantly evaluated the appropriateness of our underlying pedagogical theory of constructivism. As our design became stronger, we turned our attention to evaluating our pedagogy. In River City, students are immersed in conducting an authentic task, similar to 'learning on the job.' This seemed more similar to situated learning than constructivism. Situated learning requires real-world contexts, activities, and assessments coupled with guidance based on expert modeling, situated mentoring, and gradually increasing participation. MUVEs are a promising medium for creating and studying situated learning because they can support immersive experiences (incorporating modeling and mentoring) about problems and contexts similar to the real world. Based on the previous implementations, River City was redesigned to allow comparison of situated and constructivist learning theories. Both in and out of MUVEs, insights obtained by this comparison may enhance educators' and researchers' understanding and application of learning theories and may increase students' abilities to transfer knowledge from academic to real-world settings. Consequently, in our next iteration, we extended our MUVE curriculum to compare other learning theories to guided social constructivism. A major benefit of a DBR approach is that it promotes evaluation and redesign of the underlying theory.

Fourth Cycle

Implementation. Based upon what we learned from the first pilot study, we developed two variations of the River City curriculum. Variant GSC centers on the original guided social constructivist (GSC) model of

learning, in which guided inquiry experiences in the MUVE alternate with in-class interpretive sessions. Variant EMC shifts the learning model to center on expert modeling and coaching (EMC), with expert agents embedded in the MUVE and experts collaborating with teachers in facilitating the in-class interpretive sessions. Our third "control" condition utilizes a curriculum in which the same content and skills are taught in equivalent time to comparable students in a paper-based format without technology, using a guided social constructivist-based pedagogy. Where possible, teachers offer both the experimental and the control curricula.

To control for threats to validity, both variants were randomly assigned to students within a single classroom, with teachers instructed to minimize cross-contamination of treatments. We also created approximately eight hours of professional development for teachers, focused on content review, alternative pedagogical strategies based on different theories of learning, facilitation strategies while students are using the MUVE, and interpretive strategies for leading class discussions. This was designed as a direct response to teacher feedback in the first series of implementations.

As a result of the previous pilot studies and attendant refinements, we scaled up our implementation of the River City curriculum in spring 2004 with eleven teachers and more than 1000 students spread over two states and three school districts.

Findings. We are now in the midst of analyzing data from this implementation and early results are promising. Preliminary findings show that both students and teachers were highly engaged. All of the teachers who responded to the post-implementation survey said they would like to use the River City curriculum again. In interviews and focus groups, students said they 'felt like a scientist for the first time' and asked when River City would be available for purchase. In some of the urban classrooms in the Midwest where low attendance and disruptive behavior are daily struggles for teachers, we found that student attendance improved and disruptive behavior dropped during the three-week implementation.

Interesting patterns are emerging about which students did best under our various pedagogical conditions. More specifically, of the nearly 300 students who have been analyzed to date, students in the two experimental treatments improved their biological knowledge by 32% for GSC and 35% for EMC. Control students also improved, but by only 17%. Improvements were also seen across the board for knowledge and application of scientific processes. In this case, the control students improved slightly more than the other two groups: 20% for the control, 18% for the GSC group, and 16% for the EMC group.

Given the complexity of the MUVE environment, we

are looking at multiple measurements of student learning. For example, after conducting their experiment, students are asked to write a letter to the mayor explaining their experiment, findings, and recommendations. Preliminary analysis of students' written letters to the mayor of River City suggest that students demonstrate an understanding of the process of the scientific method that was not well captured in the science inquiry post-test measures. For example, students who scored low on the science inquiry post-test wrote letters that were of similar quality to those written by students who scored higher on the post-test. Interestingly, more of the lower-performing test students met the criteria of providing suggested interventions or further research than students who scored higher on the inquiry test questions. This suggests that the complexity of the MUVE treatment creates intricate patterns of learning more appropriately measured with an authentic activity, such as writing an experimental report.

To assess the success of our changes to the environment, we asked students about the tools and avatar choices. Students were overwhelmingly positive about both, listing their favorites amongst the choices of avatars. However, uninitiated, they again mentioned the desire to see themselves on the map of River City.

Implications. Based on our interesting findings comparing GSC and EMC, we will be adding a new theoretical treatment in our next implementation. Variant LPP shifts the learning model to focus on legitimate peripheral participation, in which the entire community of practice in the MUVE works on problem-solving, and students learn more from observation of somewhat more advanced participants (avatars, computer-based agents) than via direct guidance by experts.

For this fourth cycle, we expanded the capabilities of our water sampling station to allow students to take random water samples. Students are now able to click on any of fourteen water stations and bring up an image similar to what they might see in a modern day microscope (see Figure 3). They can now take multiple samples and test for bacteria in the water as a scientist would in the real world. Given the success of this tool on students' engagement, we are currently developing three other tools that help students conduct tests related to the other diseases: a mosquito catcher, blood tests, and throat swabs. We plan to explore the effect of tool use and ability to conduct tests on students learning and feeling like a scientist in future implementations.

Based on student feedback and improvements in the technology, the map will now track individual student movement in the world. This will allow students to 'see' themselves and researchers to track student exploration. This will help us gain understanding of how students' interactions in the world affect their learning.

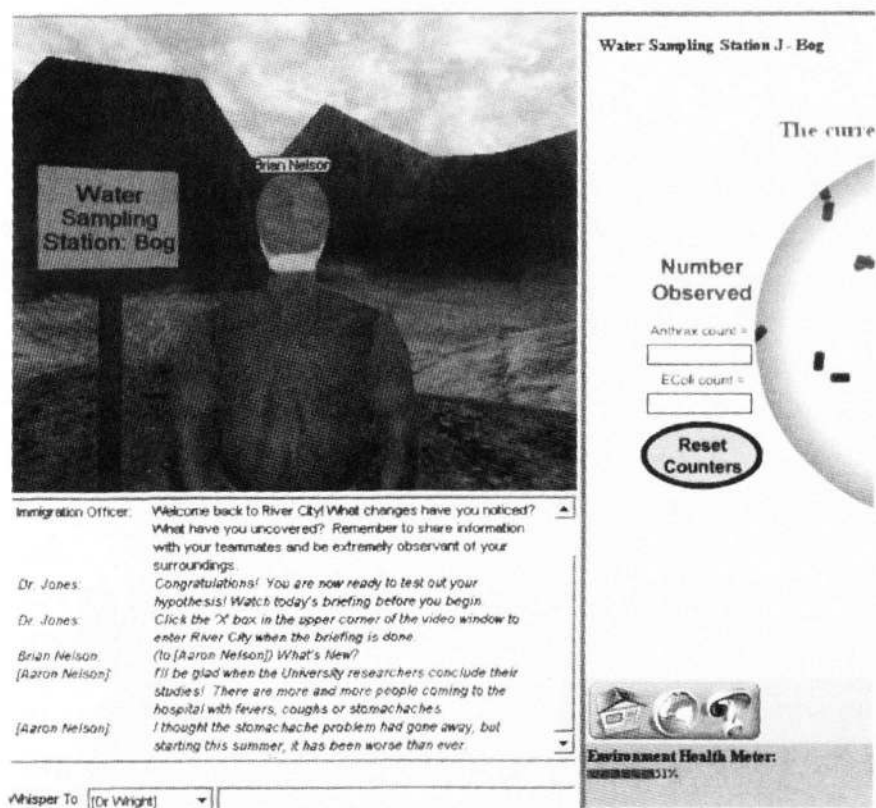


Figure 3. Taking a water sample.

Assessment of learning continues to be problematic for us. Given our differential pattern of results from qualitative and quantitative sources, we are continuing to look into better ways at the end of our unit to measure all the various types student learning we see in classroom observations.

As a result of teacher request, and because scientific inquiry is difficult to enact in a classroom, we created an extensive online professional development for teachers. However, we found that very few teachers interacted with the online materials. We are redesigning our professional development, integrating face-to-face meetings with online resources and follow-up sessions. We are also including a more extensive section on the teacher's role in River City and how to facilitate whole class discussions that wrap around the activities.

Implications for Practice, Policy, Design, and Theory

An important emphasis in our research is to increase student motivation, self-efficacy, and scientific literacy. In particular, educators need help in engaging and teaching subpopulations of learners with special needs, students unmotivated by standard instructional approaches, and pupils with learning styles more visual and kinesthetic than symbolic and auditory. In a typical middle-school classroom faced with a diverse set of learning styles, the teacher must alternate pedagogical strategies to aid each of these. Even under the best of circumstances, at any single moment some students' learning preferences block them from understanding

the lesson. In a MUVE, students can individualize their learning based on their own styles. Our environment/curriculum is targeted specifically to narrowing the gaps among students by helping all learners reach their full potential, especially students who are currently underperforming because of how they are taught in conventional classroom settings. DBR methodologies are providing a way of identifying which elements of our curriculum and MUVE environment are best suited to this goal.

By engaging in DBR-inspired cycles of design, implementation, analysis, and redesign, we have been able to refine both our curriculum and the MUVE environment prior to conducting formal randomized experimental trials. In each implementation cycle, quantitative data are revealing findings about the relative effectiveness of learning theories as instantiated in our environment/curriculum. Qualitative data are providing insights about the reasons underlying those comparative differences. In addition, our quantitative studies, including the use of a control curriculum, are helping us determine whether the leverage for learning and engagement provided by our work are substantial enough to merit moving beyond DBR to large-scale experimental research on implementation.

Dede (2004) states that an important challenge in design-based research is determining what constitute reasonable criteria for "success" in declaring a design finished. After several study iterations, substantial parts of the design have remained relatively unchanged from the initial implementation because our analysis indicated these were successful in meeting our objectives. Other parts of the design have changed based on feedback from the initial studies. For example, identity plays an important role in learning (Lave & Wenger, 1991), and research on identity in virtual environments suggests that females like to play with their identity (Turkle, 1997). While we improved our design to allow students to select different types of avatars, some students from our February implementation wanted to be able to design their own avatar, rather than choose a pre-designed one. While we do not expect boys and girls to have the same experience in River City, we would like them to have equally satisfying experiences. Our findings consistently show this to be true. Therefore, in our case, the ability to create one's avatar does not appear to be a condition for success. Thus, we have determined that aspect of our design is finished.

Conclusion

We believe that this type of controlled evolution of DBR is important to its acceptance as a legitimate methodology by the conservative end of the scholarly community in education. We hope to contribute to the field and legitimacy of DBR by sharing our

development strategy in the context of our MUVE science curriculum. □

References

- American Association for the Advancement of Science. (1993). *Benchmarks for science literacy*. New York: Oxford University Press.
- American Association of University Women. (1999). *Gender gaps: Where schools still fail our children*. New York: Marlowe.
- Butler, D. (2000). Gender, girls, and computer technology: What's the status now? *The Clearing House*, 73, 225–230.
- Clark, J. V. (1999). *Minorities in science and math*. Columbus, OH: ERIC Clearinghouse for Science, Mathematics, and Environmental Education.
- Dede, C. (2004). If design-based research is the answer, what is the question? *Journal of the Learning Sciences*, 13(1), 105–114.
- Dede, C., & Ketelhut, D. (2003). *Motivation, usability, and learning outcomes in a prototype museum-based multi-user virtual environment*. Presented at American Educational Research Conference, April, 2003.
- Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. New York: Cambridge University Press.
- Midgley, C., Maehr, M. L., Hruda, L. Z., Anderman, E., Anderman, L., Freeman, K. E., Gheen, M., Kaplan, A., Kumar, R., Middleton, M. J., Nelson, J., Roeser, R., & Urdan, T. (2000). *Manual for the patterns of adaptive learning scales (PALS)*, Ann Arbor, MI: University of Michigan.
- National Research Council. (2000). *How people learn: Brain, mind, experience and school: Extended edition*. Washington, DC: NRC.
- Turkle, S. (1997). *Life on the screen: Identity in the age of the Internet* (1st Touchstone ed.). New York: Simon and Schuster.
-