A Multi-user Virtual Environment for Building Higher Order Inquiry Skills in Science

Diane Jass Ketelhut, Chris Dede, Jody Clarke
Harvard University Graduate School of Education
Brian Nelson, Arizona State University

For information, please contact: Diane@alumni.brown.edu

Abstract
        This National Science Foundation funded study is investigating novel pedagogies for helping teachers infuse inquiry into a standards-based science curriculum. Using a Multi-User Virtual Environment (MUVE) as a pedagogical vehicle, teams of middle school students are asked to collaboratively solve a simulated 19th century city's problems with illness, through interaction with each others' "avatars", digital artifacts, tacit clues, and computer-based "agents" acting as mentors and colleagues in a virtual community of practice. This paper describes the results from the three implementations in 2004 with approximately 2000 students from geographical diverse urban areas. Results indicate that students do conduct inquiry and are motivated by that process. However, results from assessments vary depending on assessment strategy employed.

Problem
        For decades, science educators have worked to infuse inquiry into the K-12 curriculum (AAAS 1990, 1993; NRC, 1996). For example, the National Science Teachers Association recently issued a draft position statement recommending the use of science inquiry as a method to help students understand the processes and content of science (National Science Teachers Association, 2004). This goal is problematic for teachers when juxtaposed with requirements of preparing students for the detailed science content included in high stakes testing; in many situations, this competing push forces the emphasis in science classrooms to change from inquiry-based instruction to test-preparation (Falk & Drayton, 2004). Curricula centered on both inquiry and coverage of state and national content standards would help teachers achieve both objectives. In this paper, we provide an overview of an NSF-funded curriculum project that focuses on both inquiry and standards-based content, using novel pedagogies to help low-performing students master complex inquiry skills.
        However, curricula such as this only partially solve the problem. In addition, in order to provide teachers and schools with incentives to cover inquiry skills as well as factual content, high-stakes tests would need to include more inquiry-based questions. Unfortunately, this solution raises a different concern: Can learning from good inquiry-based projects be adequately assessed using a standardized test format? What kind of assessments will allow valid inferences about whether a student has learned how to engage in inquiry, particularly in the "front end" inquiry processes used to derive a strategy for making sense out of complexity: problem finding, hypothesis formation, experimental design? In this paper, we will discuss this conundrum and present results from our implementations that shed light on it.

Theoretical Underpinnings

Inquiry

        What is "inquiry?" The range of possible responses to this question is large. Some refer to inquiry as a set of process skills that include questioning, hypothesizing and testing while others equate it to "hands-on" learning. The National Science Education Standards (NSES) define

scientific inquiry as "the diverse ways in which scientists study the natural world and propose explanations based on the evidence derived from their work…also …the activities through which students develop knowledge and understanding of scientific ideas, as well as an understanding of how scientists study the natural world" (National Research Council, p 23). The standards go on to explain that scientific inquiry is:

> a multifaceted activity that involves making observations; posing questions; examining books and other sources of information to see what is already known; planning investigations; reviewing what is already known in light of experimental evidence; using tools to gather, analyze, and interpret data; proposing answers, explanations, and predictions; and communicating the results. (National Research Council, 1996, p 23).

This description should be relatively straightforward to follow in the classroom, however, it becomes increasingly problematic as teachers attempt to infuse coverage of mandatory content with inquiry, since active learning by students is much more time consuming (yet more effective) than passive assimilation. Additionally, responses to an NSTA position paper indicate that many teachers are unclear as to how to implement inquiry in the science classroom (National Science Teachers Association, 2004). Some presume that traditional "cookbook" experiments promote inquiry learning for students (Wallace & Louden, 2002).

River City, a MUVE

Our project studies how a technology-intensive learning experience that implements problem-based inquiry science curricula can provide both deep inquiry skills and content coverage. In particular, we are working to dramatically improve the educational outcomes of the bottom third of students, pupils who even by middle school often have given up on themselves as learners. These students are disengaged from schooling and typically are difficult to motivate even by good teachers using conventional inquiry-based pedagogy. We are investigating whether educational Multi-User Virtual Environments (MUVEs), which resemble the entertainment and communication media students use outside of school, can reengage them in learning. MUVEs enable multiple simultaneous participants to access virtual contexts, to interact with digital artifacts, to represent themselves through "avatars," to communicate with other participants and with computer-based agents, and to enact collaborative learning activities of various types. This last we use to create a community of inquiry learners.

Our "River City" MUVE is centered on the NSES inquiry skills listed above, as well as on content related to national standards and assessments in biology and ecology. The virtual "world" consists of a 19th century city with a river running through it, different forms of terrain that influence water runoff, and various neighborhoods, industries, and institutions such as a hospital and a university. The students themselves populate the city, along with computer-based agents, digital objects that can include audio or video clips, and the avatars of instructors (Figure 1). Content in the right-hand interface-window shifts based on what the participant encounters or activates in the virtual environment (Figure 2).
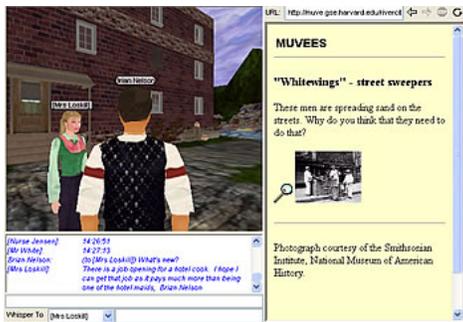
Figure 1

Figure 2

Inquiry and River City

In River City, students engage in all aspects of inquiry as defined by the NSES. These aspects are listed below, and we have mapped each onto where in River City the behavior can be observed:

1. "Making observations" – students move around the world, making visual and auditory observations about the city and its inhabitants.
2. "Posing questions" – students can ask a question of the computerized residents of River City and elicit information that often offers a clue about the problems.
3. "Examining books and other sources of information to see what is already known" – students can access information from books in the River City library as well as from guidance hints, embedded clues in digitized historical images, and the hospital admissions record.
4. "Using tools to gather, analyze, and interpret data" – students can gather data from two tools: a water sampling tool and a 'bug-catching' tool (see figure 2). Each tool is activated by a student click to draw a sample; the student then counts bacteria in a microscope-like screen.
5. "Planning investigations"—students are guided through a generalized process of the scientific method, culminating in creating a unique experiment to test their hypothesis about the problems in River City. They not only design the hypothesis, but also the procedure and data-collection methodology.
6. "Reviewing what is already known in light of experimental evidence"—students spend over a week in River City gathering evidence on the problem from multiple sources, including embedded experts in the form of hospital doctors and university researchers, prior to conducting their own experiments. They must use that information to design their experiments. Once they have analyzed the results of their own experiments, they must compare it with what they hypothesized earlier and to what the embedded experts told them. This process is made transparent in the performance assessments.
7. "Proposing answers, explanations, and predictions"—students create a hypothesis based on collecting evidence to predict what they think is causing a piece of the problem in River City. They re-evaluate that hypothesis in the light of the results of their experiment.
8. "Communicating the results"—at the end of the project, students take part in a classroom-based research conference, delineating their thinking, experiment and results. During this conference, students try to piece together all student results to understand the larger picture of what is making everyone ill in River City.

Students work in teams to gather data, develop hypotheses regarding one of three strands of illness in the town (water-borne, air-borne, and insect-borne) and then to test their hypothesis. These three disease strands are integrated with historical, social and geographical content, allowing students to experience the inquiry skills involved in disentangling multi-causal problems embedded within a complex environment. After testing their hypothesis, students analyze their data using graphs and tables and then write an authentic lab report on their findings in the form of a "Letter to the Mayor of River City." Finally, at the end of the project, students compare their research with other teams of students in their class to delineate the many potential hypotheses and causal relationships embedded in the virtual environment.

Design and Procedure

Research questions

The research questions on which this analysis is centered are:

1. Do students engage in inquiry (as defined by the NSES) in River City?
2. When compared to the "control" version, what types of significant gains in affect and learning for both content and inquiry do versions GSC, EMC and LPP produce?
3. How do results on inquiry learning compare between standardized type testing and performance assessments?
4. What are MUVEs' strengths and limits in facilitating classroom-based inquiry learning?

Sample

This paper will look at results from large-scale implementations of approximately 2000 students in major urban areas in New England and the Midwest as well as one mid-atlantic suburban area; schools in these areas had high proportions of ESL and free-and-reduced-lunch pupils. A total of 8 schools, 12 teachers and 61 classes are involved in this analysis.

Procedures

In order to explore the type of learning best supported by MUVEs used for inquiry learning, we developed three variations of the River City curriculum for these implementations. Variant GSC centers on a guided social constructivist (GSC) model of learning-by-doing, in which guided inquiry experiences in the MUVE alternate with in-class interpretive sessions led by the teacher. Variant EMC shifts the learning model to a situated pedagogy with expert modeling and coaching (EMC) based on expert agents embedded in the MUVE. Finally, variant LPP also uses a situated learning model but based on a community of practice. These three River City variants were compared to a "control" condition that utilized a paper-based curriculum in which the same content and skills were taught in equivalent time to comparable students without using computers, via a guided social constructivist-based pedagogy. The control curriculum (EI) included features similar to River City, such as a historical scenario and unknown disease transmission. In addition to experimental design and analysis, this curriculum also included physical experimentation. This type of control curriculum enables us to focus on the strengths and limits of MUVEs, as well as the types of pedagogy best supported by this medium.

The three computer-based variants (GSC, EMC and LPP) were randomly assigned to students within each classroom, with teachers instructed to minimize cross-contamination of treatments. Some implementations only had two of these three variants assigned. The paper-based control treatment was randomly assigned to whole classes. Each teacher offered both the computer-based treatments and the control.

River City incorporates an underlying database that captured individual student activity in the MUVE with a timestamp, allowing us to analyze students' microbehaviors throughout the implementation. After designing and conducting their experiments, students in both the control and River City treatments were asked to write letters to the Mayor of River City in which they discussed their hypothesis, experimental design, results and recommendations for solving the city's health problem.

Both qualitative and quantitative data were collected from students and teachers over the three-week implementation period. Pre- and post-intervention, the students completed an affective measure that was adapted from three different surveys, Self-Efficacy in Technology and Science (Ketelhut, 2005), Patterns for Adaptive Learning Survey (Midgley, C. 2000), and the Test of Science Related Attitudes (Fraser, 1981). This modified version has scales to evaluate students' efficacy of technology use (videogame, computer, chat, etc), science self-efficacy, thoughtfulness of inquiry, science enjoyment, and career interest in science. To assess understanding and content knowledge (science inquiry skills, science process skills, biology), we administered a self-designed content test, (with sections modified from Dillashaw and Okey, 1980), pre- and post-intervention. This content test was redesigned after the first implementation and thus those results will not be compared on that measure.

We conducted semi-structured interviews with a subsample of students pre-, during, and post-intervention. The students were chosen by their teacher and represented both low and high achievement. Interviews were conducted in the school during the students' free period. All interviews were audio or video recorded and transcribed verbatim.

To support teachers, we created a professional development program with some online portions, 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. The teachers collected demographic data and rated their expectations of students' successes and motivation with the project. Teachers responded to a pre- and post-questionnaire regarding their methods, comfort with technology, and reflections on using the MUVE in their science class.
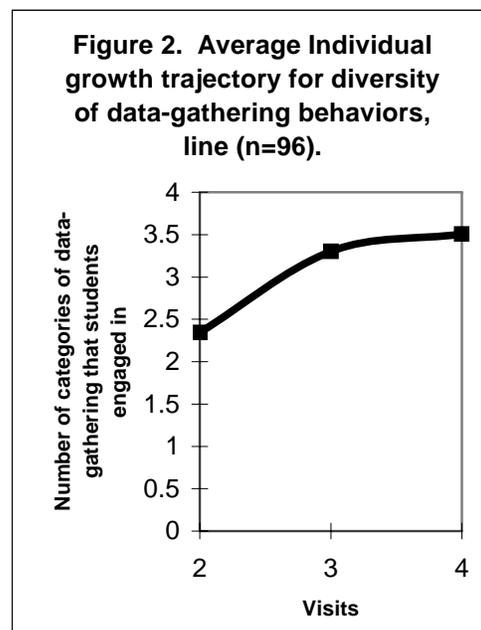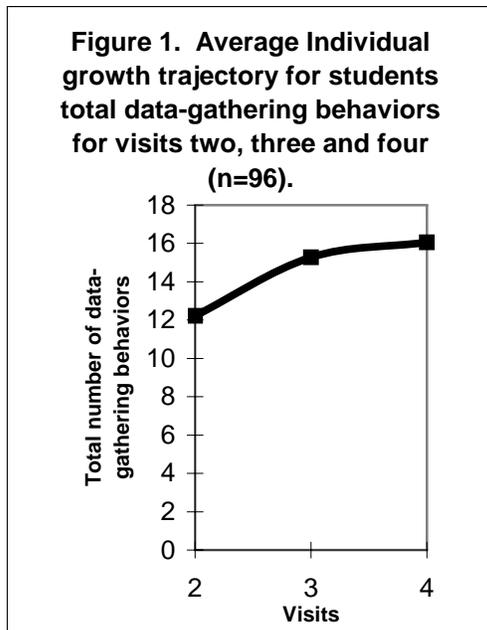
Findings

The quantitative data was analyzed with SAS. Descriptive statistics, correlations and multi-level regression models were run. A significance level of $p \leq .05$ was used, unless otherwise indicated; and checks for linearity, normality and homoscedasticity were performed at intervals. No clear violations were noted.

Inquiry Engagement:

Previously, research on individual student inquiry behaviors would have required a nearly one-to-one ratio between observer and student in order to document what processes students went through while engaged in an activity. As technology has advanced, researchers have utilized technologies such as video recording to begin to capture microbehaviors from larger groups of students. Now, many forms of interactive technology, including MUVEs, incorporate a database that records all student utterances and interactions that take place within the environment in which the student is learning. An example of this level of detail allows us to answer our first research question. The first four inquiry behaviors listed by the NSES related directly to student actions in River City. To understand whether students were engaged in scientific inquiry and in what way, we analyzed their data-gathering behaviors as shown in the database. First, we were interested in

understanding whether they were engaged. Figure 1 shows the average trajectory of total data-gathering behaviors for a student across the three main data-gathering visits (visits 2-4) to River City of a subsample of our students. As can be seen in this, students initially show in visit 2 an average of 12 data-gathering behaviors, which rises to close to 16 by the fourth visit.

Figure 1. Average Individual growth trajectory for students total data-gathering behaviors for visits two, three and four (n=96).

Figure 2. Average Individual growth trajectory for diversity of data-gathering behaviors, line (n=96).

We were also interested in whether students used a single source of data to base their experiments on or whether they used more than one, indicating an informal triangulation in their data gathering. There are a total of eight sources that students could gather data from: observations, hospital admissions record, talking to residents of River City, River City library books, guidance hints, clues in the embedded digital artifact, water and bug sampling stations. As can be seen in Figure 2, students in this subsample begin gathering data from at least 2 sources, on average, increasing to nearly 4 by the fourth visit. Thus, the technology afforded us in a MUVE allows us to confirm that students are engaged in scientific inquiry behaviors and choose to vary how they gather data across the different sources; in addition, they continue to increase their commitment throughout the data-gathering period.

To confirm the data from the database, we also analyzed the results of interviews and focus groups, looking for evidence of inquiry. Many students claimed they felt like a scientist for the first time in science class because they were "doing tests and stuff to see what was causing the sickness" (Clarke and Dede, 2005). River City has an online microscope and a bug catcher tool that students can use to take water samples and count the number of mosquitoes in an area. These tools helped students feel like they were "actually conducting an experiment." Having to come up with a hypothesis and design an experiment was motivating. Being able to "pretend to be a real scientist" allowed some students to take on a new identity as an effective science learner. Students seem to enjoy the inquiry pedagogy and like that it is "more independent working…rather than having him instruct us and telling us what to do and guiding us." They claim, "it was different by exploring by myself not being told what things to test out." According to one student, "when I was making the experiment and going around asking everything I kind of felt like a detective." Many students said that they liked the fact that it was more "difficult" and "more challenging" than their regular science class. Having to solve the problem and "figure out" why people were getting sick made students
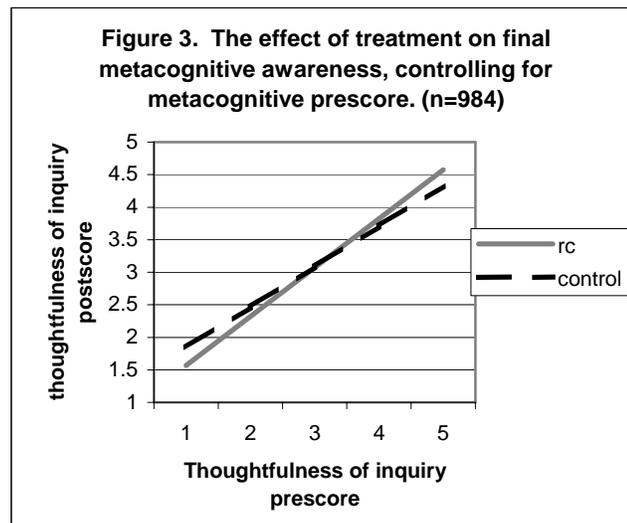
"think more" and as a result, learn more. One student claimed, "we had to figure out things and ask questions and use our brains and think really hard... because we had to figure out what was wrong."

Affective results:

For some of our implementation sites, the question of pedagogy and curriculum is meaningless as many students are rarely in class to experience it. In some of those classrooms using this technology project to deliver scientific inquiry-based curriculum, we found that student attendance improved and disruptive behavior dropped during the three-week implementation (Nelson et al, 2005).

We were also interested in characteristics that promote scientific interest and inquiry. For example, on our affective measure test, we measured *thoughtfulness of inquiry,* a measure of metacognitive awareness. This construct is important in conducting scientific inquiry, as students need to be able to reflect on their findings in order to make predictions that are evidence-based and to draw conclusions. For the first implementation, student scores on this subscale on the post-survey were significantly higher ($p<.01$) on average for River City students, in comparison to the scores for students in the control group. For example, River City students scoring an average of 1 (strongly disagree) on the scale of 1-5 for the pretest were associated with scores of 1.8-1.9 on the post test, nearly double their starting average score. Students in the control group also improved, on average, but only to 1.3.

Figure 3 shows the average results for the second and third implementations. For these implementations, this picture is a bit more complicated as student scores on the post survey depended on their pre-survey scores. As can be seen here, all students with low pretest scores in *thoughtfulness of inquiry* improved on the posttest survey, however the control students improved slightly more than the River City students. This effect changes as the pretest score increases with River City students with average or above average pretest scores on *thoughtfulness of inquiry* scoring better than the control students on the post survey.

Figure 3. The effect of treatment on final metacognitive awareness, controlling for metacognitive prescore. (n=984)
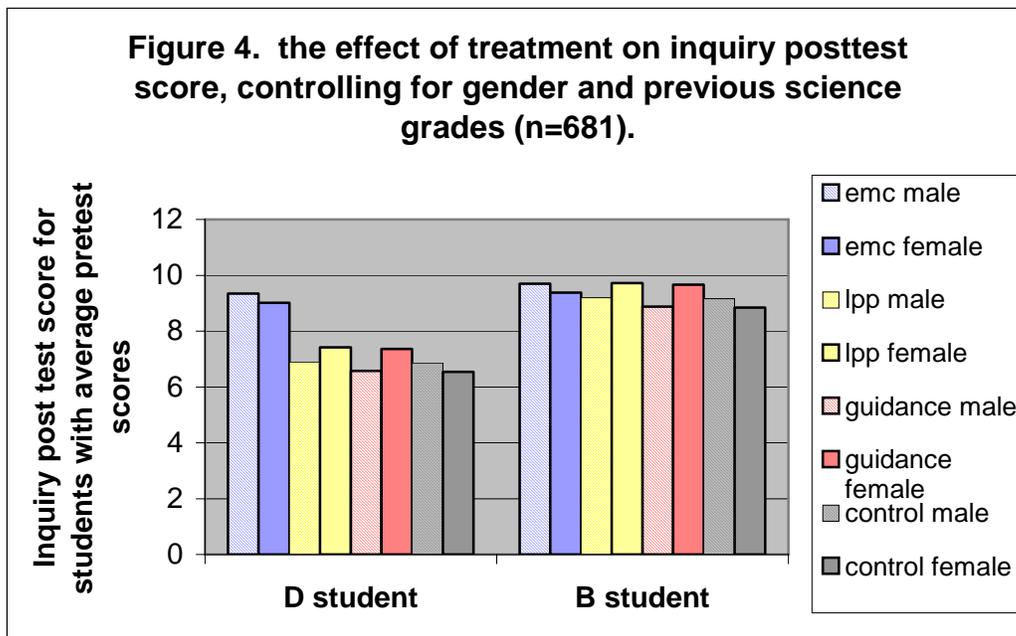
Another subscale measured *interest in a scientific career*; the gain in interest in science careers was 5% higher for students who had taken part in the River City curriculum than for those who had completed the control curriculum—a substantial gain for a 3-week implementation!

Biology Content Results:

We designed River City to help students learn standards-based content as well as scientific inquiry. Were we successful? These results are a bit more equivocal. Of the nearly 300 students who have been analyzed to date in implementation one, students in the River City experimental treatments improved their biological knowledge by 32%-35%. Control students also improved, but by only 17%. However, in implementation two and three, we saw little differences between the treatments and little growth over the course of the project. We are hoping that ongoing analysis of more recent implementations will clarify this.

Inquiry Content Results:

Our second and third research questions for these implementations revolve around whether using a technology-based inquiry project could improve inquiry learning for students, and whether the method of measurement gave different answers to that question. When using survey questions to assess inquiry, we found equivocal results. In our first implementation, improvements were seen across the board for knowledge and application of scientific processes; 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. The results for the second and third implementation showed an additional difference by gender. Figure 4 shows the results for students' inquiry scores on the standardized test-like post survey. The different colors represent the different treatments, with gray indicating the control curriculum. The GSC curriculum did not differ from the control and is not shown. The solid shaded bars represent female student scores with the slashed bars representing male student scores. Overall, students in River City treatments outperformed students in the control treatments. However, in addition to that there were two other interesting effects seen. First, for all treatments except for LPP, boys outperformed girls on the inquiry survey questions. Interestingly, the treatment that was based on a community of practice model better supported girls as indicated by their higher scores for that treatment only. The second interesting piece of this was that while most treatments maintained the outcome differences expected between students with low versus high previous histories in science (as indicated by their science grades), students in the EMC treatment did nearly equally as well regardless of their entering grades in science.

Figure 4. the effect of treatment on inquiry posttest score, controlling for gender and previous science grades (n=681).

Since we wondered how difficult it was to measure inquiry with a multiple-choice test, we also analyzed students' "Letter to the Mayor" for evidence of inquiry elements. In our first implementation, instructions varied somewhat between the River City curriculum and the control curriculum; as a result, detailed comparison of the letters between treatments for this implementation may not be productive. Therefore, we looked for similar demonstrations of student understanding of the processes of inquiry and for motivation. The letters written for the control curriculum often: were much shorter in length, did not demonstrate motivation or engagement, did not mention the experiment, and did not explicitly recognize the interconnectedness of the chosen problem with other possible causes of the larger problem. Analysis of the letters' evidence of inquiry found that students taking part in the MUVE-based curriculum earned scores more than double that of their paper-based control peers, on average (p<.01).

For the next implementations, the instructions were identical which allowed more detailed comparison between the letters. Results are shown in Table 1. This table shows various aspects of inquiry on which the letters were scored. "✶" indicates an area that students in one treatment scored significantly higher than students in treatments marked "-". As can be seen in Table 1, students in the guided social constructivist (GSC) treatment had higher scores in nearly every category, whereas students in the control treatment did not do significantly better on any aspect of the letters to the mayor than did the River City treatment students.

Table 1. Coded areas of the "letters to the mayor" that showed significant differences (p<.05) by treatment in student scores relative to one or more of the other treatments (n=173).

Areas that differed significantly by treatment (p<.05)	GSC	EMC	LPP	Control
Overall quality	✶	—	[gray]	—
Summarizing the problem	[gray]	[gray]	✶	—
Awareness that different symptoms were related to different diseases	✶	—	✶	—
Stating a testable hypothesis	✶	—	✶	[gray]
Collecting evidence to test hypothesis	✶	—	—	—
Understanding the vector of disease transmission	✶	✶	—	[gray]
Stating a conclusion	✶	[gray]	[gray]	—

Key: ✶ = Treatment that on average had highest scores in this category

    — = Treatments that on average had worse scores in this category relative to ✶treatments

    [gray] = Treatments that on average were not significantly different from the others in this category

Further analysis of students' letters to the mayor of River City suggest that students demonstrate an understanding of the process of inquiry 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. In addition, in their letters both low- and high-performing students demonstrated a clear causal relationship between the problem and the reason(s) for the problem. As another illustration, in their

letters low-performing content students matched the high-performing content students around criterion of stating an opinion regarding the cause of the problem and/or the outcome of the experiment. These differences from the survey results is further indicated by the high success that GSC students had on the letters to the mayor despite the fact that on the analysis of the survey results, GSC students scored similarly to the control students and worse than other River City students.

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. This also brings to question whether inquiry can be assessed with standardized tests, and if not, what effect this will have on its integration into the standards-based classroom.

River City versus the Control Curriculum on Improving Inquiry:

Our last research question wonders whether River City can facilitate teaching with inquiry from the perspective of the teacher? We are concerned that so many science teachers are unfamiliar with methods for teaching with inquiry. We designed River City in the hopes that it could support their learning about scientific inquiry. We are just beginning the analysis of the teachers involved in the project; however, initial analysis seems to indicate it does. For example, one teacher reported that "Students seem to learn more deeply about science and problem solving in the simulation or manipulation setting than in traditional book education, as evidenced by their class discussions" (Galas & Ketelhut, in press). In addition, in focus groups and interview, teachers recognize how this project grabbed the interest of all of their students. Nearly all teachers want to implement the project in following years.

Conclusion

Scientific inquiry is a difficult construct for teachers to implement without support, and the current emphasis on content coverage via high stakes tests often reinforces presentational pedagogies. Our project is showing that MUVE-based curricula can teach standards-based biological content infused with complex inquiry skills better than good traditional approaches do. While analysis of this sizable dataset is still underway, our preliminary findings show that students learned biology content, that students and teachers were highly engaged, that student attendance improved, that disruptive behavior dropped, that students were building 21st century skills in virtual communication and expression, and importantly, that using this type of technology in the classroom can facilitate good inquiry learning.

References

American Association for the Advancement of Science. (1990). *Science for All Americans*. New York: Oxford University Press.

American Association for the Advancement of Science. (1993). *Project 2061: Benchmarks for science literacy*. New York: Oxford University Press.

Clarke, J., & Dede, C. (2005, April). *Making learning meaningful: An Exploratory Study of Using Multi User Virtual Environments (MUVES) in Middle School Science*. Paper presented at the American Educational Research Association Conference, Montreal.

Dillashaw, F. G., & Okey, J. R. (1980). Test of integrated process skills for secondary science students. *Science Education, 64*(5), 601-608.

Falk, J., & Drayton, B. (2004). State Testing and Inquiry-based Science: are they Complementary or Competing Reforms? *Journal of Educational Change, 5*, 345-387.

Fraser, B. (1981). *TOSRA: Test of Science Related Attitudes*. Australian Council for Educational Research, Hawthorne, VIC.

Galas, C & Ketelhut, D. (in press). River City, The MUVE. *Leading and Learning with Technology*.

Hinrichsen, J., & Jarrett, D. (1999). *Science Inqury for the Classroom: a Literature Review*. Portland: Northwest Regional Educational Laboratory.

Ketelhut, D. (2005, April). *Assessing Science Self-Efficacy in a Virtual Environment: a Measurement Pilot*. Paper presented at the National Association of Research in Science Teaching Conference, Dallas.

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. (1996). *National Science Education Standards: observe, interact, change, learn*. Washington, DC: National Academy Press.

National Science Teachers Association. (2004). *NSTA Position Statement: Scientific Inquiry* (Draft), [Internet]. NSTA. Available: http://www.nsta.org/main/forum/showthread.php?t=1175 [2004, August 9].

Nelson, B., Ketelhut, D. J., Clarke, J., Bowman, C., & Dede, C. (2005). Design-Based Research Strategies for Developing a Scientific Inquiry Curriculum in a Multi-User Virtual Environment. *Educational Technology, 45*(1), 21-27.

Wallace, J., & Louden, W. (2002). Introduction to "Laboratories". In J. Wallace & W. Louden (Eds.), *Dilemmas of Science Teaching* (pp. 36-37). New York: RoutledgeFalmer