British Journal of Educational Technology Vol 41 No 1 2010 56–68
doi:10.1111/j.1467-8535.2009.01036.x

A multi-user virtual environment for building and assessing higher order inquiry skills in science

Diane Jass Ketelhut, Brian C. Nelson, Jody Clarke and Chris Dede

Diane Jass Ketelhut is an assistant professor of Science Education at Temple University. She is currently investigating the use of virtual environments for science assessment. Brian C. Nelson is an assistant professor of Educational Technology at Arizona State University. His current research examines learning, assessment, and engagement issues involved in the use of immersive computer games for STEM education. Jody Clarke is a research associate at Harvard Graduate School of Education. Her research focuses on how immersive technologies enable new approaches and methods for assessing and understanding learning. Chris Dede is the Timothy E. Wirth professor in Learning Technologies at Harvard Graduate School of Education; his fields of scholarship include emerging technologies, policy, and leadership. Address for correspondence: Dr Diane Jass Ketelhut, 1301 Cecil B. Moore Ave, Ritter Hall 450, CITE Department, Temple University, Philadelphia, PA 19122, USA. Email: djk@temple.edu

Abstract

This study investigated 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 collaboratively solved problems around disease in a virtual town called River City. The students interacted with 'avatars' of other students, digital artefacts and computer-based 'agents' acting as mentors and colleagues in a virtual community of practice set during the time period when bacteria were just being discovered. This paper describes the results from three implementations of the River City virtual environment in 2004–05 with approximately 2000 students from geographically diverse urban areas. The results indicated that students were able to conduct inquiry in virtual worlds and were motivated by that process. However, the results from the assessments varied depending on the assessment strategy employed.

Introduction

For decades, science educators have worked to infuse inquiry into the K-12 curriculum (American Association for the Advancement of Science (AAAS), 1990, 1993; National Research Council (NRC), 1996). For example, the National Science Teachers Association (NSTA) in the US 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 (NSTA, 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 centred on both inquiry and coverage of state and national content standards would help teachers achieve both objectives.

However, curricula such as these only partially solve the problem. 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 standardised 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 and experimental design? In this paper, we provide an overview of a National Science Foundation-funded curriculum project that focuses on both inquiry- and standards-based content, using novel pedagogies embedded in a virtual environment to help low-performing students master complex inquiry skills. In addition, we discuss the conundrum related to standardised assessment approaches used with virtual environment-based curricula, and we 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, hypothesising 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' (NRC, 1996, p. 23).

However, a problem arises when teachers attempt to infuse coverage of mandatory content with active learning of inquiry because many assume that such activity by students is much more time-consuming (yet more effective) than passive assimilation (Trautmann, MaKinster & Avery, 2004). Additionally, responses to an NSTA position paper indicate that many teachers are unclear as to how to implement inquiry in the science classroom (NSTA, 2004). Some teachers presume that traditional 'cookbook' experiments promote inquiry learning for students (Wallace & Louden, 2002); others understand that inquiry is far more complex than ensuring students can follow prespecified recipes.

River City, a multi-user virtual environment (MUVE)

The River City project is studying how a virtual environment-based 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

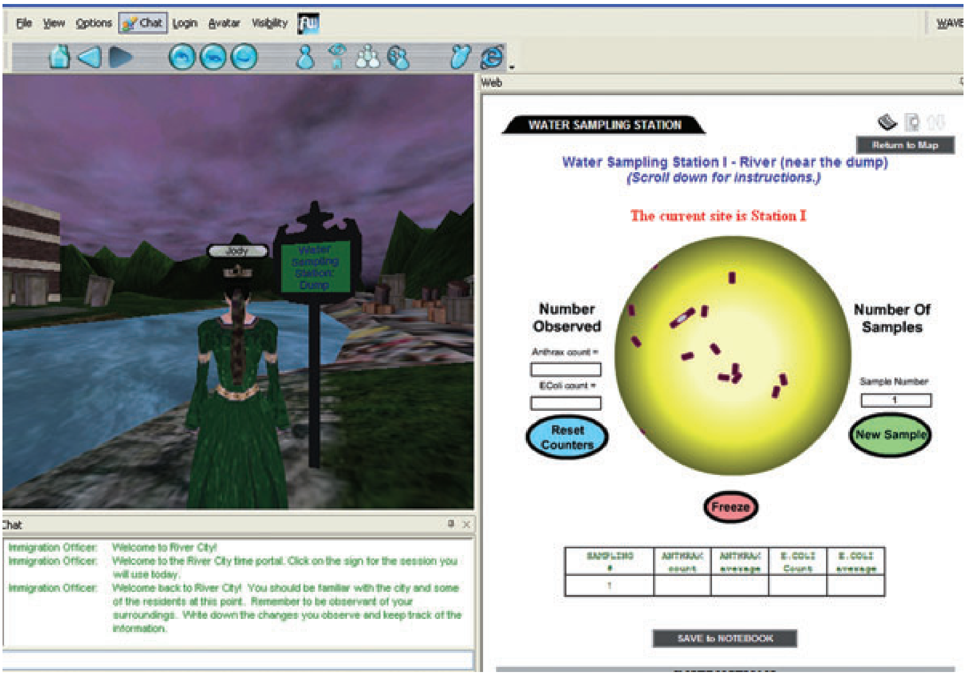

Figure 1: River City interface

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 MUVEs, which resemble the entertainment and communication media students use outside of school, can re-engage them in learning. MUVEs enable multiple simultaneous participants to access virtual contexts, interact with digital artefacts, represent themselves through ‘avatars’, communicate with other participants and with computer-based agents, and enact collaborative learning activities of various types in order to create a community of inquiry learners.

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

River City development

As we describe in detail elsewhere in an extensive history of the project (Ketelhut, Clarke & Nelson, in press), River City began as a 'Multi-user Virtual Environment Experiential Simulator' (MUVEES), a two-year study to create and evaluate graphical virtual environments that use digitised museum resources to enhance middle school students' motivation and learning about science. To accomplish the goals of that study, we built our own MUVE system based on the Sense8 WorldToolKit (no longer available). In 2004, we migrated the River City MUVE to a commercial MUVE engine from ActiveWorlds, Inc. (http://www.activeworlds.com). This update allowed us to more easily embed realistic visual and auditory objects throughout our virtual town and more easily support interactive inquiry tools in the MUVE.

Inquiry and River City

In River City, students engage in all aspects of inquiry as defined by the NSES. These aspects are listed in the next section, and we have mapped each onto how and where in River City the behaviour 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 computerised 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 the River City library, guidance hints, embedded clues in digitised historical images, and the hospital admissions record.
4. 'Using tools to gather, analyse, and interpret data'—students can gather data from two tools: a water sampling tool and a 'bug-catching' tool (see Figure 1). Each tool is activated by a student clicking an icon to draw a sample; the student then counts bacteria in a screen that is similar to a microscope.
5. 'Planning investigations'—students are guided through a generalised process of the scientific method, culminating in creating a controlled experiment to test their hypothesis about the problems in River City.
6. 'Reviewing what is already known in light of experimental evidence'—students gather evidence on the problem from multiple sources, including embedded agents in the form of hospital doctors and university researchers prior to conducting their own experiments.
7. 'Proposing answers, explanations, and predictions'—students create a hypothesis based on collecting evidence to predict what they think is causing part of the problems 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 (Ketelhut, 2007).

Students work in teams to gather data, to develop hypotheses regarding one of three strands of illness in the town (water-borne, airborne and insect-borne) and then to test their hypothesis. The 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 analyse their data using graphs and tables, and then write an authentic lab report on their findings in 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. Previous research on the River City project explored design-based research strategies (Clarke, Dede, Ketelhut & Nelson, 2006; Nelson, Ketelhut, Clarke, Bowman & Dede, 2005), issues of scale (Clarke & Dede, 2009), and the impact of guidance (Nelson, 2007), self-efficacy (Ketelhut, 2007) and learning styles (Dieterle, 2009) on learning and engagement.

In order to explore the type of learning best supported by virtual environments used for inquiry learning, we developed three variations of the River City curriculum for these implementations. Variant guided social constructivist (GSC) centres on a GSC model of learning-by-doing, in which guided inquiry experiences in the MUVE alternate with in-class interpretive sessions led by the teacher. In the GSC curriculum, student teams explore River City, aided by the use of embedded visual and auditory clues that provide situated context for making sense of the issues facing the citizens of the town. Variant expert modelling and coaching (EMC) shifts the learning model to a situated pedagogy with EMC based on expert agents embedded in the MUVE. In the EMC curriculum, student teams can ask questions to 'live' characters in the world, played by trained graduate student assistants on the project. These expert agents were trained to offer support to learners in River City in making sense of the data they gather, without giving direct instruction on specific answers or methods. Finally, variant LPP (legitimate peripheral participation) also uses a situated learning model but based on a community of practice in which newcomers to the community are given tasks designed to expose them to the practices and goals of the community. As players gain experience in the community, they are given increasingly complex tasks (Lave & Wenger, 1991). In the LPP curriculum, student teams are given specific jobs to perform by an in-world scientist–agent on each visit to the city.

These three River City variants were compared with a 'control' condition that utilised a paper-based curriculum in which the same content and skills were taught in equivalent time to comparable students without using computers via a GSC-based pedagogy. The control curriculum, epidemiological investigations (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. This paper focuses on findings related to three aspects of scientific inquiry: engagement, the impact of learning theories on outcomes and types of assessment.

Design and procedure

Research questions

The research questions on which this analysis is centred are the following:

1. Do students engage in inquiry (as defined by the NSES) in River City?
2. When compared with the 'control' version of the River City curriculum, what types of significant gains in affect and learning for both content and inquiry do versions GSC, EMC and LPP produce?
3. How do the results on inquiry learning compare between a standardised type of testing and performance assessments?

Sample

This study examines the results of approximately 2000 students. The students were spread across eight schools, 12 teachers and 61 classrooms in major urban areas in the Northeast and Midwest, and a suburban district in Mid-Atlantic US. The schools in these areas had high proportions of English-as-a-second-language (ESL) and free-and-reduced-lunch pupils.

Procedures

The three computer-based variants (GSC, EMC and LPP) of River City were randomly assigned to the students within each classroom, with teachers instructed to minimise 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 captures individual student activity in the virtual environment with a timestamp, allowing us to analyse students' behaviours throughout the implementation. After designing and conducting their experiments, the 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 problems.

Both qualitative and quantitative data were collected from the students and the teachers over the 3-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 *et al*, 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 & 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.

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

The teachers participated in a professional development programme that focused on content, pedagogy, learning theories and facilitation strategies. The teachers collected demographic data and rated their expectations of the students' successes and motivation with the project. The teachers responded to a pre- and post-questionnaire regarding their methods, comfort with technology and reflections on using the MUVE in their science class.

Data analyses

Quantitative analyses were conducted using multi-level modelling and individual growth modelling. These models allowed us to control for natural variations at the class and teacher levels in class size, culture and pedagogy. Qualitative analysis was a multi-step process. First, a rubric was developed for coding the letters to the mayor of River City. Multiple raters coded a subset of the letters and reached 80% agreement. The codes were then discussed and finalised. Once consensus was met, one researcher coded the rest. An analysis of student interviews happened in two stages. First, open coding techniques (Strauss & Corbin, 1998) were used to code the transcripts of the individual students. Then a second round of coding was performed to look specifically for learning, motivation and collaboration.

Findings

Inquiry engagement

To answer our first research question on whether students were engaged in scientific inquiry, we analysed their data-gathering behaviours as shown in the database (Ketelhut, 2007). First, we were interested in understanding whether students were engaged in the processes of inquiry. Figure 2 shows the average trajectory of the total data-gathering behaviours across the three main data-gathering visits (visits 2–4) to River City of a subsample of our students. As seen in this, the students initially show in visit 2 an average of 12 data-gathering behaviours, which rises to close to 16 by the fourth visit.

We were also interested in whether students used a single source of data on which to base their experiments or whether they used more than one, indicating an informal triangulation in their data gathering. There are a total of eight types of evidential activities in which students could engage: observations, hospital admissions record, talking to residents of River City, River City library books, guidance hints, clues in embedded digital artefacts, and water and bug sampling stations. As seen in Figure 3, the students in this subsample began gathering data from at least two sources on average, increasing to nearly four by the fourth visit. Thus, the technology afforded to us in the MUVE allows us to confirm that the students are engaged in scientific inquiry

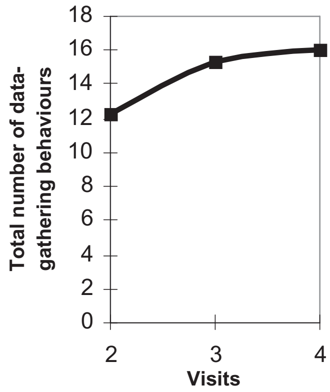

Figure 2: Average individual growth trajectory for students total data-gathering behaviours for visits 2, 3 and 4 (n = 96)

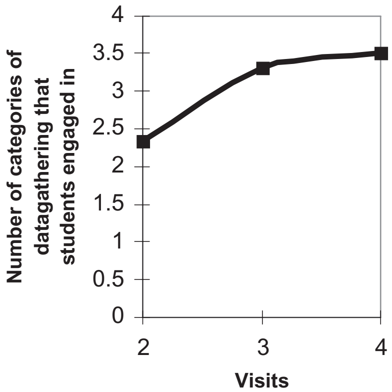

Figure 3: Average individual growth trajectory for diversity of data gathering behaviours, line (n = 96)

behaviours and choose to vary how they gather data across the different sources. In addition, they continued to increase their commitment to the activities of inquiry throughout the data-gathering period.

To confirm the data from the database, we also analysed the results of interviews and focus groups, looking for evidence of inquiry. Many of the students claimed that they felt similar to a scientist for the first time during the River City curriculum because they were 'doing tests and stuff to see what was causing the sickness' (Clarke & Dede, 2005). The virtual microscope and bug catcher tools helped the students feel as if 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 of the students to take on a new identity as an effective science learner. The students using the River City virtual environment enjoyed the inquiry pedagogy and liked that it was 'more independent working ... rather than having him instruct us and telling us what to do and guiding us'. The students claimed, '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 of the 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 the 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 the implementation sites, attendance rates during typical school days were quite low: questions of pedagogy and curriculum are meaningless when many students are rarely in class to experience them. In some of our River City classrooms, we found that student attendance improved and disruptive behaviour dropped during the implementation (Nelson *et al*, 2005).

We were also interested in the characteristics of our virtual environment-based curriculum that promote scientific interest and inquiry. For example, on our affective measure test, we assessed thoughtfulness of inquiry, a measure of students' 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 in order to draw conclusions. A subsample of the students in this study, on average, scored higher on this measure after participating in the River City curriculum ($p < 0.01$) than students in the paper-based control curriculum. For example, the River City students scoring an average of 1 (they strongly disagree that they are metacognitively aware) on the scale of 1–5 for the pretest were associated with scores of 1.8–1.9 on the posttest, nearly double their starting average score. The students in the control group also improved on average, but only to 1.3. Later implementations, however, had more neutral findings, indicating that this is an area in need of more research.

Another subscale measured interest in a scientific career; the gain in interest in science careers was 5% higher for the 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. These results are a bit more equivocal. The students in the River City experimental treatments in one site improved their biological knowledge by 32–35%

($n = 300$). The control students also improved, but by only 17%. However, in the other sites, we saw little differences between the treatments and little growth over the course of the project.

Inquiry content results

The second and third research questions for these implementations revolve around whether using a virtual environment-based inquiry project could improve inquiry learning for students as compared with a control curriculum, what the impact of three different pedagogical learning theories were on inquiry learning and whether varying the method of measurement gave different answers to those questions. When using survey questions to assess inquiry, we found few differences between the three treatments, or compared with the control. In one site, improvements were seen across the board for knowledge and application of scientific processes; 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. The results for the other sites showed an additional difference by gender.

Looking at their inquiry scores on the standardised post-survey that was similar to a test, we found that overall, the students in the LPP and EMC River City treatments outperformed the students in the control treatments (although the GSC curriculum did not differ from the control at all). However, in addition to this primary result, there were two other interesting effects seen. First, for all the treatments except for LPP, the boys outperformed the girls on the inquiry survey questions. Interestingly, the LPP treatment that was based on a collaborative community of practice model better supported the girls, as indicated by their higher scores for that treatment only. The second interesting finding is that for all the treatments except EMC, we saw that the students who had above average science grades did better than those with below average previous science grades. The students in the EMC treatment with below average previous science grades did nearly equally as well on the posttest survey as the above average students.

Because we wondered how difficult it was to measure inquiry with a multiple-choice test, we also analysed the students' end-of-implementation 'Letter to the Mayor' for evidence of inquiry elements. In our first implementation, the instructions given for writing the letters varied somewhat between the River City curriculum and the control curriculum; as a result, a 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 were typically much shorter in length, did not demonstrate motivation or engagement, did not mention the experiment and did not explicitly recognise the interconnectedness of the chosen problem with other possible causes of the larger problem. An analysis of the letters for evidence of inquiry found that on average ($p < 0.01$), the students taking part in the MUVE-based curriculum earned scores more than double that of their paper-based control peers.

For the next implementations, the instructions for completing the letters were identical in all of the treatments, which allowed for a more detailed comparison between the letters. The letters were coded to show the various aspects of inquiry (problem statement, testable hypothesis, evidence collected to allow evaluation of hypothesis, conclusion offered), overall quality and indications of understanding of the biology involved (symptoms and disease connection, understanding the vectors of disease transmission). In contrast to the findings from the post-survey, the students in the GSC treatment had the highest scores in nearly every category except problem statement, whereas the 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 ($p < 0.05$). The EMC and LPP students did better than the control curriculum on a few categories, with the LPP students performing the best of all of the treatments on the problem statement category (Ketelhut, Dede, Clarke & Nelson, 2007).

A close analysis of the students' letters to the mayor of River City suggests that this form of assessment allows for students to demonstrate knowledge of inquiry processes in ways that differ from or are not possible on the multiple-choice test. For example, the letters written by the students who scored low on the science inquiry posttest versus those written by the students who scored higher on the posttest were indistinguishable. In addition, in their letters, both the low- and high-performing students demonstrated a clear causal relationship between the problem and the reason(s) for the problem and scored similarly on the ability to state an opinion regarding the cause of the problem and/or the outcome of the experiment. These differences between letter writing and multiple-choice measures is further indicated by the high success that the GSC students had on the Letters to the Mayor—despite the fact that on the analysis of the test results, the GSC students scored similarly to the control students and worse than the other River City students.

Interestingly, the students who received a lower score on the multiple-choice test were more likely to suggest future interventions or further research than the students who scored higher on the inquiry test questions. Perhaps the complexity of the virtual environment-based River City curriculum is more appropriately measured with authentic activities, such as writing an experimental report. This also brings to question whether scientific inquiry can be assessed with standardised tests, and, if not, what effect this will have on its integration into the standards-based classroom.

Conclusion

We set out in this study to investigate the best methods for designing a virtual environment for scientific inquiry and for assessing the understanding of inquiry. Our results show an intricate pattern of understanding that is complicated by the method of assessment. Using a standardised post-survey that was similar to a test, we found few differences between the River City group and the control group, except that the students with poor grades in science do best when taught scientific inquiry with an expert mentoring and modelling version of River City. We also found that girls tend to do worse than boys, except those in the community of practice River City variant. However, when we look at a performance assessment that mimics a lab report, we see a very different story: students in the guided social constructivist version of River City show a stronger understanding of scientific inquiry than all the other students.

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 River City project is showing that virtual environment-based curricula can teach standards-based biological content infused with complex inquiry skills as well or better than good traditional approaches do. Our findings also suggest that designers should consider their purposes for designing a virtual environment and choose a pedagogical learning theory that matches that purpose. Finally, in a culture that values high-stakes assessments, we argue that our findings indicate that high-stakes tests used in isolation of other information can erroneously categorise students' understanding.

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). *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.

Clarke, J. & Dede, C. (2009). Design for scalability: a case study of the River City curriculum. *Journal of Science Education and Technology*, *18*, 4, 353–365.

Clarke, J., Dede, C., Ketelhut, D. J. & Nelson, B. (2006). A design-based research strategy to promote scalability for educational innovations. *Educational Technology*, *46*, 3, 27–36.

Dieterle, E. (2009). Neomillennial learning styles and River City. *Children, Youth and Environments*, *19*, 1, 245–278.

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*. Hawthorne, VIC: Australian Council for Educational Research.

Ketelhut, D. (2005). *Assessing science self-efficacy in a virtual environment: a measurement pilot*. Paper presented at the National Association of Research in Science Teaching Conference, Dallas.

Ketelhut, D., Clarke, J. & Nelson, B. (in press). The development of River City, a multi-user virtual environment-based scientific inquiry curriculum: historical and design evolutions. In M. Jacobson (Ed), *Designs for Learning Environments of the Future: International perspectives from the learning sciences*. New York: Springer Publishing Company.

Ketelhut, D. J. (2007). The impact of student self-efficacy on scientific inquiry skills: an exploratory investigation in River City, a multi-user virtual environment. *Journal of Science Education and Technology*, *16*, 1, 99–111.

Ketelhut, D. J., Dede, C., Clarke, J. & Nelson, B. (2007). Studying situated learning in a multi-user virtual environment. In E. Baker, J. Dickieson, W. Wulfeck & H. O'Neil (Eds), *Assessment of problem solving using simulations* (pp. 37–58). Mahwah, NJ: Lawrence Erlbaum Associates.

Lave, J. & Wenger, E. (1991). *Situated learning: legitimate peripheral participation*. Cambridge: Cambridge University Press.

Midgley, C., Maehr, M. L., Hruda, L. Z., Anderman, E., Anderman, L., Freeman, K. E. *et al* (2000). *Manual for the Patterns of Adaptive Learning Scales (PALS)*. Ann Arbor, MI: University of Michigan.

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

National Science Teachers Association (NSTA) (2004). *NSTA position statement: scientific inquiry* (Draft), [Internet]. NSTA. Retrieved October 23, 2009, from http://www.nsta.org/about/positions/inquiry.aspx.

Nelson, B. (2007). Exploring the use of individualized, reflective guidance in an educational multi user virtual environment. *Journal of Science Education and Technology*, *16*, 1, 83–97.

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.

Strauss, A. & Corbin, J. (1998). *Basics of qualitative research: techniques and procedures for developing grounded theory*. Thousand Oaks, CA: Sage.

Trautmann, N., MaKinster, J. & Avery, L. (2004). *What makes inquiry so hard? (and why is it worth it?)*. Paper presented at the annual meeting of the National Association for Research in Science Teaching, Vancouver, BC, Canada.

Wallace, J. & Louden, W. (2002). Introduction to 'laboratories'. In J. Wallace & W. Louden (Eds), *Dilemmas of science teaching* (pp. 36–37). New York: RoutledgeFalmer.