

A Design-Based Research Strategy to Promote Scalability for Educational Innovations

Jody Clarke
Chris Dede
Diane Jass Ketelhut

Harvard Graduate School of Education

Brian Nelson
Arizona State University

Ways for Information Technology to Aid in Bringing Products and Services to Scale

"Scaling up" involves adapting an innovation successful in some local setting to effective usage in a wide range of contexts. In contrast to experiences in other sectors of society, scaling up successful programs

Jody Clarke is a doctoral candidate in Learning and Teaching at the Harvard Graduate School of Education. Her research interests are in the Learning Sciences and emerging technologies. Jody is project manager for Chris Dede's NSF funded grant focused on scaling up a multi-user virtual environment (MUVE) designed around scientific inquiry (e-mail: clarkejo@gse.harvard.edu).

Chris Dede is the Timothy E. Wirth Professor of Learning Technologies at Harvard's Graduate School of Education. His fields of scholarship include emerging technologies, policy, and leadership. His funded research includes a grant from the National Science Foundation to aid middle school students learning science via multi-user virtual environments and a Star Schools grant from the U.S. Department of Education to help high school students with math and literacy skills using wireless mobile devices to create augmented reality simulations. Chris is the co-editor of *Scaling Up Success: Lessons Learned from Technology-based Educational Innovation*, published by Jossey-Bass in 2005. He is a Contributing Editor of *Educational Technology* (e-mail: Chris_Dede@harvard.edu).

has proved very difficult in education (Dede, Honan, & Peters, 2005). For example, the automation or one-size-fits-all model does not fit when scaling up in education because a pedagogical strategy that is successful in one particular classroom setting with one particular group of students frequently will not succeed in a different classroom with other students. Scaling educational innovations without realizing that their effectiveness is often eroded by variations in implementation context may result in a "replica trap": repeating everywhere what worked locally without taking into account individual variations in needs and assets (Wiske & Perkins, 2005).

Outside the field of education, information technology has aided in bringing product or service innovations to scale primarily in two complementary ways: automation and individualization. Automation simplifies and standardizes a product or service so that necessary tasks to supply it require only pre-set routine actions by people or machines. Through automation, for example, a factory can use information technology coupled with machines to generate mass-produced products (e.g., identically configured clocks) cheaply, efficiently, and reliably. Similarly, through automated processes and associated standardized protocols for employee actions, a whole chain of restaurants can simultaneously and successfully implement a new process for frying food. Typically, applied automation achieves scale via the lowest common denominator, the one-size-fits-all design and implementation strategies.

In contrast, individualization produces variants of products tailored to a wide spectrum of styles and

Diane Jass Ketelhut is a lecturer and doctoral candidate in Learning and Teaching at the Harvard Graduate School of Education. Her research interests are in scientific inquiry, academic self-efficacy and how technology can improve science education. Diane is the Chair of the Graduate Student Council of AERA. She holds certification in secondary school science and was a science curriculum specialist and teacher for grades 5-12 for 12 years, prior to coming to Harvard. She received a BS in Bio-Medical Sciences from Brown University and an MEd in Curriculum and Instruction from the University of Virginia (e-mail: diane_ketelhut@gse.harvard.edu).

Brian Nelson is an Assistant Professor of Educational Technology in the Division of Psychology in Education at Arizona State University. His current research centers on the theory, design, and implementation of educational multi-user virtual environments for science inquiry, and MUVE-based adaptive guidance systems. He is also working with the Project Pathways group at ASU on research into the design of learning supports for English language learning (ELL) students studying math and science (e-mail: Brian_Nelson@asu.edu).

tastes. For example, many software applications allow users to customize their appearance, toolbars, features, and modes of processing. Information technology enables designers to embed ways that users can co-create the specific product or process they are seeking. Individualization achieves scale by meeting a spectrum of customer needs with a customizable product or service—often at some cost in terms of price, complexity of co-design, and challenges in usage compared to alternative products or services mass-produced through automation.

Recent advances in technology are creating an emerging “fusion” option for scale, a design and implementation strategy that combines the virtues of automation and individualization. As an illustration, consider the iPod, Apple Computer’s portable digital music player (<http://www.apple.com/ipod/color/>). The iPod is mass produced and identical when removed from its packing by each consumer. How the purchaser configures this device for usage, however, is very individualized. The iPod is generally used to organize and listen to music, audio books, and podcasts. Yet, this same device can function as an audio recorder, a fast external hard drive, a photo library display device, and a personal assistant to view and synchronize data such as a calendar and contacts. A variety of accessories allow even more potential uses, as well as customizing the look and feel of the device to the owner’s personality.

iPods are an example of a technology that is cost-competitive and produced through automation, but easily adaptable to a wide range of uses and styles based on individual needs and tastes. While the technical capabilities of iPods are uniform, users individualize their experiences, taking advantage of some functions while ignoring others. Information technologies enable the sophisticated design and production that makes this fusion of automation and individualization possible.

In this article, we describe how Design-Based Researchers in the Learning Sciences can build on a fusion of automating and individualizing in order to take an educational innovation to scale. As a case study, we discuss our ongoing Design-Based Research¹ in scaling up a multi-user virtual environment curriculum to enable effective, standard usage across a spectrum of educational contexts, while supporting individualization where necessary to adapt to local classroom, school, and district conditions. In the following sections, we first briefly describe the River City curriculum and present findings from implementations to date. Then we delineate and illustrate our

strategy for scaling up based on fusing automation and individualization. We conclude with a description of a proposed “scalability index” that would allow measuring the scalability of an educational innovation across a wide spectrum of contexts.

Design of and Findings from the River City Curriculum

River City is a multi-user virtual environment (MUVE) designed to teach middle school science. MUVEs enable multiple simultaneous participants to access virtual contexts, interact with digital artifacts (such as online microscopes and pictures), represent themselves through “avatars,” communicate with other participants and with computer-based agents, and enact collaborative learning activities of various types (Nelson, Ketelhut, Clarke, Bowman, & Dede, 2005). The River City curriculum 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.

The virtual “world” is a city, set in the late 1800s, and concentrated around a river that runs from the mountains downstream to a dump and a bog. Like most 19th century industrial towns, it contains various neighborhoods, industries, and institutions, such as a hospital and a university (see Figure 1).

Upon entering the city, the students’ avatars can interact with computer-based agents (residents of the city), digital objects (pictures and video clips), and the avatars of other students. In exploring, students also

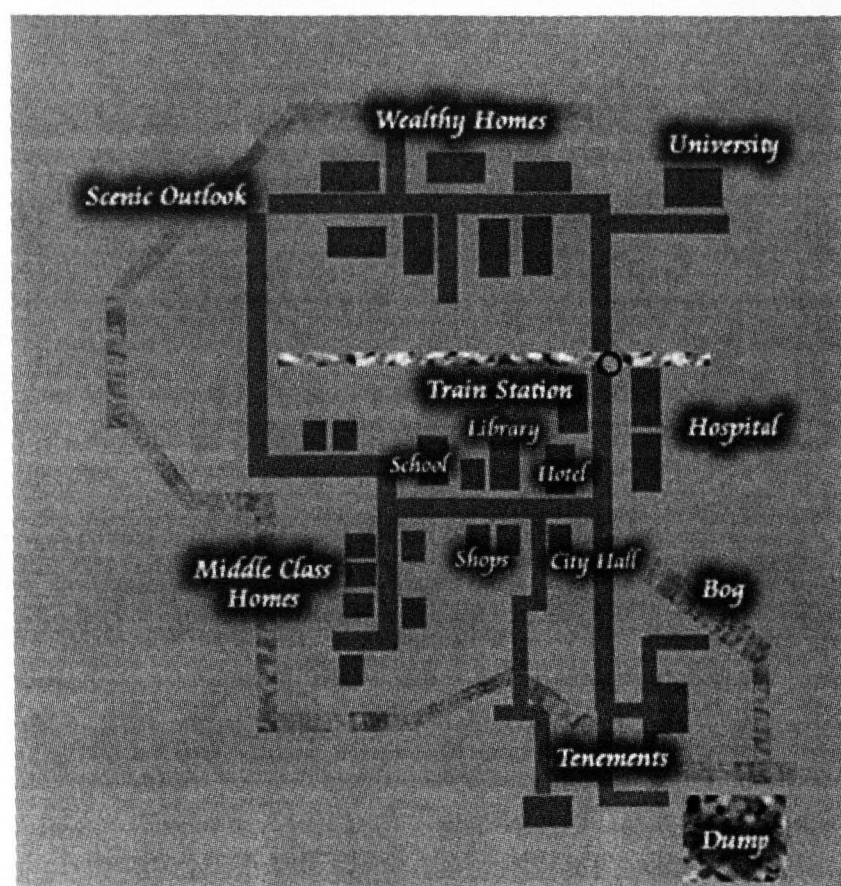
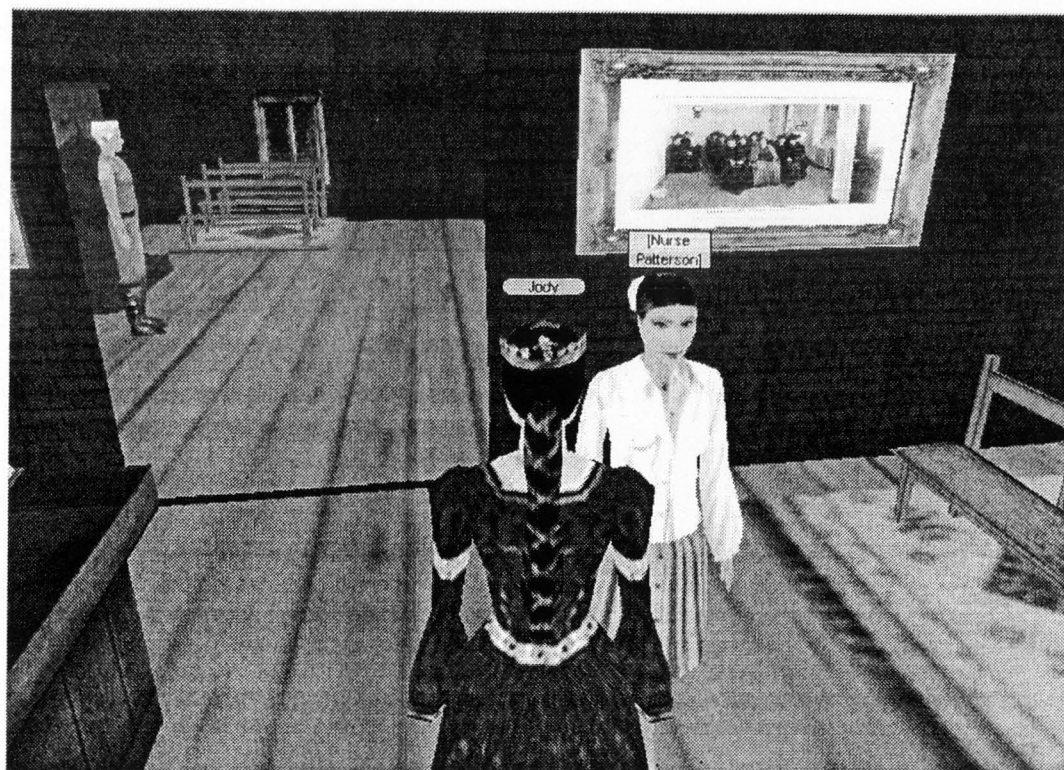


Figure 1. Map of River City.

¹If you want to know more about how our work in Design-Based Research, please see the January, 2005, special issue of *Educational Technology on Design-Based Research*.



Immigration Officer: Welcome to River City!
 Immigration Officer: Welcome to the River City time portal. Click on the sign for the session you will use today.
 Immigration Officer: Welcome to River City! You will be attending the University here, and your role is two-fold: to learn how to become a great scientific investigator and to help us uncover the mystery of why so many River City residents have been getting sick.
 Jody: (to [Nurse Patterson]) What's New?
 [Nurse Patterson]: Lately, it is exhausting working here as it has been more crowded. Also, the daughter of my neighbors on Second Street has recently become sick to her stomach.
 Jody: (to [Nurse Patterson]) Thanks.
 [Nurse Patterson]: You are welcome.

River Ci

Waiting room

The waiting room is very crowded today. I wonder where everyone is from and what their symptoms are.



Photograph courtesy of the Smithsonian Institute, National Museum of American History



Environment Health Meter:
 This meter shows the health of the environment in River City.



Whisper To: [Aaron Nelson]

Figure 2. Talking with an agent.

encounter various visual and auditory stimuli, such as mosquitoes buzzing and people coughing that provide tacit clues as to possible causes of illness. Content in the right-hand interface-window shifts based on what the student encounters or activates in the virtual environment, such as a dialogue with an agent (Figure 2) or a virtual microscope that allows examination of water samples (Figure 3).

Students work in teams of three or four to develop and test hypotheses about why residents are ill. Three different illnesses (water-borne, air-borne, and insect-borne) are integrated with historical, social, and geographical content, allowing students to develop and practice the inquiry skills involved in disentangling multi-causal problems embedded within a complex environment (Ketelhut, Clarke, Dede, Nelson, & Bowman, 2005). A final sharing day at the end of the project allows students to compare their research with that of other teams of students in their class and to piece together some of the many potential hypotheses and causal relationships embedded in the virtual environment.

Findings from Our River City Implementations to Date

Utilizing Design-Based Research strategies, we have conducted numerous studies of the River City MUVE to determine if virtual environments can simulate real-world experimentation and provide students with engaging, meaningful learning experiences that enhance scientific literacy. For reasons of space, we discuss below only some research findings from our 2004 implementations of this curriculum. We present these results to illustrate how we obtain insights about what contextual factors pose barriers to scale and what design elements we could modify to overcome those barriers.

Spring, 2004: Second Generation of the River City MUVE. Based on prior small-scale studies, in 2003 we developed a "second generation" version of the River City MUVE and designed two pedagogical variants based, respectively, on guided social constructivism (GSC) and embedded modeling and coaching (EMC). In the spring of 2004, we conducted a large-scale study of these River City MUVE variants with 11 teachers and

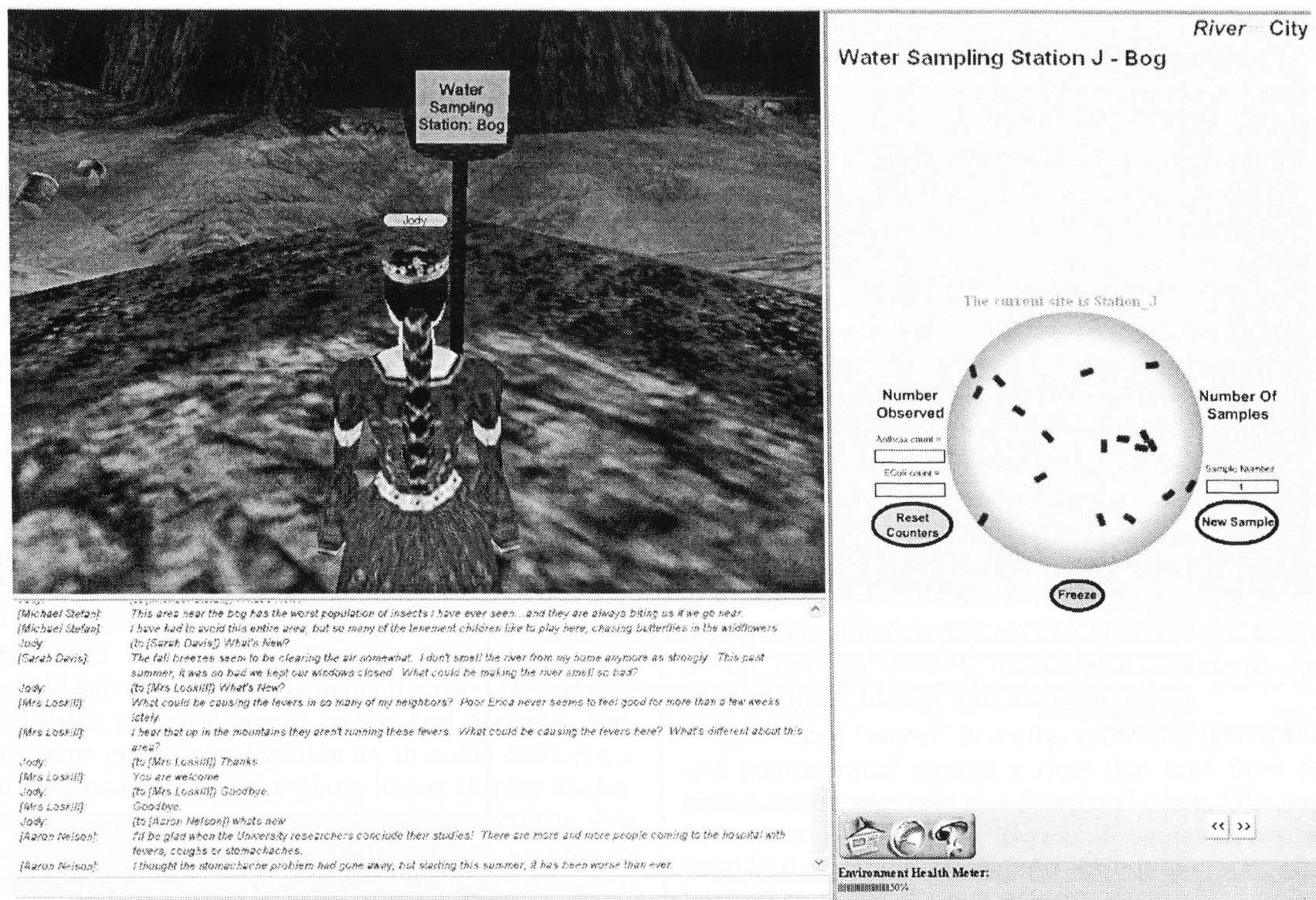


Figure 3. Virtual microscope.

over 1000 students in urban public middle schools in Wisconsin and Massachusetts with high proportions of ESL and low SES students. A control curriculum with similar content and pedagogy, but delivered completely on paper, was also developed and randomly assigned to whole classes, with all but one teacher offering both the computer-based treatments and control. Through this experimental approach, we could study the value of the MUVE as a medium for learning (Dede et al., 2004).

Within the experimental classrooms, students were randomly assigned to one of the two variants of the River City treatment (GSC or EMC). We collected quantitative and qualitative data similar to the pilot studies in the form of student pre- and post-surveys (an affective measure that assessed student motivation, self-efficacy, and interest in science careers; and a content measure that assessed content knowledge in science inquiry and disease transmission), teacher pre- and post-surveys, teacher expectation of student performance, log files of student activities in the MUVE, and pre- and post-interviews and focus groups with students.

Our results from this implementation supported earlier findings that students engaged in scientific inquiry and built higher-order skills in virtual communication and expression (Nelson et al., 2005). Both students and teachers were highly engaged; girls and boys showed similar patterns of improvement; student attendance improved; and disruptive behavior dropped. 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' (Clarke & Dede, 2005) and asked when River City would be available for purchase. One of our school districts had absentee rates approaching 50% during the time frame of our implementation. For the single participating teacher in that district, absentee rates decreased by 35% from the first to last week of the project.

Results of a randomly chosen representative subgroup of students from four of the 11 teachers were analyzed with multi-level modeling using students' class assignment as the grouping variable. The examination of the results indicates that, on average,

students in a guided social constructivist experimental group (GSC) achieved 16% higher scores on the posttest in biology than students in the control group. Similar results were seen from the affective measures. Student scores for thoughtfulness of inquiry on the post-survey were significantly higher ($p < .01$) on average for both experimental groups (GSC and EMC), in comparison to the scores for students in the control group (Ketelhut *et al.*, 2005).

In addition to the pre- and post-implementation surveys, students also wrote letters to the fictional mayor of River City outlining their hypotheses and experiment results, and offering suggestions for how to deal with the illness in the virtual town. Analysis of the letters' quality on a 26-point scale found that students in the experimental groups (GSC and EMC), on average, earned scores more than double that of the students in the control group (Ketelhut *et al.*, 2005).

Fall, 2004: More learning variants of the River City MUVE. We modified the River City curriculum again, adding an additional learning variant as well as an individual guidance system. The third learning variant, legitimate peripheral participation (LPP), is based on Lave and Wenger's (1991) concept of learning as enculturation into a community of practice. The individualized, computer-based guidance system (IGS) with high and low variants utilizes personalized interaction histories and log files collected on each student's activities to offer real-time, customized support in the form of hints (Nelson, 2005).

In the fall of 2004, we conducted another implementation of River City with five teachers and approximately 490 seventh-grade students in New York State. Once again, the control curriculum was randomly assigned to whole classes, with each teacher offering both the computer-based treatments and control. Within the experimental classrooms, students were randomly assigned to one of two variants of the River City treatments (GSC, LPP) or the IGS system in either High (extensive) or Low variants. As in the spring implementation, we collected quantitative and qualitative data in the form of student pre- and post-surveys (a revised affective measure that assessed student motivation, self-efficacy, and interest in science careers; and a revised content measure that assessed content knowledge in science inquiry and disease transmission), teacher pre- and post-surveys, teacher expectation of student performance, log files of student activities in the MUVE, and randomly selected student pre- and post-interviews. Quantitative data were analyzed with multi-level modeling using students' class assignment as the grouping variable.

We found that students who increased their self-efficacy as science learners also earned higher-score gains on the science content test, on average. The magnitude of this relationship was greatest for low SES

students in the River City group that began with low self-efficacy (Figure 4).



Figure 4. The effect of gain in self-efficacy over the course of the study on gain in total content, controlling for SES and treatment for boys—results for girls are parallel but slightly higher ($n = 424$).

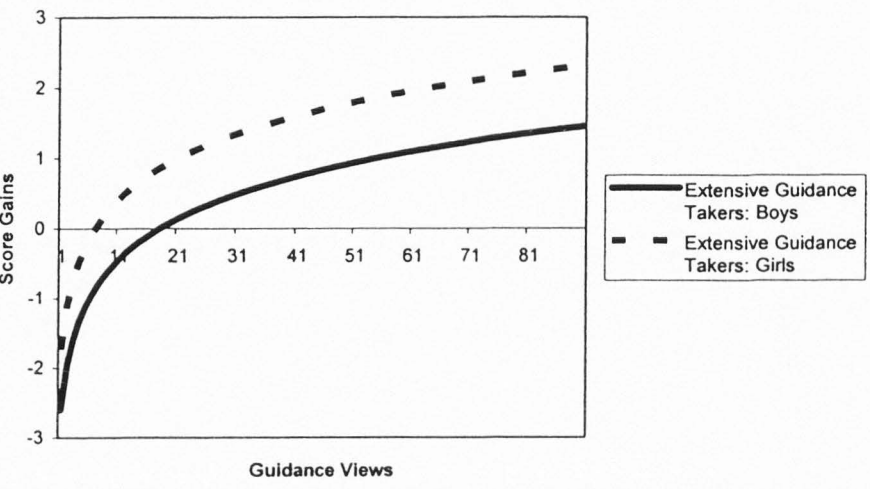


Figure 5. The fitted relationship between levels of guidance system use and content test score gains by students exposed to extensive levels of guidance who chose to “take up” the guidance at least one time in a MUVE-based curriculum, by gender ($n = 272$).

We also found a strong positive link with learning outcomes for students who accessed the guidance system (Nelson, 2005). Students in the high guidance group who accessed more guidance messages earned higher-score gains on the science content test, on average, than those who viewed fewer hints. In addition, we found an interaction between gender and guidance use. Girls using the guidance system outperformed boys, on average, at each level of guidance-message viewing (Figure 5).

We are currently analyzing data gathered from a large-scale implementations conducted in late fall, 2004 with seven teachers and more than 600 students

in North Carolina, Wisconsin, and Massachusetts, as well as from implementations in late spring, 2005.

In addition to results on student learning, our Design-Based Research strategies are helping us to discover which factors we need to modify in order to improve scaling up of our design. As a result of naturalistic variation among teachers, students, technology infrastructures, and school settings, we have an opportunity to assess how factors related to each of these variables affect learning outcomes. For example, we can compare student learning outcomes from teachers with strong backgrounds in science to those with weak backgrounds in science. Based on our assessments to date of these factors, we are developing strategies for which parts of the River City intervention to standardize (automation) and which to make customizable (individualization), as a means to successfully scale our innovation across a wide spectrum of educational settings. We discuss some of these strategies below.

Designing the River City Curriculum for Scalability

If one is not engaged in full-fledged systemic reform of a school system, scaling up requires designing educational innovations to function effectively across a range of relatively inhospitable settings (Dede, 2004). Scalability into typical school sites that are not partners in innovation requires designing interventions that are automated, but allow for individualization in order to retain some efficacy in contexts in which major conditions for success are absent or attenuated. In making judgments about scalability of an intervention, differentiating the intervention's design from its "conditions for success" is important (Dede, in press). For instance, the effective use of antibiotics illustrates the concept of "conditions for success": Antibiotics are a powerful "design," but worshiping the vial that holds them or rubbing the ground-up pills all over one's body or taking all the pills at once are ineffective strategies for usage—only administering pills at specified intervals works as an implementation strategy. A huge challenge that educators face, and one of the reasons this field makes slower progress than venues like medicine, is the complexity of conditions for success required in effective interventions; nothing powerful in facilitating learning is as simple as an inoculation in medicine.

Under these circumstances, major intended aspects of an innovation's design may not be enacted as intended by its developers. Developing a design for scalability into contexts in which "important, but not essential" conditions for success are weakened or lacking requires adding options that individualize the innovation when parts of its intended enactment are missing. Such design strategies are exemplified in our River City MUVE research, which, through Design-Based Research, is identifying conditions for success

likely to be attenuated in many contexts and evolving the curriculum's design to allow for individualization that enhances effectiveness under those circumstances.

In particular, in our research to date we have identified four factors important in the enactment of our River City MUVE curriculum:

- **teacher preparation** (including teacher's knowledge of science and content-specific pedagogy, as well as fluency with educational technology);
- **class size** (affecting the degree of individualization and interaction possible);
- **learner academic achievement** (including factors such as students' perceived self-efficacy in learning science and foundational knowledge in science, technology, and literacy); and
- **learner engagement** (illustrated by indices such as student attendance at school and teachers' perceptions of student motivation and classroom behavior).

Findings from our prior studies provide insights into how to design for scalability when the implementation context is weak in terms of one or more of these conditions.

For example, in our previous implementations of the River City MUVE, we delivered professional development in several ways. Initially, the professional development material was delivered online via a Web portal to allow teachers to access it on their own schedule. Many of our teachers were implementing the curriculum at remote sites, and, due to teachers' busy schedules, coordinating times to meet synchronously in the River City world was generally not feasible. Unfortunately, some of the teachers ignored all or most of the professional development. As a result, problems arose during the implementation because these same teachers did not understand the purpose and process of the curricular intervention, the inquiry skills and content, or the necessary pedagogical strategies for leading the small group and whole class interpretive discussions. While this sounds grim, in practice our curricular intervention worked fairly well even in these situations, as can be seen from the results listed above for the Spring 2004 implementation. Thus, we concluded that the River City MUVE is designed for scalability, creating curricular interventions so compelling for students and with sufficient internal guidance that they have a fulfilling, self-directed learning experience—albeit with reduced educational outcomes—even with an unprepared teacher.

In response to varied participation in the online portal, and the teacher-preparation condition for success, we evolved the professional development portion of the intervention to increase its scalability. For example, we produced a just-in-time, "light" version of the professional development that teachers can skim for ten minutes per day during the unit, providing essential

information needed to guide students for that stage of the learning experience. We also piloted a train-the-trainer approach where we trained a science coordinator who then went back to the district and trained five teachers. We plan to scale this further by training a number of technology and science coordinators and creating a community of practice around the trainers, who will return to their prospective districts and train teachers to use the curriculum.

Illustrations of Fusing Automation and Individualization to Enhance Scalability

In addition to fusing automated and individualized features of the River City MUVE to address scalability issues related to teacher preparation, we are also leveraging Design-Based Research strategies to respond to issues for success of large class size, learner academic achievement, and learner engagement discussed below.

Issue of Success of Class Size and Low Academic Achievement

Research has shown that low-achieving students and students from low socioeconomic backgrounds perform better academically when in smaller-sized classes (Akerhielm, 1995; Boozer & Rouse, 2001; Rice, 1999). Reducing class size requires that schools have available classroom space, access to qualified teachers, and money to pay for increased salaries and resources; the state of California has spent billions of dollars in efforts to reduce class size (Sack, 2002). Our design takes into consideration the fact that reducing class size is a complex issue in education that not every school is able to address successfully; thus, designs that retain effectiveness with large class sizes and with students with poor prior academic performance are important for scalability.

One problem inherent in large class size is that it is difficult for teachers to individualize instruction. To address issues of large class size and histories of low student academic performance, Nelson (2005) created an individualized guidance system (IGS) embedded in the River City MUVE environment to assist students in making sense of the complexity of the virtual worlds and to scaffold each student's explorations. Research has shown that lack of guidance in computer-based constructivist environments, including MUVES, can lead to student confusion (Baylor, 1999; Baylor, 2002; Moreno & Mayer, 2005; Moreno, Mayer, & Lester, 2000). In school settings where students are unaccustomed to exploratory learning and student-centered curricula, or where large class sizes make individualized instruction difficult, absence of embedded guidance in computer-based learning environments can pose powerful barriers to success (Brush & Saye, 2000; Moreno & Mayer, 2005).

The IGS system tracks students' movements and

actions in River City and stores them in a database that maintains personalized histories of each student. All the items students can interact with in River City are programmatically tagged with identification codes. Every time a student clicks on an object or 'speaks' to a River City citizen, a record of the event is stored. A guidance model, operated by an invisible software agent, is triggered after each student interaction event in the River City MUVE. A subset of events is associated with guidance scripts, and the guidance model uses these scripts to offer a specific selection of messages to each student. The scripts contain a set of rules for selecting guidance, based on a student's history of interactions with objects and citizens. The IGS does not automatically show specific guidance content, but instead displays 'hint' buttons linked to guidance messages in the right-hand interface-window (see Figures 6 and 7).

To view guidance messages, students need to click on these hint buttons. In this way, we are able to monitor IGS usage levels and patterns.

While research shows that students who use the IGS system perform better on the post-assessments than students who do not use it (Nelson, 2005), we recognize that not all students will opt to click on the 'hints.' Thus, we are in the process of planning a design that includes intelligent agents that can provide more active and individualized scaffolding for students who are not 'on task.' These automated agents will compensate for students who are at a disadvantage due to large class size or prior academic achievement by focusing on the students' individual needs. Intelligent agents, also known as animated pedagogical agents, are "lifelike characters that cohabit learning environments with students to create rich, face-to-face interactions" (Johnson, Rickel, & Lester, 2000). Research on using life-like pedagogical agents in learning environments suggests that they have positive effects on student learning, motivation (Baylor, 2002; Lester, Towns, & FitzGerald, 1999), and transfer (Moreno & Mayer, 2004; Moreno, Mayer, Lester, 2000). Including intelligent agents in River City will allow us to monitor and diagnose student performance (Baylor, 1999) and provide individualized feedback to students who need more scaffolding and guidance through an automated system.

Issue of Success of Learner Engagement

In our prior studies, we have found that autonomy and optimal level of challenge are critical elements in students' motivation for learning (de Charms, 1968; Dede, Clarke, Ketelhut, Nelson, & Bowman, 2005; Lepper & Henderlong, 2000; Malone & Lepper, 1987; Ryan & Deci, 2000). Our Design-Based Research strategies to enhance student engagement are summarized in Dede, Nelson, Ketelhut, Clarke, and Bowman (2004). While these strategies have enhanced

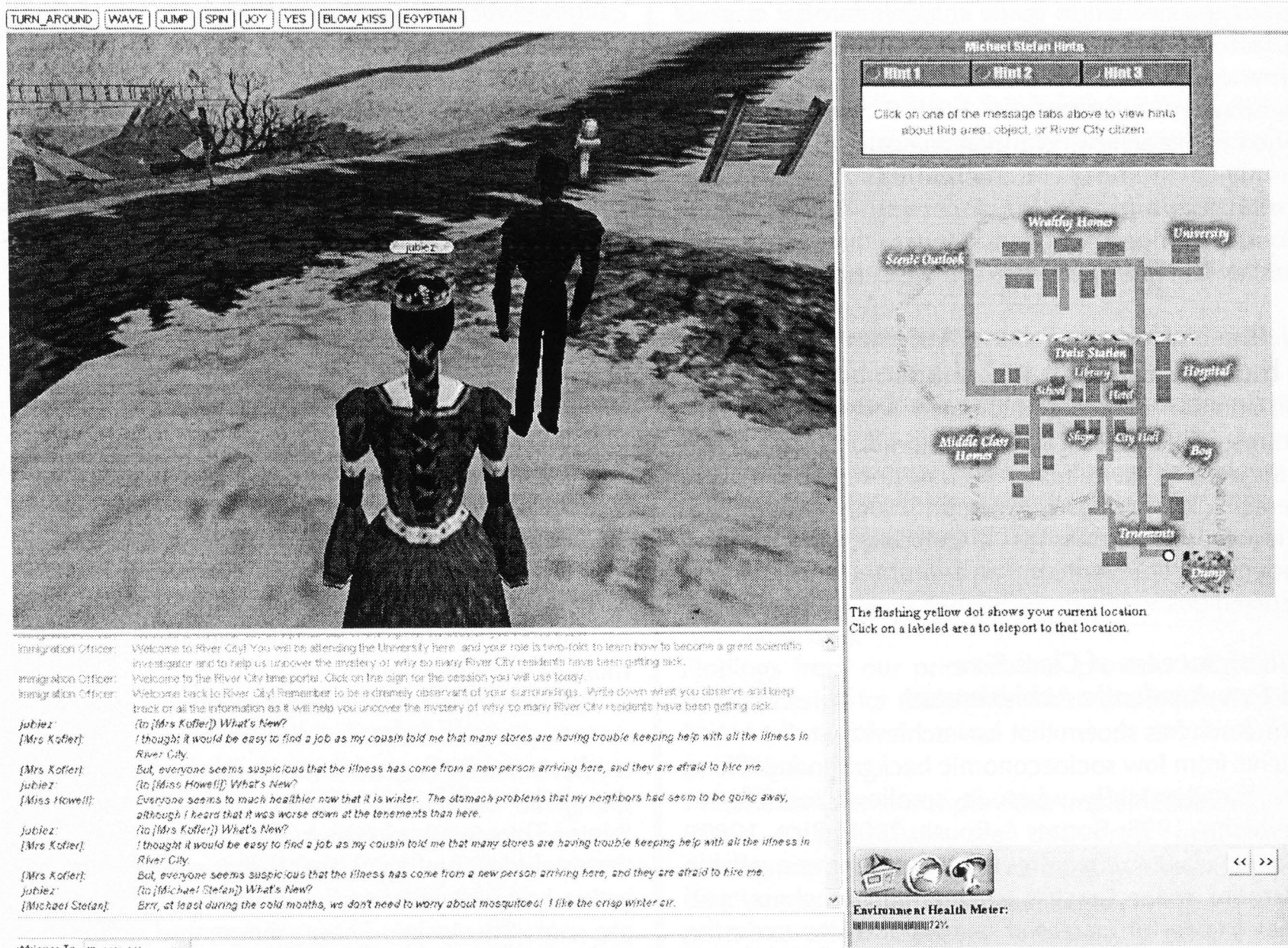


Figure 6. Location of 'Hint' button.

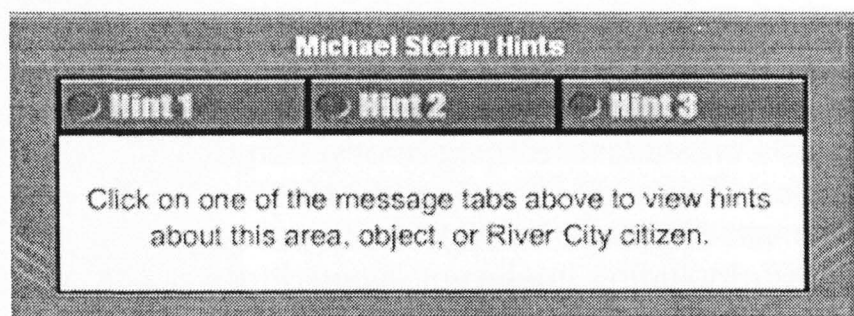


Figure 7. 'Hint' button.

student learning and engagement, we have found that some contexts require design features that engage students and motivate them to complete the daily objectives and engage in inquiry.

In order to do this we are currently designing a feature that will allow a student's avatar to gain "levels" automated to advance students through River City in pre-set stages. Like most video games, each level has associated powers and capabilities; students must master the skills of one level to advance to the next

level. An automated feature will enforce this flow of individual passage from one level to the next when students demonstrate mastery of content. In our present design, students enter River City through a time portal on six separate occasions. These six different "worlds" will be adapted into levels. In order to advance a level, a student will have to complete certain curricular objectives, such as talking to certain residents, visiting specified places in River City, or even helping another team member who is struggling.

These "levels" will reward achievement of various curricular objectives with enhanced "powers" in the MUVE, each linked to academic content. In videogames, the attainment of higher levels with greater capabilities is a major force in participant engagement. One example of the new type of power students will attain at higher levels is a "marauder's map," similar to the magical object in the Harry Potter series. A common epidemiological practice is to map the spread of disease by putting color-coded thumbtacks on a map of the affected area. We plan to create an electronic version of this that allows students

to see a visual representation of the symptoms that they encounter through talking to residents. This will be instrumental in helping students see the "hot spots" of the various diseases and to discover patterns in the spread of disease.

As the discussion above illustrates, in our design of River City we are fusing automation and individualization to address issues related to scalability. We can infuse various forms of individualization to handle variations in local context, while relying on the standardized nature of mediated immersion in this virtual environment to produce similar, foundational learning experiences across all participants. But how will we know if we are succeeding in our attempts to design for scalability?

A Proposed "Scalability Index" that Estimates Relative Scalability of Innovations

Knowing if we are succeeding in our attempts to design for scalability is difficult, as there are not any standardized methods for measuring the scalability of an innovation in the field of education. However, we speculate it might be possible through the development of a quantitative index that measures the relative scalability of an innovation across a wide spectrum of variations in context. By identifying factors within the intervention's context that represent important conditions for success and summarizing the extent to which the effect of the intervention is sensitive to variation in each, the proposed "scalability index" would provide prospective adopters of an innovation with a better sense of what its likely effectiveness would be in their own particular circumstances. It would also be of potential value to researchers in the Learning Sciences as it would allow one to determine which innovations retain much of their effectiveness under adverse conditions.

An initial step that is essential to creating a viable scalability index is the careful specification of a sensible framework of contextual factors that represent possible general conditions for success of educational innovations. Various scholars have already begun researching this issue. For example, Russell, Bebell, and O'Dwyer (2003) have studied a variety of factors thought to influence the conditions for success of the implementation of instructional technology in school districts. Fortunately, for many types of innovations, we believe that a relatively small set of contextual factors is often very influential in determining effectiveness. This leads us to conclude that examining scalability in the context of this subset of powerful conditions for success may still yield a workable index. We believe that potential influential factors to be included in the subset include teachers' knowledge of content and pedagogy, students' socioeconomic and linguistic backgrounds,

students' mobility and absenteeism, and (for technology-based innovations) the extent and reliability of the computer/networking infrastructure.

Statistical Approaches to the Creation of a Scalability Index

The evaluation of the sensitivity of an intervention's impact to select contextual conditions is really a question of statistical *interactions*. In evaluating the sensitivity to the conditions for success, one might ask: Is the effect of the intervention dependent on the selected contextual conditions? Is the intervention more effective for children of lower SES, or higher? Does the impact of the intervention depend on important teacher qualities or on features of the classroom and school infrastructure? In a single study, such questions are usually addressed by interactions between the treatment and its conditions for success in the statistical model.

One approach, then, to exploring the feasibility of creating a scalability index is to ensure that such interactions are included in the statistical models that underpin the data-analyses conducted to assess the implementation of educational interventions. If the interactions have a statistically significant effect, then we know that the effect of the treatment is sensitive to the conditions that participated in the interaction. Having successfully tested for the presence of such an interaction—with student SES, teacher quality or educational infrastructure—one can then estimate the several effect sizes² that can be anticipated for the intervention under each of the interacting conditions and pool them into a global index of scalability that captures the extent to which the intervention's effect size is sensitive to variation in the conditions for success. Whether or not such a scalability index is feasible as a generalizable measure is a larger issue we are currently studying in the context of our research on specific issues in the scalability of River City.

Conclusion

Bringing a technology innovation to scale in education requires a design that is flexible enough to be used in a variety of contexts and robust enough to retain effectiveness in settings that lack conditions for

²Cohen (1988) states that effect size is "the degree to which the phenomenon is present in the population" or "the degree to which the null hypothesis is false" (pp.9–10). Effect size is used in order to determine the "power" of an intervention or how large a sample one might need to get the "power" that they want.

success. All such designs for scale have some limitations. For example, we can design our River City learning experience to engage unmotivated students and encourage them to attend school regularly, but that design will not reach learners so unengaged that they refuse to attend school even to experience our curriculum. Within these limits and as illustrated through our work with River City, we believe that a fusion of automation and individualization is an effective approach to Design-Based Research for scalability. □

References

- Akerhielm, K. (1995). Does class size matter? *Economics of Education Review*, 14(3), 229–241.
- Baylor, A. L. (1999). Intelligent agents as cognitive tools. *Educational Technology*, 39(2), 36–40.
- Baylor, A. L. (2002). Agent-based learning environments for investigating teaching and learning. *Journal of Educational Computing Research*, 26(3), 249–270.
- Boozer, M., & Rouse, C. (2001, July). Intraclass variation in class size: Patterns and implications. *Journal of Urban Economics*, 50(1), 163–189.
- Brush, T., & Saye, J. (2000). Implementation and evaluation of a student-centered learning unit: A case study. *Educational Technology Research & Development*, 48(3), 79–100.
- 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.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- deCharms, R. (1968). *Personal causation*. New York: Academic Press.
- Dede, C. (2004). *Design for defenestration: A strategy for scaling up promising research-based innovations*. Chicago: NORC.
- Dede, C., Honan, J., & Peters, L. (Eds.). (2005). *Scaling up success: Lessons learned from technology-based educational innovation*. San Francisco: Jossey-Bass.
- Dede, C., Clarke, J., Ketelhut, D., Nelson, B., & Bowman, C. (2005). Fostering motivation, learning, and transfer in multi-user virtual environments. Paper presented at the American Educational Research Association Conference, Montreal.
- Dede, C., Nelson, B., Ketelhut, D., Clarke, J., & Bowman, C. (2004). Design-based research strategies for studying situated learning in a multi-user virtual environment. *Proceedings of the 2004 International Conference on Learning Sciences* (pp. 158–165). Mahwah, NJ: Lawrence Erlbaum Associates.
- Johnson, W. L., Rickel, J., & Lester, J. (2000). Animated pedagogical agents: Face-to-face interaction in interactive learning environments. *International Journal of Artificial Intelligence in Education*, 11, 47–78.
- Ketelhut, D., Clarke, J., Dede, C., Nelson, B., & Bowman, C. (2005). Inquiry teaching for depth and coverage via multi-user virtual environments. Paper presented at the National Association for Research in Science Teaching Conference.
- Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. New York: Cambridge University Press.
- Lepper, M., & Henderlong, J. (2000). Turning "play" into "work" and "work" into "play": 25 years of research on intrinsic versus extrinsic motivation. In *Intrinsic and extrinsic motivation: The search for optimal motivation and performance* (pp. 257–307). San Diego: Academic Press.
- Lester, J. C., Towns, S. G., & FitzGerald, P. J. (1999). Achieving affective impact: Visual emotive communication in lifelike pedagogical agents. *The International Journal of Artificial Intelligence in Education*, 10(3–4), 278–291.
- Malone, T. W., & Lepper, M. R. (1987). Making learning fun: A taxonomy of intrinsic motivations for learning. In R. E. Snow & M. J. Farr (Eds.), *Aptitude, learning, and instruction: III. Cognitive and affective process analyses* (pp. 223–253). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Moreno, R., & Mayer, R. E. (2004). Personalized messages that promote science learning in virtual environments. *Journal of Educational Psychology*, 96, 165–173.
- Moreno, R., & Mayer, R. E. (2005). Guidance, reflection, and interactivity in an agent-based multimedia game. *Educational Psychology*, 97, 117–128.
- Moreno, R., Mayer, R. E., & Lester, J. C. (2000). Life-like pedagogical agents in constructivist multimedia environments: Cognitive consequences of their interaction. ED-MEDIA 2000 proceedings (pp. 741–746). Charlottesville, VA: AACE Press.
- Nelson, B. (2005). Investigating the impact of individualized, reflective guidance on student learning in an educational multi-user virtual environment. Unpublished dissertation, Harvard University.
- Nelson, B., Ketelhut, D. J., Clarke, J., Bowman, C., & Dede, C. (2005, Jan.–Feb.). Design-based research strategies for developing a scientific inquiry curriculum in a multi-user virtual environment. *Educational Technology*, 45(1), 21–27.
- Rice, J. K. (1999, Summer). The impact of class size on instructional strategies and the use of time in high school mathematics and science courses. *Educational Evaluation and Policy Analysis*, 21(2), 215–229.
- Russell, M., Bebell, D., & O'Dwyer, L. (2003). *Use, support, and effect of instructional technology study: An overview of the USEIT study and the participating districts*. Boston: Technology and Assessment Study Collaborative; <http://www.bc.edu/research/intasc/studies/USEIT/description.shtml>
- Ryan, R., & Deci, E. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55(1), 68–78.
- Sack, J. L. (2002, Feb. 27). Smaller classes under scrutiny in California schools. *Education Week*.
- Wiske, M. S., & Perkins, D. (2005). Dewey goes digital: Scaling up constructivist pedagogies and the promise of new technologies. In C. Dede, J. Honan, & L. Peters, (Eds.), *Scaling up success: Lessons learned from technology-based educational innovation*. San Francisco: Jossey-Bass.