Robust Designs for Scalability

Jody Clarke
Chris Dede
Harvard University

Paper provided for discussion at the AECT Research Symposium
June 22-25, 2006, Bloomington, Indiana

Abstract:

One-size-fits-all educational innovations do not work because they ignore contextual factors that determine an intervention's efficacy in a particular local situation. Identifying variables within the intervention's setting that represent important conditions for success and summarizing the extent to which the impact of the intervention is attenuated by variation in them can provide prospective adopters of the innovation a better sense of what level of effectiveness they are likely to enjoy in their own particular circumstances. This study presents a research framework on how to conduct such an analysis and how to design educational innovations for scalability through enhancing their adaptability for effective usage in a wide variety of settings. The River City MUVE, a technology-based curriculum designed to enhance engagement and learning in middle school science, is presented as a case study.

Key discussion points or questions for session:

- What other conditions of success are important in technology innovations?
- What types of research studies will help to determine the full range of conditions for success?
- What are some strategies for collecting contextual data and determining conditions for success across various projects?
- What are other heuristics for robust design?
- As a field, how can we share heuristics for robust design and scalability?

Introduction

"Scaling up" involves adapting an innovation successful in a local setting to effective usage in a wide range of contexts. In contrast to experiences in other sectors of society, scaling up successful programs is very difficult in education (Dede, Honan, and Peters, 2005). For example, the 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 different students. Research findings typically show substantial influence of contextual variables (e.g., the teacher's content preparation, students' socioeconomic status) in shaping the desirability, practicality, and effectiveness of educational interventions. Identifying variables within the intervention's setting that represent important conditions for success and summarizing the extent to which the impact of the intervention is attenuated by variation in such variables can provide prospective adopters of the innovation a better sense of the level of effectiveness they are likely to enjoy in their own particular circumstances. In addition, this type of analysis can help designers improve the robustness of their innovations by developing "hybrid" versions optimized for success in particular types of settings.

In this paper, we discuss a research framework for how to design for scale in education. As a case study, we offer our research on the River City MUVE curriculum, a technology-based innovation designed to enhance engagement and learning in middle school science. Our intention in presenting these ideas is to provide an analytic framework that can be used to document and compare the impact that various contextual factors have on the success of an intervention across various projects. Heuristics from such analyses may aid designers in developing treatments more effective and scalable, particularly in inhospitable settings. Our

suggestions are intended to help policymakers, practitioners, and researchers anticipate the level of resources that might reasonably be invested in scaling up an innovation for effectiveness in a particular context.

In the sections that follow, we first document the importance of designing educational innovations for scalability and of making informed judgments about the suitability of an innovation for a particular context. Next, we present a conceptual framework for depicting various dimensions important for scalability. Then, we identify critical contextual features that research documents are important in the effectiveness of technology-based educational interventions. Following this, we describe quantitative analytic methods for determining the attenuation of an intervention's effectiveness by local variation in its conditions for success. With an overall model established for research on scalability, we present early results of applying this framework to a particular case, our River City curriculum. We then discuss implications for robust design that retains effectiveness under inhospitable conditions. Finally, we discuss conclusions and implications for further research.

The Importance of an Innovation's Scalability

Numerous studies have documented that it can be difficult to scale up promising innovations from the fertile, greenhouse environments in which they were conceived to the barren contexts that exist in schools with poor facilities, overwhelmed teachers, and struggling students (Dede, Honan, and Peters, 2005). Adapting a locally successful innovation to a wide variety of settings—while maintaining its effectiveness, affordability, and sustainability—is even more challenging.

In general, the more complex the innovation and the wider the range of contexts, the more likely a new practice is to fail the attempt to cross the chasm between its original setting

and other sites where its implementation could potentially prove valuable (Moore, 1999). In other words, scalable designs for educational transformation must avoid what Wiske and Perkins (2005) term the "replica trap": the erroneous strategy of trying to repeat everywhere what worked locally, without taking account of local variations in needs and environments. Without advances in design for scalability and in the assessment of an innovation's suitability for a particular context, education will continue to waste substantial resources implementing interventions that fail despite promise shown elsewhere.

Dimensions of Scale

In the context of innovations in teaching/curriculum, Coburn (2003) defines scale as encompassing four interrelated dimensions: depth, sustainability, spread, and shift in reform ownership. "Depth" refers to deep and consequential change in classroom practice, altering teachers' beliefs, norms of social interaction, and pedagogical principles as enacted in the curriculum. "Sustainability" involves maintaining these consequential changes over substantial periods of time, and "spread" is based on the diffusion of the innovation to large numbers of classrooms and schools. "Shift" requires districts, schools, and teachers to assume ownership of the innovation, deepening, sustaining, and spreading its impacts. We propose a fifth dimension to extend Colburn's framework, "evolution." "Evolution" is when the adopters of an innovation revise it and adapt it in such a way that it is influential in reshaping the thinking of its designers. This in turn creates a community of practice between adopters and designers whereby the innovation evolves.

Viewing the process of scaling from a design perspective suggests various types of activities to achieve scale along each dimension:

- *Depth:* evaluation and research to understand and enhance causes of effectiveness
- *Sustainability:* robust-design to enable adapting to inhospitable contexts
- *Spread:* modifying to retain effectiveness while reducing resources and expertise required
- *Shift:* moving beyond "brand" to support users as co-evaluators, co-designers, and co-scalers
- *Evolution:* learning from users' adaptations to rethink the innovation's design model

In particular, design for *sustainability* centers on the issue of contextual variation and involves designing educational innovations to function effectively across a range of relatively inhospitable settings (Dede, 2006). This is in contrast to models for effective transfer of an innovation to another context that involve partnering with a particular school or district to make that setting a conducive site for adapting a particular design. Scalability into typical school sites that are not partners in innovation necessitates developing interventions that are "ruggedized" to retain substantial efficacy in relatively barren contexts, in which some conditions for success are absent or attenuated. Under these circumstances, major aspects of an innovation's design may not be enacted as intended by its developers. While all the dimensions above relate to designing to scale, this study focuses on the sustainability dimension and robust-design.

We do not expect that interventions created for use in multiple settings through robust-design strategies will outperform an intervention designed for specific classrooms that have all the necessary conditions for success. Similarly, a biotechnologist would not expect to design a plant strain specifically tailored to arid climates that in typical contexts somehow also outperforms those strains that have evolved for more ideal conditions of moisture. The strengths of ruggedized interventions are likely weaknesses under better circumstances; for example, high levels of support for learner help and engagement that aid unengaged pupils with low prior preparation could well be intrusive overhead for better-prepared, already motivated students. We are currently conducting design-based research on our River City MUVE curriculum(described

later in the paper) to explore whether robust-design can produce the educational equivalent of plant strains tailored to harsh conditions that are productive where the usual version of that plant would wither and die.

However, the robust-design approach has intrinsic limits, as some essential conditions that affect the success of an educational innovation cannot be remediated through ruggedizing. As an illustration of an essential condition for success whose absence no design strategy can remediate, for implementations of the River City MUVE curriculum in some urban sites, student attendance rates at classes typically averaged about 50% prior to the intervention. Although attendance in science class improved by 60% during the implementation of the curriculum, an encouraging measure of its motivational effectiveness through robust-design, clearly the curriculum nonetheless had little value for those students who seldom came to school during its enactment. Further, in the shadow of high stakes testing and accountability measures mandated by the federal No Child Left Behind legislation, persuading schools to make available multiple weeks of curricular time for a single intervention is very hard. Essential conditions for success such as student presence and district willingness to implement pose challenges beyond what can be overcome by the best robust-designs.

However, design-based researchers can potentially still get some leverage on these essential factors. For example, as we will discuss later, the River City MUVE curriculum is engaging for students and teachers, uses standards-based content and skills linked to high stakes tests, and shows strong outcomes with sub-populations of concern to schools worried about making adequate yearly progress across all their types of students (Dede, Clarke, Ketelhut, Nelson, Bowman, 2005). These capabilities help surmount issues of student involvement and

district interest, giving our intervention traction in settings with low student attendance and a focus on test-preparation.

Identifying Critical Contextual Features

Leveraging the sustainability dimension of scale involves identifying critical contextual features of the design's effectiveness and testing for their interaction with the treatment. How does one determine these conditions for success? What type of data collection is necessary? Based on relevant theory, research, and practice, we are developing a taxonomy of critical contextual variables for technology-based educational innovations (Dede, Willett, Clarke, Moody, in press). In order to build this taxonomy, we are drawing both from the professional literature for practitioners on technology integration (http://www.ncrel.org/engauge/intro/intro.htm) and from the research literature (empirical studies, theoretical, meta-analyses, evaluations) on the types of factors that investigators have identified as crucial for their technology-based innovations.

For example, we examined all of the studies of educational interventions published between 2002 and 2004 in three leading educational technology journals (*Journal of Learning Sciences, Journal of Educational Computing Research*, and the *Journal of Science, Education, and Technology*), as well as other prominent work in the field (Waxman et al, 2003; Russell, 2003; Means and Penuel, 2005; Pearson et al., 2005), compiling a list of the variables included in the analyses. We are triangulating these various sources and approaches and have developed the following "work-in-progress" list of contextual variables:

Student Level Variables
- Developmental level (i.e., age)
- Socio-economic status (SES)
- Race or ethnicity
- Gender

- Level of general education (i.e., grade level in school)
- Previous academic performance (includes achievement scores, GPA, class ranking)
- Absentee record
- Teacher expectation of performance
- Major (if higher education)
- Affective measures (e.g., level of engagement and of self-efficacy)
- Technology familiarity and fluency (includes level of comfort or anxiety, technology self-efficacy)
- Collaboration skills (for innovations involving group work)
- Population of students (general population, special ed)
- Students' home technology access and usage

Teacher Level Variables

- Years of teaching experience
- Academic specializations (e.g., certified in science)
- Professional development related to the innovation
- Support related to the innovation
- Technology familiarity and fluency
- Teacher's Pedagogical Beliefs
- Teacher's Beliefs about using technology in classroom
- Teacher Demographic Characteristics
- Teacher ownership of content/making personally relevant to students
- Teacher comfort level using innovation
- Teachers' assessment that activities are congruent with the content of standards and high-stakes tests

Technology Infrastructure Conditions

- Access to educational materials:
- Location of technology (lab outside classroom, lab in classroom, computer cart brought in, students school laptops)
- Reliability and quality of technology (machine level, network/server connection reliability)
- Access to technology equipment
- Type of technology used (personal computers, networked laboratories, multimedia, other

School/Class Variables:

- School characteristics (e.g., public/private)
- Type of class schedule (rotating, block)
- Length of unit: amount of time students have to use technology
- Type of class (honors, remedial, college prep)
- Sample size (number of students/classes/teachers)

- Support for teachers from an administrator who is positive about the program and can help garner resources, reduce institutional barriers (such as securing transportation to study sites), and showcase activities
- Support and mentoring from other teachers

By identifying features 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 (e.g., how each feature interacts with the treatment), prospective adopters of the innovation will have a better sense of its likely overall effectiveness in their particular circumstances. Also, policymakers will gain a better sense of which innovations are most promising in their potential for adaptability to a wide range of settings, and researchers can contrast the scalability of various types of innovations, thereby attaining insights into how to improve design and implementation. In order to summarize the extent to which the intervention is sensitive to contextual variables, we propose utilizing a scalability index, calculated for comparability across projects parallel to current measures such as effect size.

Developing a 'Scalability' Index for Innovations

The establishment of the Education Sciences Reform Act of 2002, in addition to creating the U.S. Education Department's Institute of Education Sciences (IES), is fostering a shift in what kinds of evidence constitute "proof" of an innovation's effect. A report for IES prepared by the Coalition for Evidence-Based Policy (2003) states that "evidence-based" research entails randomized controlled experiments that report the size of a treatment's effect (effect size). There are various statistical methods for calculating an effect size (ES). The most common is Cohen's d (Cohen, 1988), where the difference between the mean of two groups is divided by the pooled standard deviation of the two groups:

$d = M_1 - M_2 / \sigma_{pooled}$

Cohen (1988) offered conventional definitions of the ES for describing the magnitude of the effect (p.40):

Small: d=.20
Medium: d=.50
Large: d=.80

It is important to note that in the social sciences, even the most successful interventions have small effect sizes (Light, Singer, Willett, 1990).

However, publishing one overall effect size does not allow policy makers to determine the likely effectiveness of the innovation in their local setting where students, teachers, and resources may vary from the conditions for success of the innovation, ideal conditions under which its effect size was calculated. Simply reporting this metric does not provide decision makers and designers with enough information to leverage the sustainability dimension of scale. Therefore, we are currently exploring the utility of a "Scalability Index" that estimates the relative sensitivity of innovations to attenuation in various dimensions that represent their conditions for success. For example, if the effect of an innovation was found to have an interaction with class size (number of students in the classroom), then an index that summarized the extent to which the innovation's effect was attenuated when used in classrooms with more than, say, 30 students would provide valuable input into decision making. The proposed index would summarize, along each dimension of its major conditions for success, the degree to which the educational effectiveness of the design is robust to a shortfall in that dimension.

We are currently developing analytic strategies for creating our proposed "scalability index" based on Cohen's (1988) f^2 standardized framework for estimating the size of the effect associated with interactions. We are applying this to estimate the interaction between a

treatment and a condition for success. One approach to calculate Cohen's f^2 is to use the R^2 statistic, which measures the proportion of the variation in the outcome that is explained by the predictors or predicted by the predictors in a regression model. The f^2 is also a measure of proportion in variation of the regression model with the contextual variables and their interaction with treatment:

$$f^2 = \frac{R^2_{full} - R^2_{reduced}}{1 - R^2_{full}}$$

One can compare the *square root* of the f^2 index to Cohen's conventional definitions of the ES for describing the magnitude of the effect, as discussed earlier (Cohen, 1988, p.40). Using this method, conditions for success that generate large values of the f^2 statistic in interaction with the treatment would suggest that the impact of the intervention is very sensitive to this particular condition for success. We plan to present these scalability index data in tables where the effect size of the interactions between treatment and various conditions are provided (presenting the square root of the f^2 statistic). Then, decision makers can carefully consider the impact of the particular condition before implementing the intervention at new sites. In addition, designers can use this information to develop "hybrid" versions of the innovation where the design is flexible enough to be optimized for success under various conditions.

River City MUVE as a Case Study

Leveraging the sustainability dimension of scale involves identifying critical contextual features of the design's effectiveness and testing for their interaction with treatment. In order to design for scale, designers should develop flexible models with variants adapted to a spectrum of
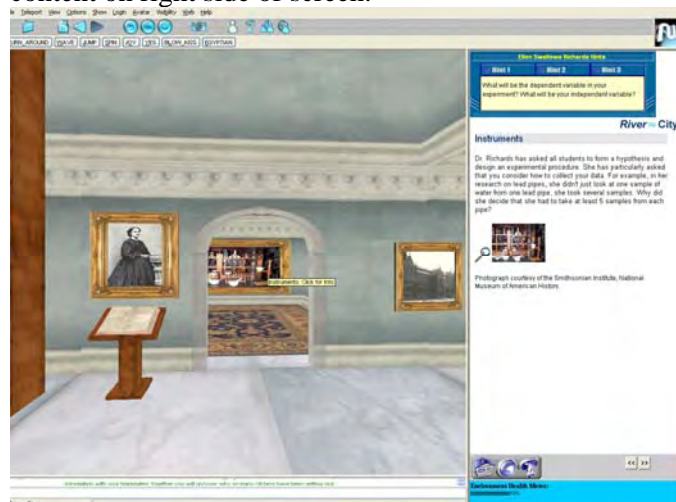
implementation conditions. As a case study illustrating this design strategy, we offer our own

research on the scaling up of River City, a technology-based curriculum.

River City is a multi-user virtual environment (MUVE) designed to teach middle school

science (http://muve.gse.harvard.edu/rivercityproject/). 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, (see figure 1) and enact collaborative learning activities of various

types (Nelson, Ketelhut, Clarke, Bowman, Dede, 2005). When participants click on an object in

the world, the content appears in the right hand interface (see figure 2).

Figure1: Talking to River City Residents

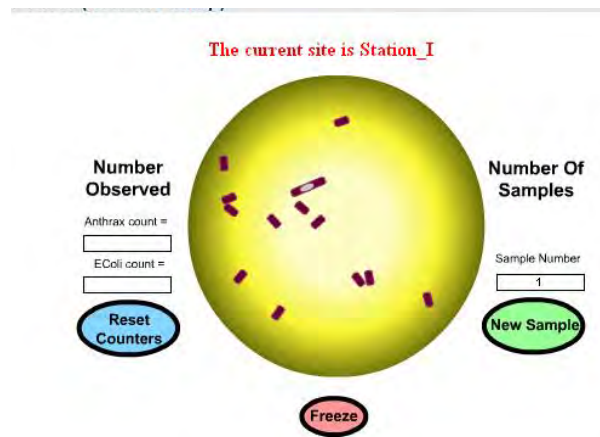Figure2: View of 3-D environment and web-based content on right side of screen.

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. Middle school students travel back in time to the 19th century and use their 21st century

knowledge and skills to help the Mayor figure out why the residents of River City are getting

sick (see figures 6-8 at end of paper for more views of the environment).

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). Students work in teams of three or four to develop and test their hypotheses about why residents are ill. Students learn about and use to the tools of scientists as part of their scientific exploration. For example, they can use the virtual microscope to take water samples (see figures 3 and 4 below) from one of the water sampling stations in the city.

Figure 3: Taking a water sample with the virtual microscope. Students click on the "Water Sampling Station" green sign and a microscope appears in the right hand interface with moving EColi.

Figure 4: Close up of Microscope. Students click "Freeze" and count the number of EColi and Anthrax in the water. After counting up the EColi, students can take a new water sample and a new count.

Research Design

Utilizing design based research strategies, we have conducted a series of studies in which we explored the types of learning that MUVEs appear to foster. We present results from these studies to illustrate how we identified contextual factors and what design elements we modified to overcome any barriers. At varied intervals, we implemented different treatments of River City based on alternative theories of how people learn: guided social constructivism, expert mentoring and coaching, legitimate peripheral participation (communities of practice), and levels of guidance. For a detailed description of these different studies and treatments, please see (Nelson, Ketelhut, Clarke, Bowman, 2005 and Clarke, Dede, Ketelhut, and Nelson, 2006). To date, we have worked with over 70 teachers and 6000 students primarily in urban public middle schools with high proportions of ESL low SES students. This paper focuses on a sub-sample of 7 teachers and more than 600 students who implemented River City in fall 2004.

We employed quasi-experimental design strategies where students were assigned to either treatment or control at the classroom level. Within the treatment classrooms, students were randomly assigned to one of our learning variants at the student level. The control

curriculum was similar in pedagogy and content, yet delivered via paper-based curriculum. Each teacher offered both the computer-based treatments and the control.

We collected demographic data on students from teachers, including age, prior science grades, free and reduced lunch status (as measure of SES), native language, race, gender, reading level, attendance, and teacher expectation of students along 5 dimensions: content mastery, motivation, behavior, absence, and technology mastery. (Please note that some teachers did not supply this information; therefore, we were unable to test for interactions between some of these variables and treatment. We discuss later how we have modified our implementation to prevent this lack of data in the future.)

Students were administered an affective measure that was adapted from three different surveys, Self-Efficacy in Technology and Science (Ketelhut, 2004), Patterns for Adaptive Learning Survey (Midgley, 2000), and the Test of Science Related Attitudes (Fraser, 1981) pre- and post-intervention. This modified version has scales to evaluate students' science efficacy, thoughtfulness of inquiry, science enjoyment, career interest in science, etc. At the end of this survey, they were also asked about their technology access and usage outside of school. To assess understanding and content knowledge (science inquiry skills, science process skills, biology), we administered a content test pre- and post-intervention, with sections modified from Dillashaw and Okey, (1980).

Teachers were administered questionnaires pre and post intervention that asked about their years of teaching experience, academic specializations, professional development in technology, technology fluency and familiarity, pedagogical beliefs and practices around science and technology, comfort level with technology, comfort with project, and opinion of level of support provided by our research team.

The backend architecture of the MUVE environment is a database. This database has an event table that captures detailed moment-by-moment individual student activity in the environment, such as which residents they talked to, what pictures they clicked on, how long they spent in a given location, etc. These data-rich files were captured for all students in the experimental treatment. In addition, as a final performance demonstration, all students completed a report to the Mayor of River City that explained their experiment and findings.

The quantitative data was analyzed with SAS using multiple regression techniques. Checks for linearity and homoscedasticity were performed. No violations were detected. A significance level of $p < .10$ was used.

At its core, the evaluation of the sensitivity of an intervention's impact to select contextual conditions is a question of statistical interactions. In evaluating the sensitivity to the conditions for success, one asks: Is the effect of the intervention dependent on the selected contextual conditions? Is the intervention more effective for children of lower SES, or higher? In our research, such questions were addressed by the inclusion in the statistical models of interactions between the treatment and its conditions for success. We present results from these studies to illustrate how we identify contextual factors and what design elements we modify to overcome any barriers.
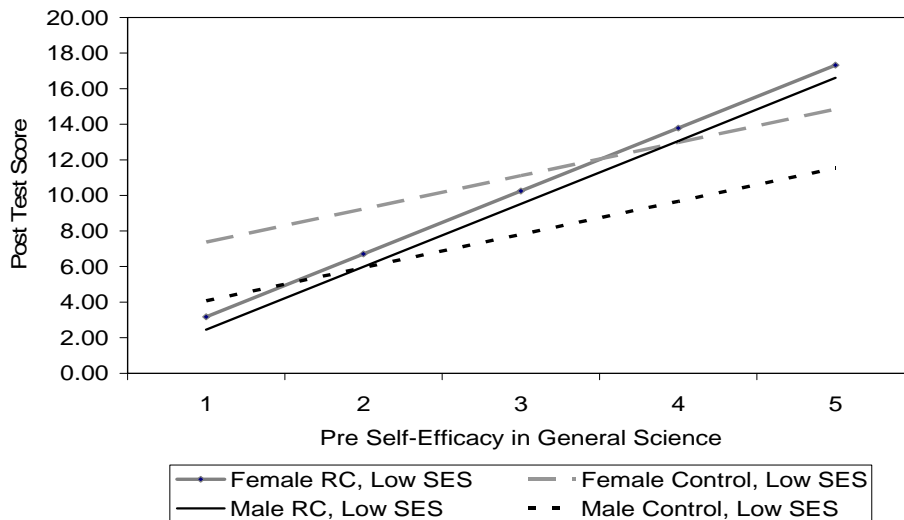
Findings

Results from this implementation supported earlier findings that students in the River City treatment are engaged in scientific inquiry and in higher order thinking skills (see Nelson et al, 2005; Clarke et al, 2006). We found a positive relationship between students' starting self-efficacy in general science and their post test score. The magnitude of this relationship was greater for students in the River City treatment. We did not find any significant differences

between the four River City treatments and therefore present analyses looking at the River City treatment versus the control curriculum. As seen in figure 5 below, the impact of the River City treatment on students' post-test score, controlling for SES, differed depending on students' gender and entering self-efficacy in general science. SES explains variance in the outcome, total post test score (t= -7.53***), but does not have a statistical interaction with the treatment. Therefore SES is not a condition for success but is included as a control variable because it helps to explain the variance. For example, for students with low SES, girls in the control who entered the project with low self-efficacy in general science performed better on the post-test than girls in River City with similar low self-efficacy in general science, on average. However, for students with low SES, girls in River City who entered the project with high self-efficacy (~4) outperformed everyone on the post test score, on average. Within the each of the two treatments, for students with low SES, girls outperformed boys on the post test, on average.

Figure 5. Effect of Treatment on Post Test Score, Controlling for Male, SES, and Pre Self-Efficacy in General Science (n=362).

17

The results for students with high SES are parallel but slightly higher (the interaction between SES and treatment was not significant). The interactions between some of the contextual variables and treatment suggest that there are conditional factors that may be considered conditions for success when implementing the River City project.

We are just starting to explore and identify these contextual factors. For the sake of space, in this paper we focus on student level variables from our working list presented above. In order to know whether or not these variables were conditions for success, we performed analyses to test whether or not they had a statistically significant interaction with treatment. Table 1 summarizes the results of the tests for interactions and the extent to which the effect of the intervention is sensitive to variation in each. Column 1 lists the student contextual variables, and column 2 states whether or not there was a statistical interaction. Statistical interactions were measured through multiple regression analysis; Column 3 presents the t value and p value of the parameter estimate of the interaction in the model. Column 4 presents the square root of the f^2 index, calculated from the R^2 statistic, using the formula explained above.

Table 1: Student Contextual Variables and their Statistical Interaction with Treatment

Student Level Contextual Variables	Interaction with Treatment	Parameter Estimate (t value, p value)	$\sqrt{f^2}$
Student Demographics			
Socio-economic status (SES)	No	t= -0.06	
Race	No	t= 0.28	
Gender	Yes	t=1.69~	0.13
Native English Speaker	No	t=-0.81	
Level of general education (i.e., grade level in school)	No[+]	t=0.70	
Affective measures Entering the Project			
Enjoyment of Science	Yes	t=2.84**	0.24
Self-Efficacy in General Science	Yes	t=2.11*	0.38
Career Interest in Science	Yes	t=3.89***	0.32
Thoughtfulness of Inquiry in Science	No	t=0.92	
Science Inquiry	No	t=1.60	
Students home and technology access and usage			
Internet access at home	No	t=0.60	
Use chat programs	No	t= 0.30	
Play computer games	No	t=0.53	
Play video games	No	t=.12	

Key: ~ p<.10; * p<.05; ** p<.01; *** p<.001

[+]Please note we did not detect an interaction between treatment and any grade level, but due to space we only present the t value for the interaction between 7th grade and treatment.

As can be seen in the rows 4, 8, 9, and 10 of column 2, only four of the student contextual variables represent important conditions for success, as determined by statistical interaction with treatment in this sample. According to Cohen's conventions, all of the effects of the interactions with treatments in this sample are "small." This is not surprising. As mentioned above, in education, the effect sizes of even the most successful interventions tend to be "small" (Singer & Willett).

Looking at column 4, we can create an index from least to highest sensitivity, where Gender is the least sensitive (.13) and "Entering Self-Efficacy in General Science" is the most

sensitive (.38). In addition to providing insights on student learning in the project, these findings are helping us to discover which factors we need to modify in order to improve the 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. In the following section, we discuss how this analysis of the conditions for success is helping us to make our design more robust, so that we can scale our innovation across a wide variety of educational settings.

Designing Framework for Scalability and Ruggedization

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. As described above, we are 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 prior research, (Clarke, Dede, Ketelhut, Nelson, 2006) we identified four factors important in the enactment of the River City curriculum and discussed how we were evolving the curriculum's design to allow for individualization that enhances the effectiveness under those conditions. Our design process has evolved, and therefore the following section provides an updated look of these conditions for success combined with new findings presented above. Our first 2 conditions for success are concerned with teacher level conditions around teacher ownership of content and teacher comfort level using innovation. Our discussion for how we made the design robust for these conditions focuses on the infrastructure and back-end

architecture of the design. Our other conditions for success focus on student conditions for success and design features of the curriculum.

Teacher Conditions for Success

Professional development

Through design-based research, we have been testing various ways of providing professional development for teachers (Clarke et al, 2006). We provided online training for some teachers, but found that most of them did not access it. We also knew that, as we scaled, we would not be able to provide just-in-time advice for teachers. Therefore, we are now testing how a "train the trainer" model of professional development scales. In this model, members of our research team provide professional development for trainers. Each trainer works with up to 10 teachers in a district, providing professional development and monitoring their implementations. The trainer is available to provide "just in time" support for teachers and also serve as the "eyes and ears" of the research team in classrooms where the project is implemented.

Teacher Ownership and Comfort Level

We have found that integrating the River City project, a technology intensive project, pushes the limits of teacher's comfort level. In the post surveys and interviews, some teachers have expressed that they felt more comfortable with the control curriculum because they felt like they had more "control." We have been searching for ways to provide teachers with more autonomy and empowerment in the River City project, a condition that we feel leads to success of the implementation. Our result is a "Teacher Dashboard" that provides teachers with all the tools and mechanics necessarily to successfully implement the River City project. This dashboard houses numerous resources and functions under one location (web page), so that

teachers only need to create a single book mark in their internet browser (the site is password protected).

Through the "Teacher dashboard", teachers can now create student accounts and passwords for the River City program. In the past, teachers did not always fulfill their obligations to provide us with the demographic data we need in order to looks for various conditions for success. Being able to determine the conditions for success relies heavily on the collection of appropriate data. In the results presented above, missing data on student prior academic achievement, reading scores, and demographics led to sample size about 25% smaller than it should have been. Therefore, with the "Teacher Dashboard", when the teacher creates each student account, they must enter demographic data about each student. After creating a class of students, they then assign the students to teams of 3. In the past, we created student accounts and randomly assigned student teams. These two steps are meant to provide teachers with more control over using the project with their students and also make the project easier to scale as the teachers rely less on the research team. We no longer receive any emails from teachers regarding student logins or complaints about team assignments.

As students move through the River City project, they advance through 6 different worlds chronologically that represent different seasons: October 1878, January 1879, April 1879, July 1879, the control world (controlled portion of their experiment) and their experimental world (testing their experiment). This allows them to collect data on change over time in River City. When students first enter the River City program, they enter a "time portal" and then click on the world to which they wish to travel back in time. While students are supposed to travel back and work through the project chronologically, this has not always been the case. Teachers have had very little control over the ability to make sure students are in the right world during a given

class period. Therefore, we added a "world access" feature that is now available in the "Teacher Dashboard." Teachers have the option to set world access at the class or student level. While this might seem minute, early feedback indicates that it has made a difference in terms of teachers feeling as though they have control over the curriculum and students' chronological movement through it. Most importantly, this feature is customizable—teachers do not even have to use it.

Our last infrastructure-related design for scale is another customizable feature meant to provide teachers with more control over student learning in the project. As mentioned above, the back end of the River City environment is a database. We use these data files to track students learning trajectory in the world. However, until now teachers have never had access to these data. While we encourage teachers to review student lab books and keep up with student progress via written work in the project, teachers have had little detailed knowledge about what each student is doing in the program itself. Inside River City, students work in teams and communicate with their team via a text-based chat. We are in the process of adding a feature to the "Teacher Dashboard" that allows teachers to run reports on the "team chat" at the student, team, and/or class level. A report will be generated that shows the world, avatar names, and team chat of the students. This optional feature will allow teachers to monitor their students' progress (whether they are actually on task) and language (whether or not they are using bad language). Again, this feature will be customizable for teachers and optional in usage.

In addition to these modifications in the architecture of the design, we have added resources such as "Individualized Professional Development" documents and "Day-by-Day" lesson planning for teachers to use as resources and "quick guides." We have also added short videos (less than a minute) that model what students are supposed to do each day in the

curriculum. For example, before students enter "January 1879," a teacher can show students the "January 1879 Video." The video reminds students what they have been doing in the project and connects it to what they will be doing in January—providing context and building a continuum in the curriculum. Each video has a similar look and feel. All footage was captured from the River City environment, and each video starts with the Mayor of River City (an avatar from the environment) talking to the students. These videos were created as a model for teachers of how to introduce each day's lesson, or as an introduction for students before they begin their daily activities.

Student Affective Conditions for Success

Our next set of robust design strategies are concerned with conditions for success measured at the student level. We have found that students enter the project with varying levels of engagement in science, self-efficacy in general science, background knowledge, and career interest in science. While we have designed our project to engage all students, we hope to reach students who are not engaged in science, don't feel good about their ability to do science, and have a history of low academic performance in science. Through design-based research, we have been studying ways to engage our target student population, but also to maintain engagement for students who are already performing well in science. We are developing three different strategies to meet these needs: powers, pedagogical agents, and roles.

Powers

Similar to features of videogame play that reward experiences and accomplishments by giving participants special powers, we have designed a system of powers that reward student learning and exploring in the River City environment. These powers provide students with access to further curricular information that will help them more deeply understand the spread of

disease in River City. As mentioned above, students travel through River City chronologically. They enter the city during four different seasons, collecting information about the spread of disease in the town. Next, they then develop a hypothesis based on their research and design an experiment to test their hypothesis. They then go back into River City (the control and experimental worlds) to test their experiment. Accompanying each world is a list of curricular objectives that guide students through the inquiry process. These curricular objectives have been modified into activities such that completion of them leads towards the attainment of powers.

For example, we want students to explore the different areas of River City and gather information about how the 3 diseases are more prevalent in different areas of the city. Therefore, in the spring world (April), some of the requirements to obtain powers involve visiting a certain location and talking to the residents in that location, or clicking on pictures or objects in that location. As an illustration, students learn a lot of important information when they visit the hospital. Once inside, they can talk to Nurse Patterson, Doctor Aaron Nelson, review the admissions records, and click on pictures that provide historical information about 19th century hospitals. Therefore, as a requirement for one step towards achieving powers for this given world, we have back-end Boolean statements that scan the database and register whether at least 1 team member has visited the hospital and interacted with a resident, object, or picture. Just like in a videogame, students are not told the requirements for powers, nor are they told that they exist. However, they are presented with the curricular objectives in their lab books and use the lab book to guide their discovery. Once a combination of team members has completed the specified curricular tasks, they earn "powers" and are teleported to a secret mansion in the city.

This secret mansion contains extra curriculum and is only accessible for teams of students who have earned "powers" by completing the curricular objectives. The first power

earns students access to the first floor entryway of the building. Each successive world's powers is another floor of the building; so students who achieve powers in, for example, April 1879 will have access to the hallway, the second story and the third story. Therefore, if students missed attaining powers in some previous level of the world, they can make up the missed learning by later attaining powers in a different level.

The "powers" for April 1879 involve access to curriculum that presents students with a historical look at the tools of scientists from the 19th century to 2005. The third floor of the mansion is a museum, and a sign welcomes students to click on the various tools. For example, they can click on an 1880 version of a microscope and then on a modern day microscope to see how much the tool has evolved and enables us to detect such things as microbes. The River City world has modern day microscopes that enable students to take water samples, but having this extra curriculum provides some insight about why the scientists in the 19th century were not able to see diseases caused by bacteria.

Students learn that, as better and better microscopes were invented, scientists were able to see microbes more clearly. However, even with the modern microscope, it isn't possible to see inside the body. Students learn about a modern tool that was invented that allows doctors and scientists to see into the body: a CAT scan. Students can click on the names of patients listed in the April hospital records to see what a CAT scan of their lungs looked like. When names are clicked on, if the person has tuberculosis (one of the 3 diseases in River City), then the CAT scan shows a diseased lung. If they do not have tuberculosis, then the CAT scan shows a healthy lung.

These powers are engaging for students while still providing rich content that furthers their knowledge. Powers illustrate an adaptation of the curriculum to aid students with low

motivation to succeed. We are now starting studies to determine if the inclusion of powers makes River City more effective in settings where many students are turned off to school.

Pedagogical Agents

In addition to maintaining student engagement, we have found it important to help some students gain more self-efficacy in their ability to do science as they move through the project. Therefore, we are in the process of developing pedagogical agents: "wise fools" that will work alongside students as they move through the River City curriculum. These "wise fools" will ask students questions that encourage them to reflect on their learning. Each team will have one "wise fool" working alongside them, agents who know to which of the residents team members have talked and on what artifacts and objects they have clicked. Using strategies from the literature on metacognition (Brown, 1987; Campione, 1987; Schoenfeld, 1992; White and Frederickson, 1998) and reciprocal teaching (Palincsar and Brown, 1984; Palincsar and Brown, 1985; Palincsar, 1986; Palincsar, Ransom, Derber, 1989), the "wise fools" will ask students for information and encourage student learning along the way. They will also offer information about the project (in addition to the guidance and hints developed by Nelson, 2005). We hypothesize that these conversations with the "wise fools" will help students learn and increase their self-efficacy as they move through the project.

Roles and collaboration

Collaboration is an important element in students' learning in River City. However, close examination of log files and interviews with students reveals that many students struggle with collaboration (Clarke, draft). Therefore, through design-based research we plan to develop curricular strategies that will encourage students to collaborate with their teammates. Under these conditions, each particular teammate will have access to different levels of information. In

order to fully understand what is causing the problems, the entire team will need to pool their information and findings. For example, one teammate might ask a particular resident how they are. The resident will respond differently depending on the team member who asks the question. To teammate A, the resident might respond, "I have a slight fever and my stomach aches. I was feeling fine yesterday, and was playing near the bog with my friend Emily." However, the resident might respond differently to teammate B, "I have a slight fever and my stomach aches." Or to teammate C, "I am sad. I could not go and play near the bog today with my friend Emily because I am sick. We have been playing by the bog all week." Just like in the real world, who you are determines the type of response you might receive from a person. We hypothesize that including these roles will engage students in trying to figure out what each of their teammates have discovered and foster richer collaboration via jigsaw pedagogy.

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 its conditions for success; this may involve developing variants that are the equivalent of hybrid plants designed for inhospitable locales. Designing an innovation for sustainability and scale is a multi-stage process. One must first identify possible contextual variables that lead to conditions for success of the innovation. Next one must collect data on these various conditions. Then one must test for the statistical interaction between these conditions and the treatment. Finally, one can develop variants of the design model effective under these conditions. As an illustration, through design-based research strategies we are determining what contextual variables constitute conditions for success in implementing River City and developing heuristics for robust variants in settings where those factors are attenuated or missing.

References

Brown, A. (1987). Metacognition, executive control, self-regulation, and other more

mysterious mechanisms. In F. E. Weinert & R. H. Kluwe (Eds.), *Metacognition, motivation, and understanding* (pp.60-108). Hillsdale, NJ: Erlbaum.

Campione, J. (1987). Metacognitive components of instructional research with problem

learners. In F. E. Weinert & R. H. Kluwe (Eds.), *Metacognition, motivation, and understanding* (pp. 117-140). Hillsdale, NJ: Erlbaum.

Clarke, J. (draft). Making Learning Meaningful: An Exploratory Study of Using MUVEs to

teach Middle School Science. Unpublished Qualifying Paper presented to the Harvard Graduate School of Education, Cambridge, MA.

Clarke, J., Dede, C., Ketelhut, D., Nelson, B. (2006). A Design-Based Research Strategy to

Promote Scalability for Educational Innovations. *Educational Technology*, 46(3), 27-36.

Coalition for Evidence-based Policy. (2003). *Identifying and Implementing Educational

Practices Supported by Rigorous Evidence: A User-Friendly Guide.* Washington, DC: Institute of Education Sciences, U.S. Department of Education.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences.* Hillsdale, NJ:

Lawrence Erlbaum Associates.

Dede, C., Honan, J., & Peters. L., (Eds). (2005). *Scaling Up Success: Lessons Learned from

Technology-Based Educational Innovation.* New York: Jossey-Bass.

Dede, C. (2005). Why design-based research is both important and difficult. *Educational Technology* 45, 1 (January-February), 5-8.

Dede, C. (2006). Scaling Up: Evolving Innovations beyond Ideal Settings to Challenging Contexts of Practice. In R.K. Sawyer (Ed.), Cambridge Handbook of the Learning Sciences, pp. 551-566. Cambridge, England: Cambridge University Press.

Dede, C., Willett, J.B., Clarke, J., Moody, L. (in preparation). Examining the Feasibility of a Scalability Index: Summarizing How Contextual Factors Mitigate the Impacts of Treatment in Technology Innovations.

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

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

Gallman, N. (2003) Web site update: NCREL releases new content on enGauge website. *NCREL Educational Technology News* 3(1), 3-4.

Ketelhut, D. (2004). Assessing Science Self-Efficacy in a Virtual Environment: a Measurement Pilot. Unpublished Qualifying Paper presented to the Harvard Graduate School of Education, Cambridge, MA.

Lesgold, A. (2003). Detecting technology's effects in complex school environments. In G. Haertel & B. Means (Eds.), Evaluating educational technology: Effective research designs for improving learning, pp. 38-74. New York: Teachers College Press.

Light, R. J., Singer, J. D., & Willett, J.B. (1990). *By Design: Planning Research on Higher Education*. Cambridge, MA: Harvard University Press.

Means, B., & Penuel, W.R. (2005). Scaling up technology-based educational innovations. In C. Dede, J. Honan, & L. Peters (Eds.), *Scaling Up Success: Lessons Learned from Technology-Based Educational Innovation*. New York: Jossey-Bass.

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.

Moore, G. A. (1999). *Crossing the chasm: Marketing and selling high-tech products to mainstream customers*. New York: HarperBusiness

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., 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.

Palincsar, A. S. and Brown, A. L. (1984): Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction*, 1(2), 117–175.

Palincsar, A.S., & Brown, A.L. (1985). Reciprocal teaching: Activities to promote read(ing) with your mind. In T.L. Harris & E.J. Cooper (Eds.), *Reading, thinking and concept development: Strategies for the classroom*. New York: The College Board.

Palincsar, A.S. (1986). Reciprocal teaching. In *Teaching reading as thinking*. Oak Brook, IL: North Central Regional Educational Laboratory.

Palincsar, A.S., Ranson, Derber (1989): Collaborative research and development of reciprocal

teaching. *Educational Leadership*, 46(4), 37–40.

Pearson, P. D., Ferdig, R. E., Blomeyer Jr., R. L., & Moran, J. (2005). *The Effects of Technology*

on Reading Performance in the Middle-School Grades: A Meta-Analysis With

Recommendations for Policy. Naperville, Illinois: North Central Regional Educational

Laboratory.

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,

MA: Technology and Assessment Study Collaborative. Retrieved August 29, 2004 from

http://www.bc.edu/research/intasc/studies/USEIT/description.shtml

Schoenfeld, A. H. (1992), "Learning to Think Mathematically: Problem Solving, Metacognition,

and Sense Making in Mathematics," in *Handbook of Research on Mathematics Teaching and*

Learning, ed. D. A. Grouws, NY: Macmillan, pp. 334-370.

Waxman, H. C., Lin, M.-F., & Michko, G. M. (2003). *A Meta-Analysis of the Effectiveness of*

Teaching and Learning With Technology on Student Outcomes. Naperville, Illinois: North

Central Regional Educational Laboratory.

White, B., & Frederiksen, J. (1998). Inquiry, modeling, and metacognition: Making science

accessible to all students. *Cognition and Instruction*, 16(1), 3-117.

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.

Views of River City

Figure 6: View of River City including all interface windows. The interactive map is shown on the right hand side. The yellow dot indicates where you are, and you can teleport to a location by clicking on the name. For example, to teleport to the wealthy homes, click on "Wealthy homes."
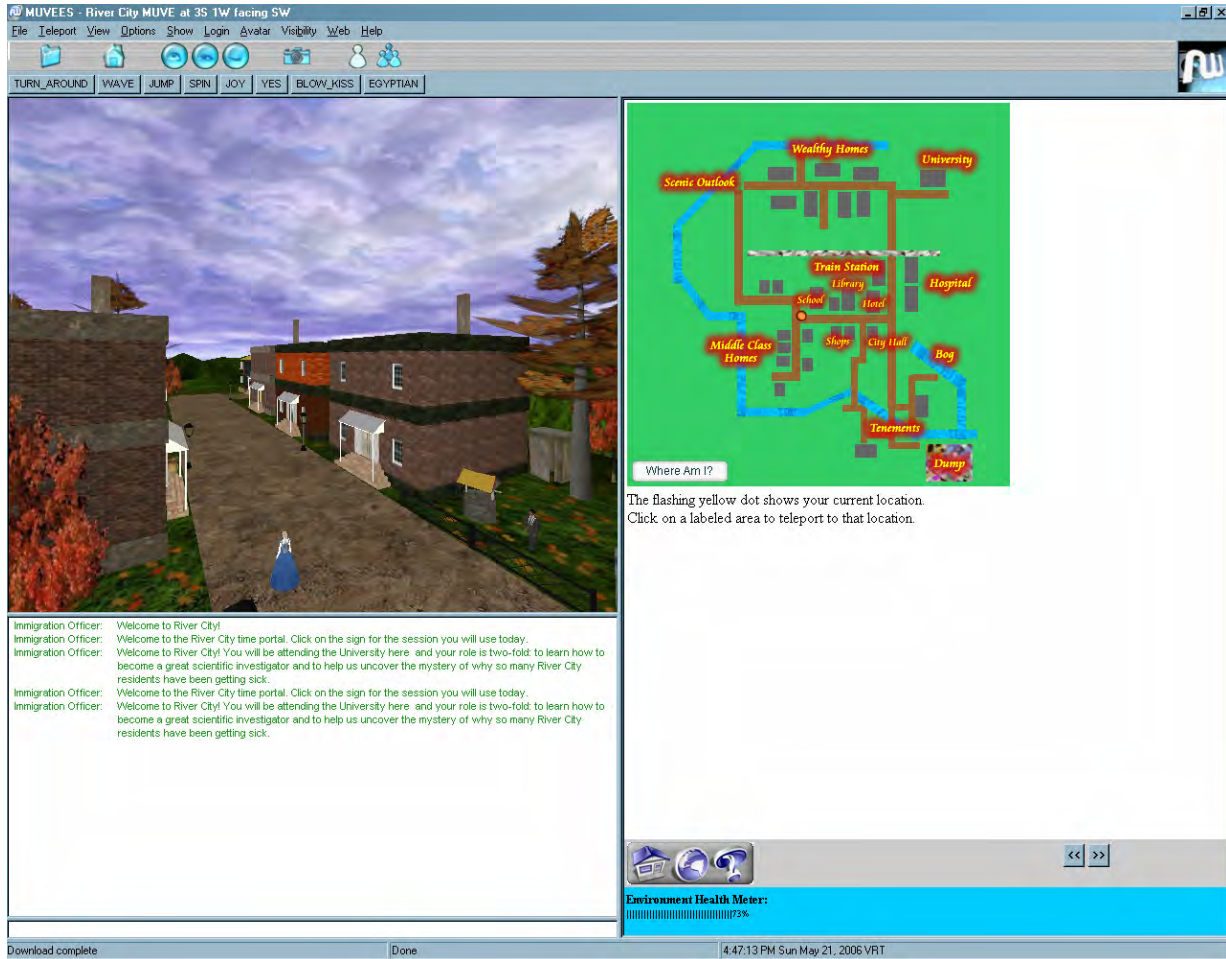
Figure 7: Aerial View of River City. This presents an aerial view of Main Street. The tenements can be seen in the background.
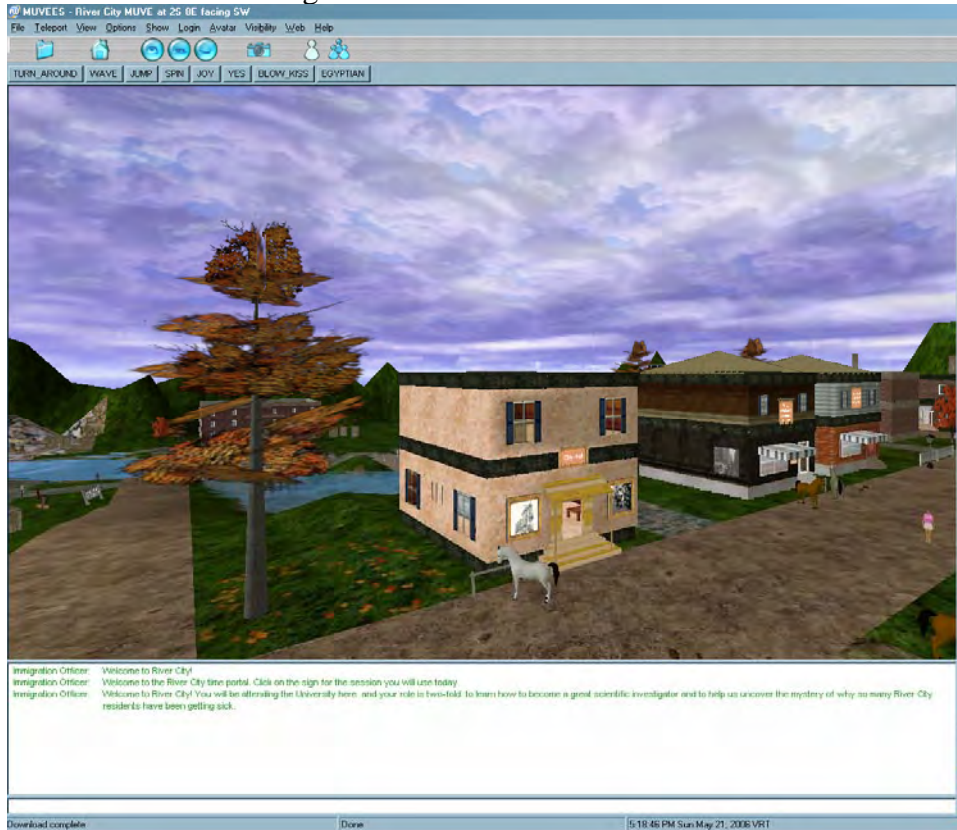
Figure 8: View of River City. This presents another view of the environment.