J Sci Educ Technol (2009) 18:353–365
DOI 10.1007/s10956-009-9156-4

Design for Scalability: A Case Study of the River City Curriculum

Jody Clarke · Chris Dede

Published online: 25 February 2009

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. This paper presents a framework on how to design educational innovations for scalability through enhancing their adaptability for effective usage in a wide variety of settings. The River City multi-user virtual environment (MUVE), a technology-based curriculum designed to enhance engagement and learning in middle school science, is presented as a case study. To date over 250 teachers and 15,000 students throughout the United States and Canada have participated in the River City curriculum. Designers creating and evolving interventions can use this scaling framework to help them increase effectiveness, sustainability, and spread.

Keywords Virtual environments · Science inquiry · Scale

Introduction

Learning is a complicated phenomenon. Numerous, often contradictory, theoretical approaches attempt to explain how and when learning happens, and many related pedagogical theories try to explicate the best way to facilitate the instructional process (National Research Council 2000). The diversity of settings in which learning occurs adds a further dimension of complexity. Yet current instructional and assessment practices instituted under the No Child Left Behind (NCLB) policy initiative tend to treat learning as if it were akin to designing a fast food restaurant, with a very limited menu of pedagogical alternatives. Part of the reason NCLB outcomes are disappointing is that this model of educational reform ignores fundamental principles of educational effectiveness and scalability by imposing policies and accountability measures whose result is a one-size-fits-all model for teaching and learning.

A limited range of options is not effective when scaling up learning and teaching. Research has documented that in education, unlike other sectors of society, the scaling of successful instructional programs from a few settings to widespread use across a range of contexts is very difficult (Dede et al. 2005a, b). In fact, research findings typically show substantial influence of contextual variables (e.g., the teacher's content preparation, students' self-efficacy, prior academic achievement) in shaping the desirability, practicality, and effectiveness of educational interventions. Therefore, achieving scale in education requires designs that can flexibly adapt to effective use in a wide variety of contexts across a spectrum of learners and teachers.

The NCLB experience, with inability to scale effectively, is hardly unique. Numerous studies have documented that it is difficult to scale up promising innovations from the fertile, greenhouse environments in which they were conceived to the often barren contexts that exist in public schools, with few resources, overwhelmed and underpaid teachers, and struggling or disengaged students (Dede et al. 2005a, b). Adapting a locally successful innovation to a wide variety of settings—while maintaining its effectiveness, affordability, and sustainability—is very 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

J. Clarke (✉) · C. Dede
Harvard University, Cambridge, MA, USA
e-mail: Jody_Clarke@mail.harvard.edu;
jec294@mail.harvard.edu

1999). 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, education will continue to waste substantial resources implementing interventions that fail despite promise shown elsewhere.

In this paper, we present a framework for how to design for scale in education, what we term "robust design." As a case study, we offer our research on the River City multi-user virtual environment (MUVE) curriculum, a technology-based innovation designed to enhance engagement and learning in middle school science. To date over 250 teachers and 15,000 students throughout the United States and Canada have participated in the River City curriculum. To achieve this degree of successful scale, our design is adaptable to meet the needs of the teachers, students, and schools that use the curriculum.

In the sections that follow, we first present a conceptual framework for depicting various dimensions important for scalability. Following this, we briefly discuss the complex relationship between technology and scaling in education settings. Then we delineate results of applying our framework to a particular case, the River City curriculum. Finally, we describe conclusions and implications for further research.

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. During "shift," districts, schools, and teachers assume ownership of the innovation, deepening, sustaining, and spreading its impacts. We propose a fifth dimension to extend Coburn'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 (Clarke and Dede 2009):

- Depth: evaluation and research (design-based 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

These dimensions do not describe a linear progression through phases, but instead delineate various types of processes developers can use to help take an innovation to scale. These developmental processes are interrelated in complex ways; for example, sustainability is fostered by spread, and evolution is accelerated by shift.

How does this framework compare with other theoretical/empirical perspectives on the concept of scaling up? Barab and Luehmann (2003), as co-editors of a special issue of Science Education on the topic of scale, discuss issues of sustainability and local adaptation as crucial for scale. They describe the role of the teacher in local adaptation as identifying local needs; critiquing the innovation in light of these needs; visualizing possible scenarios of implementation; and finally making plans or decisions regarding the implementation. The implemented experience of a science curriculum is shaped by its designers, but also by the classroom culture and the teacher's perceptions of the curriculum. Teacher perceptions include issues of accountability, subject matter, pedagogy, and logistics. Other articles in the issue focus on various aspects of teacher adoption (through diffusion), teacher adaptation, and the ways classroom culture tacitly shapes curricular usage. As with the five dimensional framework we use, the challenge for designers is framed as developing curricula that are flexibly adaptive and therefore scalable and sustainable.

An alternative conceptual framework for scale, bolstered by application to a number of NSF-funded projects, is advanced in a two-volume series edited by Schneider and McDonald (2007). In the Introduction to this series, scale-up is defined as "enactment of interventions whose efficacy has already been established in new contexts with the goal of producing similarly positive results in larger, frequently more diverse populations" (vol. 1, p. 4). In vol. 1, the primary methodological focus is not on the design processes described above, but on the theoretical and methodological questions related to standards of evidence for efficacy of an innovation, to determine if scale-up is warranted. Also, work on scale-up in other disciplines (e.g., engineering, public health, economics) is referenced as a resource for understanding operational issues of scale-up in education. In vol. 2, the emphasis shifts to case studies of NSF-funded projects charged with achieving scale. In these cases, stress is placed on how to implement innovations with fidelity and how to measure whether a particular implementation is effective.

All these frameworks are fundamentally consistent in how scaling up is defined in education and the challenges that are intrinsic to dissemination and local adaptation of innovations. The authors of both this article and the articles Barab and Luehmann gathered in their special issue of *Science Education* focus on how to design for scale; most authors in the Schneider and McDonald volumes center their chapters on how to measure efficacy and fidelity in determining whether scaling up is worthwhile and is succeeding. Within this overall conceptual landscape, this article focuses on sustainability as one aspect of the five-dimensional framework for scaling up described above.

Robust-Design for Sustainability

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.

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. For example, while apples are versatile fruit, pomologists need to adapt the design of an orchard, cultivar, and irrigation practices in order to grow apples in climates that are harsher and have shorter seasons. They would not expect these cultivars to yield more fruit than orchards in climates that have evolved for more ideal conditions. 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 be intrusive overhead for better-prepared, already motivated students. As a case example, we utilize design-based research on our River City MUVE curriculum to explore whether robust design strategies can produce the educational equivalent of fruit cultivation tailored to harsh conditions that are productive where the usual version of that fruit would die.

We must caution that 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 one teacher's science class improved by 38% 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.

That said, 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 et al. 2005a, b; Clarke et al. 2006; Nelson 2007; Ketelhut 2007). 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.

The Complex Relationship Between Technology and Scale in Educational Settings

Scaling up technology-based innovations presents particular challenges that are contextual to using technology in schools and education: adequate computers, school computer networks, computer access, and professional development. For example, when designing technology for use in K-12 public schools, the development team must consider that schools tend to have older, slower computers. Therefore, designers need to make sure curricula will run on older machines (within some reasonable limit). For example, in the early days of River City, some schools wanted to use our curriculum with their students, but could not because their computers had insufficient processor power or unsupported video cards lacking enough memory. Comparable challenges with equipment capacity now afflict projects using high-end graphical applications like Second Life or current videogames.

While River City does not require high bandwidth Internet access, an educational application that necessitates substantial high volume downloading (e.g., multiple types of streaming video) can easily overwhelm networking capacity in many schools. Further, many districts have imposed idiosyncratic policies about Internet safety requiring elaborate technical workarounds (e.g., special ports opened) for educational applications, such as River City, that need access to a server outside the district's firewall.

Curricula that require 1-1 computer access, such as the River City MUVE, are difficult to implement for schools in which computer lab time is a scarce resource generally allocated to programming or word processing classes. Further, teachers and students may require substantial support in learning how to use an application involving an unfamiliar or complicated interface. These are all challenges to scale that are particular to technology-based innovations. These challenges frequently require substantial technical support to resolve and, at worst, prohibit students and teachers from participating in that learning experience. Through the case of River City, we illustrate how we have addressed this complex relationship between technology and designing for scale.

River City MUVE as a Case Study

In order to design for scale, designers should develop flexible models with variants adapted to a spectrum of implementation conditions and learning preferences. As an illustration of robust design, we describe our research on the scaling up of River City, an NSF-funded, technology-based curriculum now functioning at scale (http://muve.gse.harvard.edu/rivercityproject/). First we describe the curriculum, and then we illustrate applying our design-for-scale framework.

River City

River City is a technology-based middle school science curriculum designed around national content standards and assessments in biology, ecology, epidemiology, and scientific inquiry (Nelson et al. 2005; Clarke et al. 2006; Nelson 2007; Ketelhut 2007). Specifically, River City focuses on the front end of inquiry: how does one identify a problem, turn the problem into a question, and generate a hypothesis. Students then design and test an experiment around their hypothesis, collect and analyze data, draw a conclusion, and present their results. The National Research Council defines science inquiry as a multifaceted activity that involves students actively making observations, posing questions, planning and conducting experiments, and communicating results (National Research Council 1996). In River City, students engage in all aspects of inquiry as defined by the NRC. The curriculum attempts foster curiosity and help students use inquiry in productive ways (American Association for the Advancement of Science 1993).

Interface Design

The technological infrastructure that delivers the curriculum is a (MUVE). MUVEs are online digital contexts where multiple participants can communicate, navigate, and share experiences (Dede 2009). A participant takes on the identity of an avatar, a virtual persona in the world and communicates with other participants' avatars via text chat and virtual gestures. In this graphical virtual context, participants also interact with computer based agents who are residents of River City (Fig. 1).

When participants click on some artifacts in the virtual world, it triggers content to appear in the right hand interface (Fig. 2). Participants can also use digital tools, such as a virtual microscope (Fig. 3).

Depicting a Nineteenth Century Virtual City

The River City virtual "world" is an industrial nineteenth century city with a river running through it. It is a historically accurate curriculum; pictures from the Smithsonian Institute are embedded within the virtual environment to

Fig. 1 Avatar of a student talking to a River City resident

Fig. 2 The River City interface

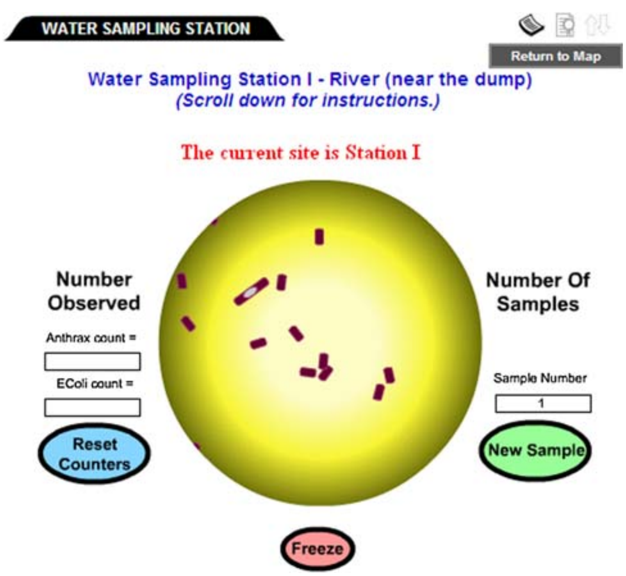

Fig. 3 Looking through the virtual microscope

help portray an accurate picture of what the time period was like. The city has a hospital, hotel, university, main shopping street and different residential neighborhoods. Students walk as their avatars and explore the virtual city observing how different forms of terrain influence water runoff in the various neighborhoods (wealthy area, middle class area and tenements) and interact with computer-based agents, known to them as the “residents of River City” (Fig. 1). Residents include university professors and graduate students who model experimental design for the visiting students, nurses, hotel workers, school children, and a newspaper reporter who is writing about the illness. Also embedded in the design are pictures, signs, charts, and audio clips (e.g., auditory clues of sick residents coughing and mosquitoes buzzing).

Gathering Data on Change Over Time in River City

The storyline of the curriculum is that students have been commissioned by the mayor of River City to travel back in time to 1878 and help her figure out why the residents of the town have fallen ill. Three different diseases, based on multiple causal factors, are simultaneously present in River City; this enables multiple pathways of inquiry. Students work in teams of three and visit River City over the course of a “year” in order to gather longitudinal data on illnesses; because the environment is a simulation, this span of time is accomplished in a few class periods. Students spend at least four consecutive class periods visiting the city during different seasons (October 1878; January 1879; April 1879; July 1879). The teams of students then spend two class periods interacting face-to-face to design their experiment; developing their hypothesis, identifying independent and dependent variables, deciding which aspect of River City to alter in order to reduce disease, and writing up a data collection procedure. Then they test their hypothesis by making a change to the River City environment. The students collect and analyze data to determine how their change influenced the spread of an illness. Students then write up their findings, draw conclusions, and make recommendations to River City’s mayor about how to stop an illness from spreading.

Using Robust Design to Develop River City

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. In this section, we describe how we have been applying this framework for scalability to designing the River City MUVE curriculum.

Depth

"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. In order to understand depth, researchers evaluate and research their design in classroom settings to test the strengths and limitations of its effectiveness. The River City research team employs design-based research methods in order to understand what conditions are more flexible and adaptable to meet needs of students and teachers in various conditions (Clarke et al. 2006; Nelson 2007; Ketelhut 2007).

Design-based research is an emerging research method whose definition, practices, and methods are still being defined by the scholarly community in order to distinguish it from other methodologies (Design-Based Research Collective 2003; Barab and Squire 2004; Collins et al. 2004; Sandoval and Bell 2004; Wang and Hannafin 2005). In this paper, we use the definition provided by the Design Based Research Collaborative (2003):

> Design-based research is an emerging paradigm for the study of learning in context through the systematic design and study of instructional strategies and tools… design-based research can help create and extend knowledge about developing, enacting, and sustaining innovative learning environments.

Design-based research is an iterative process where we engage in design, implement it in classroom settings, research the learning in context, refine our theories of learning, engage in re-design and continue the cycle of implementation. During this process, we employ both qualitative and quantitative research methods. We conduct rigorous classroom observations, interview students, and examine student work and social interactions. We administer surveys and analyze learning outcomes. Findings from these studies then inform the other dimensions of our framework.

For example, in our early implementations we conducted quasi-experimental designs where students were assigned to treatment at the classroom level. Teachers either taught River City or a paper-based control that had similar pedagogy and content but was delivered via paper and hands-on experimenting. We designed different variants of River City that offered different pedagogical approaches to learning: guided social constructivist, expert mentoring, and learning as peripheral participation. In these studies we found that students in the guided social constructivist experimental group (GSC) achieved 16% higher scores on the posttest in biology than students in the control group. These studies also showed that students' thoughtfulness of inquiry, a measure of cognitive engagement, was higher for students who participated in River City than students who participated in the control curriculum (Clarke et al. 2006). We did not find a significant difference between the different learning variants and developed a hybrid variant that contained aspects of all three pedagogical approaches (Clarke et al. 2006).

As part of these early studies, Nelson (2007) developed a guidance system and studied the role it played in students' learning in River City. He found a strong positive link with learning outcomes for students who accessed the guidance system (Nelson 2007). 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, he found an interaction between gender and guidance use. Girls using the guidance system outperformed boys, on average, at each level of guidance message viewing (Nelson 2007). These studies led us to integrate the hint system into the River City application for all students.

We also found an interesting relationship between self-efficacy and students' learning in River City (Ketelhut 2007; Ketelhut et al. 2008a, b). For example, we found that students who entered the project with low levels of self-efficacy did, on average, significantly better with River City than students who participated in the control curriculum. However, we found that students who started the project with higher self-efficacy did better in the control curriculum. These findings emphasize that one size does not fit all and that there is a need for multiple entry points and pedagogical approaches. In order to aid the learning of students who enter the project with high levels of self-efficacy, we decided to explore a feature from game design we call Powers, which is described below under sustainability.

The main purpose of these studies was to compare learning outcomes between the MUVE-based curriculum and a paper-based control (see Ketelhut 2007; Ketelhut et al. 2008a, b) and to see what contextual variables (gender, SES, affect) play a role in students learning and therefore become conditions for success (see Clarke et al. 2006; Nelson 2007; Ketelhut 2007; Clarke and Dede 2007; Ketelhut et al. 2008a, b). Results of these studies influenced the design of future iterations of our MUVE-based curriculum that were robust enough to scale into hundreds of classrooms.

Sustainability

"Sustainability" involves maintaining over substantial periods of time the consequential changes in classroom practice enabled by an innovation's depth. In order to achieve sustainability, researchers must engage in robust-design, or "hybrid" versions that are optimized for success in various settings and for different types of learners.

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. With any given pedagogical approach, some students will do well and some will do poorly. Providing an umbrella that offers different kinds of pedagogies is a rugged design strategy. While we have designed our project to engage all students, we especially 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 different strategies to meet these needs and describe them briefly below.

Unlockable Trajectories

As mentioned above, we found that students who started the curriculum with high levels of self-efficacy lost interest in the curriculum and were outperformed by students in the control curriculum (Ketelhut et al. 2008a, b). In order to reach these students we created a variation we called Powers, which allows students to unlock hidden content in the storyline. Similar to features of videogame play that reward experiences and accomplishments by giving participants special powers, we designed a system of powers that reward student learning and exploration in the River City environment. These powers provide students with access to further curricular information that help them further understand the history 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 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 invokes the attainment of powers.

For example, we want students to explore the different areas of River City and gather information about how the three 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 nineteenth century hospitals. Therefore, as a requirement for one step towards achieving powers for this given world, we check, in real time, whether at least one team member has visited the hospital and interacted with a resident, object, or picture.

Just as 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 earn access to 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 nineteenth century to the twenty-first. 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 so that now we can 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 nineteenth 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 a CAT scan of their lungs. When the student highlights a name, if that character has tuberculosis (one of the three 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 meant to be engaging for students, while still providing rich content that furthers their knowledge. Powers illustrates an adaptation of the curriculum to aid students who may get bored with the curriculum or have difficulty maintaining engagement with the curriculum (Ketelhut et al. 2008a, b).

Formative Assessment

We developed formative assessments embedded in the curriculum. The purpose was to see if we could assess students learning throughout the curriculum in a format that was not "test like." We piloted these assessments and then conducted one-on-one interviews with students where we went through the questions and received feedback. Our pilot of these questions revealed that some of the question formats were confusing and that they felt out of context. The students did not understand how the questions connected to the story line. Further, the data was not stored in a format that was easily accessible for both researchers and teachers. We also solicited feedback from teachers.

As a result, we redesigned the formative assessments to be integrated into the narrative of River City. In these assessments, a newspaper reporter interacts with students during each visit, sometimes more than once, to gather information for a newspaper article the reporter is writing about the illness in River City. This second design contains a series of multiple-choice questions and open-response items. The purpose of these new assessments is not is to only assess student learning during the curriculum, but also to provide formative feedback to teachers about student learning that they can use to inform their instruction. This will be discussed below under "Spread."

Spread

"Spread" involves the diffusion of the innovation to large numbers of classrooms and schools. In order to achieve spread, researchers should modify the design to retain effectiveness while reducing resources and expertise required for success. As we scaled our project, we knew that the best way to be efficient was to automate as many processes as possible. We did this through creating an online dashboard for teachers and automated reports.

The River City Dashboard

We have found that integrating the River City project, a technology-intensive project, pushes the limits of teachers' comfort level. In early implementations, some teachers had said 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 without using a lot of resources. Our result is a "Teacher Dashboard" that provides teachers with all the tools and mechanics necessary to successfully implement the River City project with little help from the research team. This dashboard houses numerous resources and functions under one location (web page), so that teachers only need to create a single bookmark in their internet browser (the site is password protected).

Through the "teacher dashboard," teachers can 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 look for various conditions for success. Being able to determine the conditions for success relies heavily on the collection of appropriate data. In some implementations, missing data on student prior academic achievement, reading scores, and demographics led to a 25% reduction in sample size. 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 three. In the past, we created student accounts and randomly assigned student teams. Now, these two steps both provide teachers with more control over using the project with their students and make the project easier to scale, as the teachers rely less on the research team.

Automated Email Reports

In the past, teachers who participated in our curriculum wanted to know more about what a particular student was doing in the curriculum on a given day. At the time, students interacted with the MUVE and recorded their individual work in a paper-based laboratory notebook, which they handed in to their teacher. Teachers had access to student work only via the notebooks and had little detailed knowledge about what each student was doing in the MUVE itself. Thus, teachers could keep up with student progress and learning throughout the curriculum only via written work in the notebooks.

The backend architecture of the MUVE environment is a database. Everything that learners do within the River City environment is recorded in an "event log" that lists the time and location in the virtual world, and the activity being engaged in. For example, students working in scientific teams, inside the technology, can only communicate with each other via a text-based "chat" system; the software automatically captures these chat data as they are generated. The event logs indicate where students were, with whom they communicated, what was said in these interactions, what virtual artifacts were activated, and what responses were provoked. The research team examined and analyzed data from the event logs in order to understand students' inquiry learning processes, but we had not found a way to provide this information to teachers in a timely manner that could inform instruction.

Similarly, we wanted to know more about individual student activity in a timely manner. Teachers mailed student notebooks to the research team at the end of the project. Due to the time delay and sheer volume of paper, sub-samples of these data were analyzed to inform future design, but not for formatively influencing students' learning processes. As mentioned above, embedded assessments were also developed to measure student learning unobtrusively in the curriculum. However, the first iteration of this data was not stored in a manner that made it easy to retrieve. Thus, formative assessment data was analyzed only after all implementations were complete and never shared with teachers.

As we scaled, we wanted to automate processes and find ways to put ownership of student learning in hands of teachers. We also needed to make sure we could study students learning in our curriculum. We moved beyond these suboptimal strategies by developing automated reports of student activity in the world. The student notebook was redesigned such that students had to submit all answers digitally via an online notebook. This redesign was an iterative process described below. These reports allows researchers to study metacognitive and reflective learning of students while they interact inside the MUVE. In addition, student embedded assessments were redesigned such that the data was accessible and usable by both teachers and researchers.

To make data usable for teachers, each night, three daily reports of student activity in the environment are generated and emailed to teachers: (1) notebook entries, (2) team synchronous chat transcripts, and (3) embedded formative assessments. For example, early Tuesday morning, Teacher X receives three email reports. The first email contains a summary of their students' notebooks from Monday. The second email contains transcripts of all team chat from Monday. The third email contains their students' embedded assessments from Monday. These reports allow teachers to monitor their students' progress (whether they are actually on task) and language (whether or not they are using bad language). The design of these reports has been an iterative process based on feedback from teachers. Despite the large amount of information presented in these reports, we did find that teachers used them. For example, in one study, we found that 86% percent of the teachers in the sample ($n = 73$) liked receiving both the reports of students' chat and individual work, and 81% liked receiving the daily assessments (Clarke 2008).

Another system of automated reports was set up for researchers to track different aspects of implementations, such as teacher activity in the dashboard (creating accounts), and student activity in the world. A detailed tracking system was also built to monitor individual teachers' progress through the different stages of the project. Teachers can also monitor their progress via the Teacher Dashboard. In the past, it was difficult for us as researchers to keep up with hundreds of teachers' progress through the curriculum and make sure they were fulfilling our research requirements. Now there is an automated checklist for each teacher that is viewable by both the research team and the individual teacher.

Additional Resources

In addition to these modifications in the architecture of the design, we have added resources such as "Day-by-Day" lesson planning for teachers to use as resources and "quick guides." We have also added short videos (less than a minute each) 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.

Shift

During "shift," districts, schools, and teachers assume ownership of the innovation, deepening, sustaining, and spreading its impacts. In order to achieve shift, researchers need to move beyond "brand" to support users as co-evaluators, co-designers, and co-scalers. We do not provide teachers with a scripted curriculum, but rather encourage them to adapt it to their students' needs.

At various points over the past 8 years, our focus has leaned in a particular direction (design, practice, or research); however, most often our work leans towards practice. We have worked hard to establish relationships with schools and teachers. Without strong relationships with teachers we would not be able to achieve shift. Our teachers have been co-evaluators and co-designers at every stage of implementation. We discuss some brief examples below.

Teacher-to-Teacher Curricular Help

When recruiting new teachers, they often have questions about implementing our curriculum with the students in their classrooms. When this happens, we have teachers who have implemented River City respond to questions. For example, a recent district was concerned about how to adapt our curriculum for students with special needs. Thus, we asked a teacher who had implemented River City with special needs students to share insights and information about their experience. Our teachers are experts with our curriculum, and we rely on them when questions arise about how to best implement it.

Teachers as Co-Evaluators

We have always taken teacher feedback seriously. Throughout our iterative design process, we integrate teacher feedback into our re-design. However, we also treat teachers as co-evaluators and ask them to help us pilot different features and provide extensive feedback. For example, we mentioned briefly above that we modified our student lab notebook. Teachers played a large role in the evaluation of this process. Historically, students interacted with the River City MUVE and recorded their individual work in a paper-based laboratory notebook, which they handed in to their teacher. This paper notebook served as a guide, directing students through the curriculum. However, numerous teachers complained about the amount of paper required to print these notebooks for their students. Some teachers were using their copy funds for a whole year on our project.

Interviews, observations, and discussions with teachers led us to the decision to integrate this notebook online. The design of the online notebook was an iterative process. We piloted versions of our new online design and asked River City teachers to be co-evaluators of this process. Our first phase involved reducing the number of pages by having students work from a paper guide and submit any responses to questions or activity in an online notepad. This made teachers happier, because they could reuse these notebooks and they required less paper. We worked closely with a group of teachers in the Midwest. Despite their busy schedules, they participated in telephone focus groups and provided written feedback. Teachers have also served as co-evaluators of our email reports. These relationships are invaluable to us and have led to the ownership and scale of the River City Curriculum in classrooms across the country.

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. In order to achieve scale along this dimension, researchers need to learn from users' adaptations to rethink the innovations' design model. Evolution is more than providing teachers with ownership; it is incorporating their ownership into the evolution of the curriculum. Evolution is really a product of depth, spread, and shift. Without these dimensions, we would have difficulty with evolution. We describe an example of evolution below.

Evolution of Professional Development

Like any instructional program, River City is only successful if teachers are comfortable teaching with it. Over the years our professional development and teachers' ownership of River City has evolved in such a way that it changed how we think about and deliver professional development.

During the early days when River City was being used with a few teachers we offered face-to-face and one-to-one training. However, when we started implementing with a small group of teachers in two different states, one of them remote, we were not able to deliver face-to-face professional development to all the teachers. At the time, one of the districts with which we were working was using a web-based portal. They allowed us to use the portal with all of our teachers to host our professional development materials. However, we quickly found that teachers were not accessing the materials, nor were they taking ownership of the program.

As we prepared to scale across numerous districts and states we knew we had to find a way to deliver professional development that was scalable. Our goal is to work with large districts, so we adopted a train-the-trainer approach that involved hiring local trainers to train and supervise teachers in a district.

The role of trainers was to provide professional development and ongoing support to teachers. The trainer would also serve as the "eyes and ears" of the research team in classrooms. The research team developed training materials and trained the trainers. The trainers were given the materials and told to modify them based on their own needs.

Initially, we gave the trainers PowerPoint slides and handouts that contained information about the curriculum (e.g., a cheat sheet on the diseases and causal factors). However, we also had small pockets of individual teachers who wanted to use River City with their students. We had not anticipated having a large number of individual teachers spread out across the country and could not afford to hire a trainer to work with each teacher. So, we developed separate training materials for these teachers and referred to this strategy as Individualized Professional Development (IPD). The information was housed in one electronic document and could be placed in a binder.

Originally, we treated training differently for these teachers. Teachers who worked with trainers had face-to-face instruction and ongoing support from a trainer, while IPD teachers instead received a training manual and received ongoing support from the research team. As in previous implementations, we found that sending materials to teachers was not a good method of professional development and did not lead to ownership. However, our teachers who worked with trainers were thriving.

Teachers and trainers provided feedback on documents and asked for more materials. They wanted us to develop materials that were more reflective and interactive—like our curriculum. While we put emphasis on interactive materials for students, we did not develop interactive materials for teachers. We relied on paper materials and PowerPoint presentations for professional development. However, trainers wanted us to utilize more technology and provide "just in time" training modules that teachers could watch for last-minute reminders. These conversations and meetings were critical in helping us rethink how we deliver professional development.

We reconceptualized our professional development for individual teachers and new trainers. We obtained a license to elluminate, online training software for conducting "webinars." We also modified our training for online delivery. In addition, we repurposed our IPD manual into a "just in time training" resource that has links to video training modules. These short videos provide quick demonstrations for how to create students' accounts for the simulation, set up teams of students, etc. However, rather than treat training materials differently, this "just in time training" is available to all teachers using the curriculum.

We now hold monthly online training sessions for groups of individual teachers who wish to implement the River City curriculum. The elluminate webinars have also proven valuable in updating teachers who have previously used River City and need to know about changes we have made since they were trained. We also hold train-the-trainer sessions for new and returning trainers.

These changes evolved through the practices of teachers and trainers. Without shift, they may never have been communicated to the research team.

The River City Recruitment Model

In terms both of need and scale, large urban districts are an attractive target of opportunity for any research project that hopes to reach substantial numbers of teachers and students. However, like many other groups attempting to foster innovation in urban schools, we have found that partnerships with these districts are complex and contextual. Further, under the shadow of NCLB it is challenging to find public schools willing to dedicate instructional time to an innovative project like River City. However, despite these challenges, we continue to scale our project and work with teachers across the US and Canada. Our most successful form of recruitment has been word of mouth from current users, coupled with proactive outreach through conference talks and district workshops that have a broader span of material than just River City.

When we started the River City project, our primary method of recruitment was based on prior relationships with districts through the principal investigator of the project. To this day we continue to work with a group of these districts in the Midwest. In districts where we have support from a superintendent or an administrator, we rely on district-level advertising, which varies depending on the district. Some districts prefer to advertise on websites, but one asked us to produce a flyer to be distributed to teachers via mailboxes. In the district where we have our largest number of implementations, word of mouth is a means of recruitment.

We found that cold-calling districts does not work. Without support from a high level champion such as a superintendant, it is impossible to make connections solid enough to lead to implementations at scale. Cold-calling was time-consuming and, even if successful, resulted in only a small number of classroom implementations.

However, support from a superintendent does not automatically mean easy access to schools. In one district, we had the support of the superintendent and the technology team. Some schools in the district used early iterations of River City and saw learning gains, and more teachers began asking to use River City with their students. However, the district science coordinator was not interested in having River City used in the schools. Even a direct intervention from the superintendent was unable to change that person's resistance, which blocked us from implementing at scale.

A majority of the teachers we work with find us. For example, two large organizations contacted us directly. In one case, the head of technology for an urban district had heard about River City from a technologist in a district that was using River City. However, the head of technology was our only internal champion in the district, and when that person resigned we lost the support required to involve the other parts of the district. To this day we have been unable to rekindle interest. A second group that contacted us was a regional organization of smaller districts in a large state in the Northeast United States. These districts have internal trainers who have taken ownership of

River City and incorporated it into their middle school curricula.

The last large district we recruited was the result of a conference keynote. One of the River City researchers saw a presentation by a chief learning officer in the district and approached them. Through the project's principal investigator, we were able to establish a connection. Yet, due to district bureaucracy it has taken almost 2 years to implement River City at scale in the district.

Outside of district-level initiatives, the individual teachers with whom we work all contacted the project directly via email. These teachers had learned about the project via web searches, articles, and presentations by members of the research team. In a couple of cases, we were approached by technology coordinators who discovered our project and wanted to use it with a number of teachers. Over all, though, our email contacts tend to be isolated teachers across the country.

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 iterative process that involves teachers as co-evaluators and co-designers.

Overall, the robust-design approach has intrinsic limits, as some essential conditions that affect the success of an educational innovation cannot be remediated through ruggedizing. 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.

Still, researchers can potentially get some leverage on these essential factors. For example, 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 et al. 2005a, b; Clarke et al. 2006; Nelson 2007; Ketelhut 2007). 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.

Acknowledgments The authors wish to acknowledge and thank the rest of the River City team: Diane Jass Ketelhut, Brian Nelson, Catherine Bowman, Edward Dieterle, Geordie Dukas, Eugenia Garduno, and Janet Smith for their collaboration throughout River City's evolution.

References

American Association for the Advancement of Science (1993) Benchmarks for science literacy. Oxford University Press, New York

Barab SA, Luehmann AL (2003) Building sustainable science curriculum: acknowledging and accommodating local adaptation. Sci Educ 87:454–467. doi:10.1002/sce.10083

Barab SA, Squire K (eds) (2004) Design-based research. J Learn Sci 13(1). Special issue doi:10.1207/s15327809jls1301_1

Clarke J (2008). Data as a diagnostic tool. Proceedings of the international conference of the learning sciences, Utrecht Netherlands

Clarke J, Dede C (2009) Visions of the future: learning and instructional technologies for the 21st century. In: Moller L (ed) Robust designs for scalability. Springer, New York, pp 27–48

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

Coburn CE (2003) Rethinking scale: moving beyond numbers to deep and lasting change. Educ Res 32:3–12. doi:10.3102/0013189X032006003

Collins A, Joseph D, Bielaczyc K (2004) Design research: theoretical and methodological issues. J Learn Sci 13(1):15–42. doi:10.1207/s15327809jls1301_2

Dede C (2006) Scaling up: evolving innovations beyond ideal settings to challenging contexts of practice. In: Sawyer RK (ed) Cambridge handbook of the learning sciences. Cambridge University Press, Cambridge, pp 551–566

Dede C (2009) Immersive interfaces for engagement and learning. Science 323(5910):66–69

Dede C, Honan J, Peters L (eds) (2005) Scaling up success: lessons learned from technology-based educational innovation. Jossey-Bass, New York

Dede C, Clarke J, Ketelhut DJ, 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, Canada

Design-Based Research Collective (2003) Design-based research: an emerging paradigm for educational inquiry. Educ Res 32(1):5–8. doi:10.3102/0013189X032001005

Ketelhut DJ (2007) The impact of student self-efficacy on scientific inquiry skills: an exploratory investigation in River City, a multi-user virtual environment. J Sci Educ Technol 16(1):99–111. doi:10.1007/s10956-006-9038-y

Ketelhut DJ, Clarke J, Nelson B, Dukas G (2008a) Using multi-user virtual environments to simulate authentic scientific practice and enhance student learning. In: Annetta LA (ed) Serious educational games: from theory to practice. Sense Publishers, Rotterdam, Netherlands

Ketelhut DJ, Clarke J, Nelson B, Dukas G (2008b) Rethinking pedagogy: using multi-user virtual environments to foster authentic science learning. Proceedings of the international conference of the learning sciences, Utrecht Netherlands

Moore GA (1999) Crossing the chasm: marketing and selling high-tech products to mainstream customers. HarperBusiness, New York

National Research Council (1996) National science education standards: observe, interact, change, learn. National Academy Press, Washington, DC

National Research Council (2000) How people learn: brain, mind, experience, and school. National Academy Press, Washington, DC

Nelson B (2007) Exploring the use of individualized, reflective guidance in an educational multi-user virtual environment. J Sci Educ Technol 16(1):83–97. doi:10.1007/s10956-006-9039-x

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. Educ Technol 45(1):21–27

Sandoval WA, Bell PL (2004) Design-based research methods for studying learning in context: introduction. Educ Psychol 39(4):199–201.

Schneider B, McDonald S-K (eds) (2007) Scale-up in education: ideas in principle (vol 1) and scale-up in education: issues in practice (vol 2). Rowman & Littlefield, New York

Wang F, Hannafin MJ (2005) Design-based research and technology-enhanced learning environments. Educ Technol Res Dev 53(4): 5–23. doi:10.1007/BF02504682

Wiske MS, Perkins D (2005) Dewey goes digital: scaling up constructivist pedagogies and the promise of new technologies. In: Dede C, Honan J, Peters L (eds) Scaling up success: lessons learned from technology-based educational innovation. Jossey-Bass, San Francisco