JOURNAL OF GEOSCIENCE EDUCATION **64**, 60–73 (2016)

Advantages of Computer Simulation in Enhancing Students' Learning About Landform Evolution: A Case Study Using the Grand Canyon

Wei Luo,[1,a] Jon Pelletier,[2] Kirk Duffin,[3] Carol Ormand,[4] Wei-chen Hung,[5] David J. Shernoff,[6] Xiaoming Zhai,[7] Ellen Iverson,[4] Kyle Whalley,[1] Courtney Gallaher,[1] and Walter Furness[1]

ABSTRACT

The long geological time needed for landform development and evolution poses a challenge for understanding and appreciating the processes involved. The Web-based Interactive Landform Simulation Model—Grand Canyon (WILSIM-GC, http://serc.carleton.edu/landform/) is an educational tool designed to help students better understand such processes, using the Grand Canyon as an example. Although the project is still ongoing, here, we present the initial results of using the WILSIM-GC in an introductory physical geography laboratory course. We used a quasiexperimental design to assess the efficacy of WILSIM-GC as a tool for teaching landform development and evolution. Students were assigned to a control group or a treatment group, alphabetically by last name. Pretests and posttests were administered to measure students' understanding of the concepts and processes related to Grand Canyon formation and evolution. Results show that, although both the interactive simulation and a more-traditional, paper-based exercise were effective in helping students learn landform evolution processes, there were several advantages and affordances to the simulation approach. The improvement effect from pretest to posttest scores was large for the treatment group, but small to moderate for the control group. In addition, for those questions requiring higher-level thinking, the percentage of students answering correctly was higher in the treatment group than it was in the control group. Furthermore, responses to the attitudinal survey indicate that students generally favor the interactive simulation approach. We can leverage these advantages to enhance students' learning by integrating interactive simulation exercises into curricular materials, including materials for online or hybrid courses and flipped classrooms. © *2016 National Association of Geoscience Teachers*. [DOI: 10.5408/15-080.1]

Key words: simulation model, landform evolution, learning, treatment/control experiment, Grand Canyon

INTRODUCTION

Background

Computer simulations are computer-generated, dynamic models of the real world and its processes and often represent theoretical or simplified versions of real-world components, phenomena, or processes (Smetana and Bell, 2012). As such, computer simulations offer an environment for students to explore the phenomena of the real world and better understand the science behind the phenomena. A large body of literature exists on computer simulations in science education. Smetana and Bell (2012) recently provided a comprehensive and critical review.

Ideally, computer simulations are flexible, dynamic, and interactive and thus encourage inquiry-based exploration, in which students draw their own conclusions about scientific concepts and ideas by altering values of different variables and observing their effect (Windschitl and Andre, 1998; de Jong, 2006; Perkins et al., 2012). Most researchers agree that the interactivity of computer simulation and its ability to engage students are the keys to maximizing its advantages in improving student learning (e.g., Tversky et al., 2002; Day, 2012). Interactive computer simulations give students a sense of control and ownership of their exploration and discovery, and thus, it enhances their understanding and retention of information (Podolefsky et al., 2013). These simulations offer the opportunity to re-create and visualize processes/phenomena of the real world that would take too long (e.g., geologic processes) or might be too dangerous or too complicated for a conventional classroom/laboratory setting (Akpan, 2001). Simulations also allow students to focus on the essential aspects of a process or system while eliminating extraneous variables, promoting understanding of the causal relationships between events or variables (de Jong and van Joolingen, 1998). The learning-by-doing approach can also make abstract concepts more concrete (Ramasundarm et al., 2005). The interactive engagement and immediate feedback of simulations allow students to work at their own pace and easily repeat trials and thus promote conceptual reasoning and deeper understanding (Smetana and Bell, 2012). The use of computer-based technology in classrooms is now well established, especially simulation tools that are freely available over the Internet, such as the

---

Received 18 July 2014; revised 21 May 2015 and 12 December 2015; accepted 14 December 2015; published online 15 February 2016.

[1] Department of Geography, Northern Illinois University, Davis Hall, Room 118, DeKalb, Illinois 60115, USA

[2] Department of Geosciences, University of Arizona, 1040 E. Fourth Street, Tucson, Arizona 85721, USA

[3] Department of Computer Science, Northern Illinois University, DeKalb, Illinois 60115, USA

[4] Science Education Resource Center, Carleton College, W-SERC, One North College Street, Northfield, Minnesota 55057, USA

[5] Department of Educational Technology, Research and Assessment (ETRA), Northern Illinois University, DeKalb, Illinois 60115, USA

[6] Graduate School of Applied and Professional Psychology, Rutgers University, 152 Frelinghuysen Road, Piscataway, New Jersey 08854-8020, USA

[7] Engineering, Math and Physical Science Division, College of Lake County, 19351 West Washington Street, Grayslake, Illinois 60030-1198, USA

[a] Author to whom correspondence should be addressed. Electronic mail: wluo@niu.edu

PhET collection developed by the University of Colorado (https://phet.colorado.edu/).

However, some previous studies also reported mixed or inconclusive results on the effect of simulations on enhancing students' learning (e.g., Anglin et al., 2004; Edsall and Wentz, 2007; Randy and Trundle, 2008; Scalise et al., 2011). Researchers found that traditional methods are just as effective and computer simulations alone are inadequate in helping students understand more-complex ideas because these more-complex simulations often require higher interactivity, which can potentially overwhelm students (Adams et al., 2008; Podolefsky et al., 2010a). Therefore, scaffolding is necessary to help students develop enough background knowledge so that they are not overwhelmed but are adequately equipped and ready to explore the phenomena in question (e.g., Khan, 2011; Schneps et al., 2014). There should be a balance between the level of guidance provided and the flexibility students have to explore on their own because a too strongly guided approach (i.e., step-by-step cookbook) can undermine the potential for exploration, and insufficient guidance can overwhelm students (Adams et al., 2008). Implicit scaffolding features, such as using sliding bars to limit the variable values, restricting the number of variables students can change, and setting default conditions with ideal variable values can keep students from becoming overwhelmed and forestall random interactions that may lead to confusion (Chen, 2010). Smetana and Bell (2012) concluded that "as with any other educational tool, the effectiveness of computer simulations is dependent upon the ways in which they are used" and suggested that "computer simulations are most effective when they (1) are used as supplements; (2) incorporate high-quality support structures; (3) encourage student reflection; and (4) promote cognitive dissonance." Simulations can present rich opportunities for students to experience cognitive dissonance by observing phenomena or collecting data that challenge preconceived conceptions, which then lead to conceptual change (e.g., Bell and Trundle, 2008).

Experiments directly relevant to Earth's evolution can rarely be performed in a completely controlled laboratory setting at the same spatial and temporal scales as experiments in many other sciences, such as chemistry or physics (Baker, 2014). The landforms we see today took millions of years to form, and this makes teaching students about landscape evolution a challenging task. However, landform evolution is an important part of geosciences education because landforms are usually the first natural features we observe when we study the effects of human activities on our environment (Luo et al., 2004). In addition, the ability to infer long-term processes from limited observations is truly a four-dimensional (three-dimensional space + time) skill that is often hard for students to master (e.g., Kastens et al., 2009). Therefore, computer modeling can not only have a particularly important role in testing hypotheses regarding processes in geosciences at various spatial and temporal scales but also offers a virtual laboratory environment for students to learn about these processes through interactive experiments and active exploration.

However, few studies have quantitatively compared the effects of simulations on students' learning in geoscience with other teaching methods. Edsall and Wentz (2007) reported on two experiments that compared students' performance in understanding map projections using a computer-based model versus a physical model and in understanding coastal geomorphology using geographical information system maps versus paper maps. The authors analyzed pretest and posttest scores for the control and treatment groups and concluded that both computer-based methods and traditional methods were effective at improving students' understanding. They found that computer-based approaches were appealing to students but were not, by themselves, significantly more beneficial in enabling understanding of complex concepts. Stumpf et al. (2008) compared pretest and posttest results from students who learned desert geomorphology through a virtual field trip and a traditional (real) field trip and found that the virtual field trip was statistically indistinguishable from the real field trip in establishing basic knowledge about desert geomorphology. However, their qualitative results revealed deeper personal ownership of knowledge among real field-trip participants. On the other hand, the authors also noted the tremendous advantages of the cost effectiveness of the virtual field trip (especially for physically, economically, or politically hard-to-access places) and the unique alternative learning environment it can offer for physically disabled students (Stumpf et al., 2008).

Purpose of Our Study

Given the results discussed above, it is important to conduct additional empirical studies to investigate the effectiveness and benefits of computer simulations for improving students' understanding of long-term geological processes (Edsall and Wentz, 2007). Based on the benefits of computer simulation documented in the literature, and the spatial and temporal scales of Earth Science in general and landform evolution in particular, we hypothesized that computer simulation would enhance students understanding of the long-term geological processes more than would traditional-instruction approaches. The objectives of this study were (1) to test our hypothesis quantitatively, using a quasiexperimental design; (2) to elucidate what advantages simulation models have over traditional instruction; and (3) to identify areas for improvement to be addressed in our ongoing and future work. Analyses of both quantitative and qualitative data collected from the study will be used to address objectives (2) and (3).

WEB-BASED INTERACTIVE LANDFORM SIMULATION MODEL—GRAND CANYON

The Web-Based Interactive Landform Simulation Model (WILSIM) was developed to offer an easily accessible and interactive environment for students to engage in explorative scientific inquiry and to enable and enhance students' understanding of the processes involved in landform evolution through meaningful manipulation of parameters for different scenarios (Luo et al., 2004, 2005, 2006). An earlier version adopted a rule-based model, which applied the local rules iteratively, i.e., water flows downhill and erodes the surface in proportion to slope (see Chase, 1992 and Luo et al. 2004 for details). Testing of this earlier version in classrooms showed that students' posttest scores were higher than pretest scores and that the increase was statistically significant (Luo and Konen, 2007). However, students' feedback also indicated an important limitation of the rule-based approach: it was hard to relate time and space measures in the model to those in real-world landforms. For example, time in the model was measured in terms of number of iterations (rather than millions of years) and distance in the model was measured in terms of the number of cells of the topographic grid (rather than kilometers).

To address these limitations and also to take advantage of the developments in Java (Oracle Corporation, Redwood City, CA) technology, we obtained funding from the National Science Foundation—Transforming Undergraduate Education in Science, Technology, Engineering and Mathematics (NSF-TUES) program to develop the next generation WILSIM and to simulate the landform development of the Grand Canyon. We call this new version WILSIM-GC (GC for Grand Canyon). The Grand Canyon was selected because it is arguably the most famous landform that could create heightened student interest in studying its development and help students relate the concepts and processes of geomorphology to the dramatic real-world features preserved in the Grand Canyon. This also fits our goal of primarily targeting entry-level, nonmajor students. However, the model could also be used in higher-level major courses.

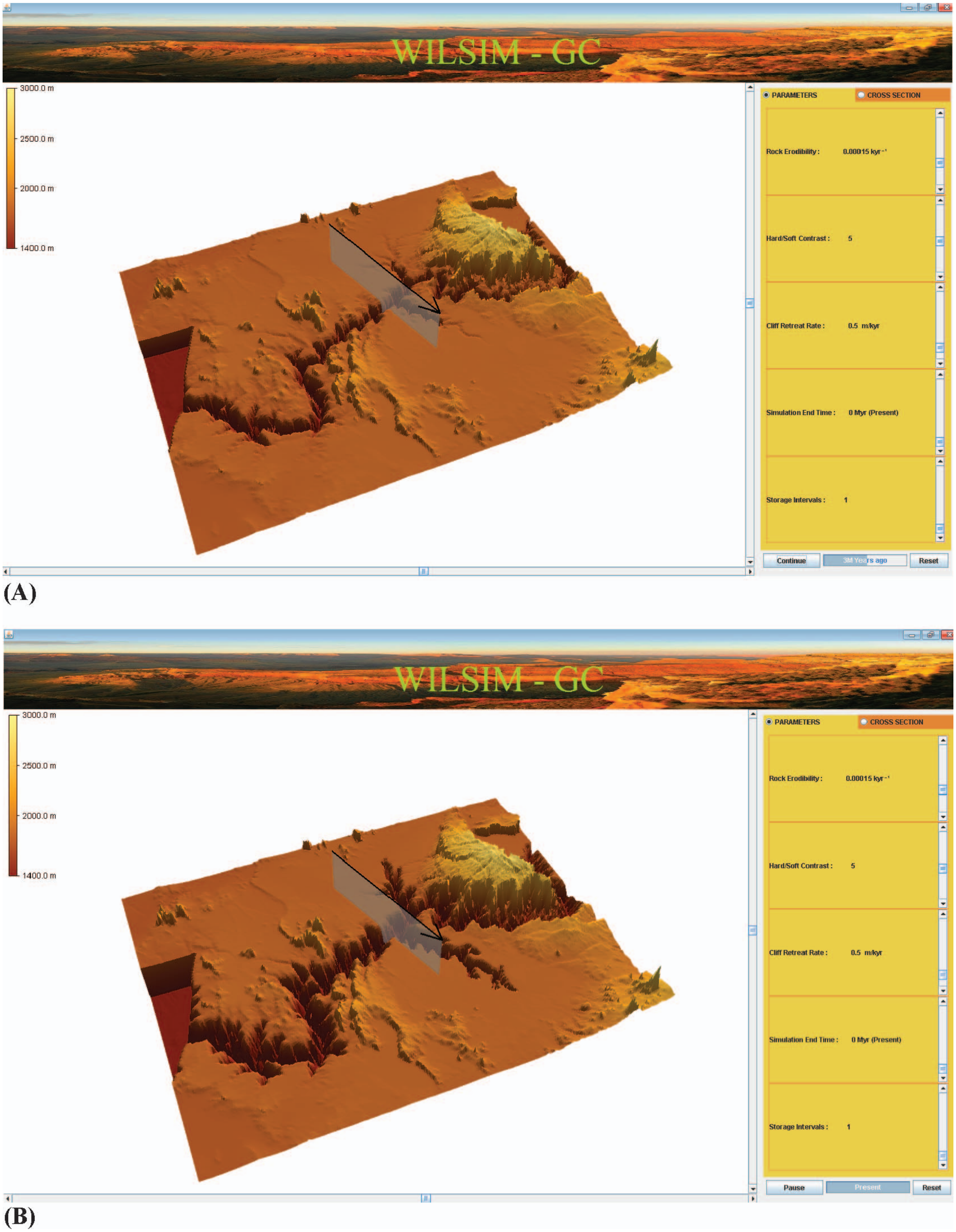
WILSIM-GC was built upon a state-of-the-art, physically based model (Pelletier, 2010) that simulates bedrock channel erosion and cliff-retreat processes responsible for the development and evolution of the Grand Canyon during the past 6 million years. The model takes into consideration the stratigraphy and geologic conditions of the Grand Canyon in a simplified way (Pelletier, 2010). The design of the model followed principles of affordances (providing visual cues such as sliding bars so that students can use to easily manipulate the model and watch the effect of their changes in real time) and constraints (limiting the range of parameter values so that students will not waste time exploring unrealistic values) advocated by Podolefsky et al. (2010b). Incision in the model is driven by the histories of key faults (Grand Wash, Hurricane, and Toroweap) as reconstructed from the geologic record (Pelletier, 2010). Students can change a number of parameters (e.g., rock erodibility, cliff retreat rate, and hard/soft contrast) and observe the effect of their changes on the resulting landform in animation, simulating 6 million years of evolution within a minute or two on an average, modern computer. The rate of bedrock channel erosion is controlled by the "rock erodibility" parameter (which encapsulates the effect of drainage area and channel slope). The weathering and failure of cliffs are included in the "cliff retreat rate" parameter based on geologic constraints. Realistic stratigraphy was included in the model to simulate the alternating soft and hard rock layers in the Grand Canyon, and students can alter the hardness contrast of the stratigraphic layers using the "hard/soft contrast" parameter. In addition, the new model takes advantage of Java OpenGL for access to ubiquitous fast-graphics hardware, a graphics library used in computer games, which is implemented as a trusted applet and allows students to save data for cross-sections and long profiles to a local computer for further analysis. Students can also change the simulation ending time to allow the model to run into the future. WILSIM-GC is a free program and can be accessed from anywhere with a Java-enabled browser, provided there is Internet connection. A screenshot of the model is shown in Fig. 1. The development and improvement of the model is ongoing. More details about the model and the most up-to-date information can be found at the project Web site: http://serc.carleton.edu/landform/.

EXPERIMENTAL DESIGN AND DESCRIPTION OF INTERVENTION ACTIVITIES

We tested WILSIM-GC with college students in GEOG 102: Survey of Physical Geography Laboratory, a one-credit-hour, general education laboratory course (accompanying the three-credit hour GEOG 101: Survey of Physical Geography Lecture), which satisfies the general education science with laboratory requirement at Northern Illinois University. The Internal Review Board human subject research approval ("exempt" status) was obtained before the start of the experiment. The laboratory course has five sections that meet at different times of the week, with a total enrollment of 54 students, but only 43 students' data were included in the final analysis (see Table I and explanation below).

We conducted a quasiexperiment using a nonrandomized control group, pretest–posttest design (Ary et al., 2010, p. 316) during a week in October 2014, which was one of the weekly laboratories contributing to students' grade in a manner similar to other laboratory classes. The experimental procedure is illustrated in Fig. 2. We divided students in each laboratory section into two groups of equal size by alphabetic order of their last names (first half in group A and second half in group B). Before the laboratory, all students were required to read some background material about the Grand Canyon posted on the course online management system BlackBoard (Blackboard Inc., Washington, DC), set up the proper Java security settings for WILSIM-GC to run on their laptops, and complete the pretest assignment. This was done to familiarize students with background information on the Grand Canyon and also to allow more time for the learning activities during the laboratory. Each laboratory session lasted 1 h and 50 min. During the laboratory, the two groups used different curricular materials to learn about the processes involved in forming the Grand Canyon: the treatment group (group A) used WILSIM-GC and the control group (group B) used traditional paper-based, written materials (the details of each set of teaching materials will be described next). Both groups then completed a posttest immediately after their respective learning activities (Fig. 2). The pretest and posttest were designed to measure students' understanding of basic landform concepts and the processes involved in the formation of the Grand Canyon and their ability to apply what they learned to a different scenario. An earlier version of the tests, consisting of 16 multiple-choice questions, was tested at a workshop with eight local community college geoscience instructors in early September 2014. Based on their feedback, the questions were revised and narrowed to 10 questions. The pretest and posttest questions were exactly the same and were administered through Blackboard. Each question was worth 10 points.

To mitigate the effect of students tending to perform better the second time they take the same test (e.g., Krumboltz et al., 1960), the correct answers were not revealed to students in either the pretest or posttest. The



Rock Erodibility : 0.00015 kyr⁻¹

Hard/Soft Contrast : 5

Cliff Retreat Rate : 0.5 m/kyr

Simulation End Time : 0 Myr (Present)

Storage Intervals : 1

Rock Erodibility : 0.00015 kyr⁻¹

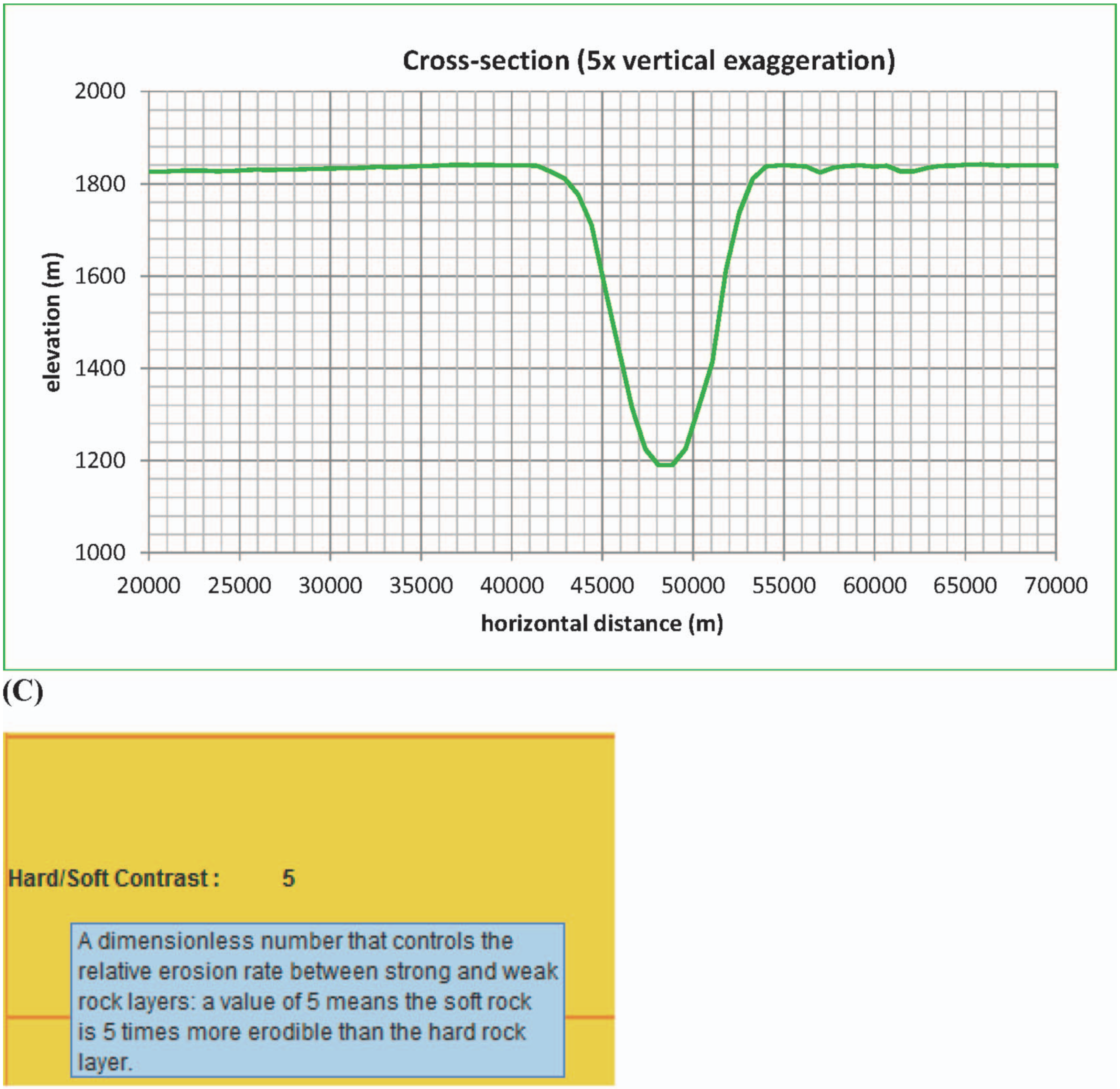
Hard/Soft Contrast : 5

Cliff Retreat Rate : 0.5 m/kyr

Simulation End Time : 0 Myr (Present)

Storage Intervals : 1

FIGURE 1: Screenshots of WILSIM-GC. (A) At about 3 million years ago (Ma). (B) At present. (C) Cross section created in Excel with saved cross-section data. (D) Help tool tip as mouse hovers over parameter name. The transparent planes in (A) and (B) with arrows show the location of the cross section.



(C)

Hard/Soft Contrast : 5

A dimensionless number that controls the relative erosion rate between strong and weak rock layers: a value of 5 means the soft rock is 5 times more erodible than the hard rock layer.

(D)

FIGURE 1: continued.

order in which the questions appeared to the students was also randomized for the two tests. To ensure that both groups had the same cumulative learning experience, the two groups switched learning activities, so that all students had the opportunity to learn from both exercises. The switch happened after the completion of the posttest so that the posttest measured the effect of each intervention (Fig. 2). At the end of the laboratory period, both groups completed a survey designed to measure the experience and their degree of satisfaction with the user-interface design of WILSIM-GC, the learning activities, and their attitude toward computer simulation in comparison to the traditional learning method.

One out of the five laboratory sections encountered Internet connection problems because of a campus network upgrade. Students in that section, as well as several students who did not complete both the pretest and posttest, were excluded from the study. Following the exclusion, the study sample ($N = 43$) consisted of a treatment group ($n = 20$) and a control group ($n = 23$) (see Table I). The demographic composition (grade level, major, ethnicity, and gender) of the 43 students included in the analysis is shown in Table II.

For the WILSIM-GC intervention (treatment group), students first ran a WILSIM-GC simulation with default parameter values and observed how the landform evolved over time in three-dimensional (3D) animation (see Fig. 1). Students also extracted the data for cross-sections of the canyon at 1 million-year intervals. The data were then brought into Microsoft Excel to plot the cross sections. To save time and to make it easy for those students who were not familiar with how to use Excel, a template Excel file was provided so that students could simply copy and paste the

TABLE I: Sample distribution.

Section[1]	Enrollment	Sample in Group A[2]	Sample in Group B[2]
1	17	7	7
2	13	4	7
4	5	2	2
5	19	7	7
Total	54	20	23

[1] Section 3 was excluded because of Internet connection problems at the time of the experiment.

[2] Students with incomplete answers were excluded. Statistics of test scores by sections were similar, but because of smaller samples in each section, they are combined into two groups in later analysis.

cross-section data into the Excel file, and cross-sections were generated automatically. The default parameter values were reasonable values for producing a landform similar to today's Grand Canyon, and that constituted the baseline scenario. Next, students explored the effects of changing the values of the rock erodibility and hard/soft contrast (from the baseline scenario) on the landform by observing its evolution over time in animation and comparing the cross sections. The changing of these variables was designed to help students understand downcutting and headward erosion. Because of the time limit of the laboratory, we only had students change these two parameters in this experiment. Detailed instructions were provided at each step, and students were asked to answer questions at various points, aimed at eliciting their understanding of the processes and concepts. For example, after changing the rock erodibility to a value lower (i.e., harder to erode) than the default value, we asked students how the width and depth of the valley changed compared with the baseline scenario and why. After changing the hard/soft contrast to 1 (i.e., no difference between stratigraphic rock layers), we asked students how the lengths of the tributary canyons changed and why (see supplemental material online at http://dx.doi.org/10.5408/15-080s1 for more details).

For the paper-based intervention (control group), students read written material explaining the different erosional processes involved in the formation of the Grand Canyon, with diagrams and pictures, including downcutting erosion, headward erosion, and knickpoint and its migration, and explaining what the typical, resulting landform would look like in a 3D perspective view and in cross section. Students were then asked to manually construct a cross section from a simplified contour map of a section of the Grand Canyon. Before that experiment, all students had already completed a laboratory class about contour maps and cross-section construction. The paper-based exercise in this experiment was simplified with a premade grid and scale, so that it was easy for students to draw the cross section. Students were also asked to answer questions designed to elicit understanding of how the shape of the cross section they constructed was related to processes they learned. For example, what is the shape of the deepest part of the cross section? What process do you think is responsible for the shape you see? Which part of the cross section is likely underlain by soft rock? Which part of the cross section is likely underlain by hard rock? Why? In parallel to the WILSIM-GC intervention, students were also asked to answer "what-if" questions. For example, what would the cross section look like if the rocks were harder to erode? Would the length of the tributary canyon increase or decrease if the rock layers were composed of the same type of hard rock (i.e., if there was no contrast in rock strength

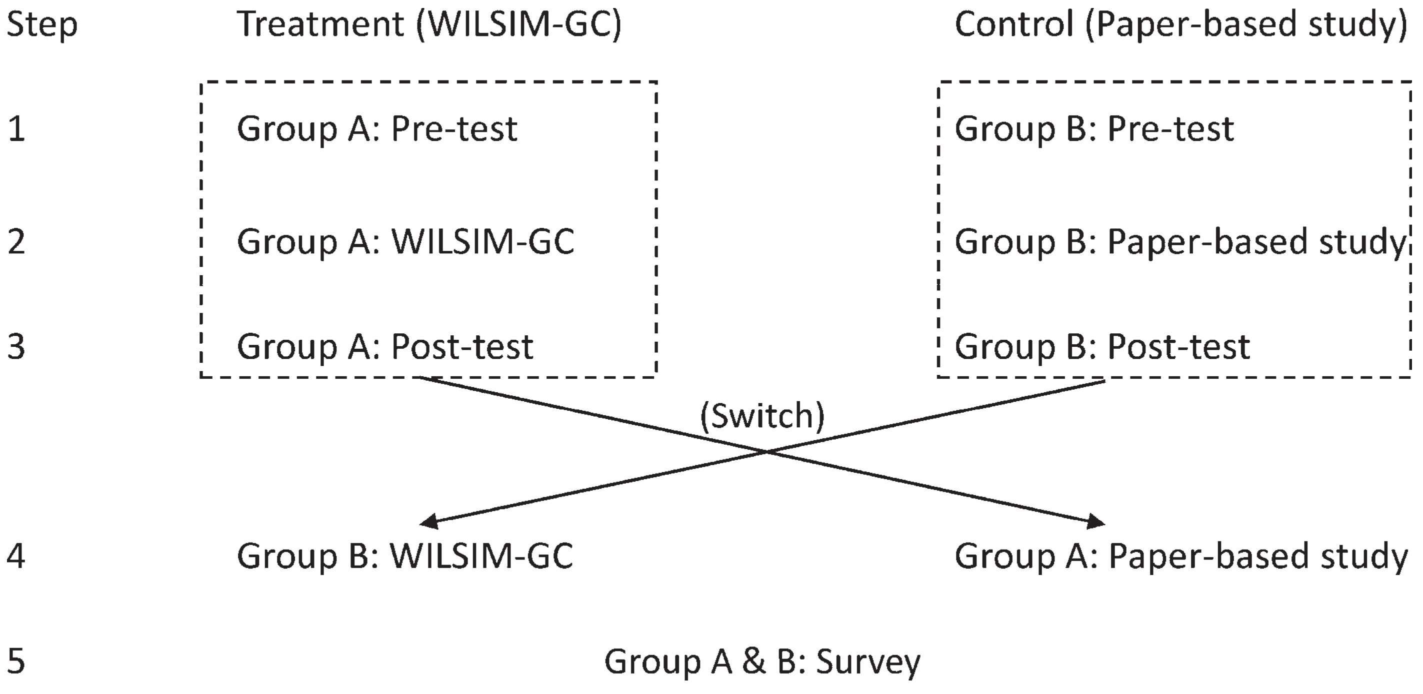


FIGURE 2: Diagram illustrating the procedure of the control/treatment experiment. Dashed boxes show the pretest and posttest comparison between the control and treatment groups. To ensure both groups had the same experience, they switched instructional conditions after completing the posttest. The attitudinal survey was conducted at the end.

TABLE II: Demographic information by groups.

	A (Treatment)	B (Control)	Total
Grade level			
Freshman	5	4	9
Sophomore	5	6	11
Junior	6	8	14
Senior	4	5	9
Major			
Geoscience	5	4	9
Non-Geoscience	15	19	34
Gender			
Male	18	13	31
Female	2	10	12
Ethnicity			
African American	3	3	6
Native American	0	0	0
Hispanic/Latino	0	4	4
Asian American	1	1	2
White/Caucasian	13	11	24
2+ Ethnicities	0	2	2

between the layers, and the erosion resistance was the same)?

Although the questions in the WILSIM-GC and paper-based laboratory materials are not exactly the same, both exercises were designed to elicit and measure students' understanding of the same underlying processes responsible for the formation of the Grand Canyon, based on effective strategies suggested in the literature (e.g., Bell and Trundle, 2008). We have made both exercises available on the Journal Web site as supplemental material. Additional curricular materials, designed based on the same principles, are available on the project Web site.

RESULTS

Data Screening and Analytic Approach

The χ^2 analysis of the demographic variables of grade level, major, and ethnicity (Table II) showed no significant differences between the treatment and control groups, confirming the two groups' homogeneity. Although there was a gender imbalance between the two groups ($p = 0.015$), *t*-test analyses showed no significant difference in the pretest and posttest scores between the two genders (either within group or between groups).

TABLE III: Pretest and posttest means and standard deviations and Cohen's *d* for treatment (simulation) and control (paper) groups.

Test	Control (Paper) ($n = 23$)		Treatment (Simulation) ($n = 20$)	
	$\bar{x}$	SD	$\bar{x}$	SD
Pretest	64.78	18.55	62.00	13.99
Posttest	72.17	15.65	76.50	13.48
Cohen's *d*	0.40		1.06	

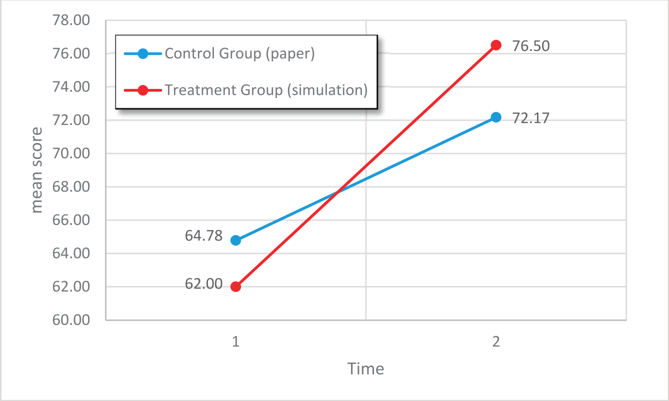


FIGURE 3: Interaction effect of experimental group and time (pretest to posttest) on test scores. Note: Time 1 = pretest; Time 2 = posttest.

We used a two-way mixed-design analysis of variance (2 × 2 ANOVA) test to identify the effect of different instruction intervention factors (using WILSIM-GC simulation versus traditional paper-based material); we examined the effect of the instruction (*simulation* versus *paper*, hereafter), time (pretest versus posttest), and the instruction × time interaction effect (i.e., pretest and posttest score differences for the simulation group compared with the paper group) on participants' test scores. A confidence level at $\alpha = 0.10$ was considered appropriate because of the small sample size (e.g., Banerjee et al., 2009). Data were screened to ensure that the assumptions of two-way ANOVA were fulfilled. Assumptions for performing parametric tests were also examined, including those for using a two-way mixed-design ANOVA with two repeated measures. This included assumptions verifying that pretest and posttest variables were continuous and approximately interval-scaled, that pretest and posttest scores were matched by participant, and that the categorical between-subjects factor (i.e., simulation versus paper group) represented independent groups. In addition, examination of studentized residuals indicated a lack of univariate outliers; Shapiro-Wilk tests and histograms verified that model residuals were normally distributed and variances between groups were homogeneous ($p = 0.596$). All common assumptions for performing the two-way mixed-design ANOVA were satisfied.

Tests of the Effect of the Intervention

Pretest and posttest means ($\bar{x}$) and standard deviations (SD) for both the simulation (treatment) and paper (control) groups are presented in Table III. The mean scores of posttests were higher than those of the pretests for both the simulation group and the paper group. Pretest scores were somewhat lower for the simulation group ($\bar{x} = 62.00$, SD = 13.99) than they were for the paper group ($\bar{x} = 64.78$, SD = 18.55). For the posttest, however, the order of means was reversed: the mean posttest score was higher for the simulation group ($\bar{x} = 76.50$, SD = 13.48) than it was for the paper group ($\bar{x} = 72.17$, SD = 15.65). This effect is illustrated in Fig. 3. The primary purpose of the two-factor

TABLE IV: Summary of two-way, mixed ANOVA for instruction (paper versus simulation) and time as a repeated measure (pretest and posttest).

Source	df	SS	MS	*F*	*p*
Within-subjects effects					
Time	1	2,563.32	2,563.32	18.54	<0.001***
Instruction × time	1	270.30	270.30	1.96	0.170
Residual	41	5,669.24	138.27		
Between-subjects effects					
Instruction	1	12.74	12.74	0.04	0.950
Residual	41	14,470.98	352.95		

***$p < 0.001$. Tests are two-tailed *p*-values were halved to reflect the results of the one-tailed tests deemed more appropriate. df = degrees of freedom, SS = sum of squares, MS = mean square.

ANOVA was to test this interaction effect for statistical significance.

Results of the 2 × 2 mixed-design ANOVA with time as a repeated measure (pretest and posttest) are presented Table IV. They show a significant within-subject effect of time [$F(1, 41) = 18.54$, $p < 0.001$]; across all participants, posttest scores were higher than pretest scores. The between-subject effect of intervention method was not significant [$F(1, 41) = 0.04$, $p = 0.95$]; mean test scores (combined across pretest and posttest) were not significantly different between the two groups. The instruction × time interaction effect was not statistically significant with a two-tailed test [$F(1, 41) = 1.96$, $p = 0.170$]. Based on previous studies, the likelihood that the paper group scored significantly higher than the simulation group was deemed negligible; thus, a one-tailed test was more appropriate. With a one-tailed test, the interaction is marginally significant (at $\alpha = 0.10$), with $p = 0.085$ and a moderate effect size ($\eta^2 = 0.05$) (Cohen, 1988). Relative to the paper group, the simulation group showed moderately greater growth from pretest to posttest; that is, the difference in scores from pretest to posttest was moderately greater for the simulation group.

In addition, the effect size captured by Cohen's *d* for the treatment group was 1.06 (>0.8), higher than it was for the control group, which was 0.40. This suggests that the WILSIM-GC intervention had high practical significance, with improvement above 1 SD, a large effect. In contrast, the improvement for the paper-based intervention was just less than 0.5 SD, a small to moderate effect (Cohen, 1992).

To see how students performed by question, we also examined the percentage of students who answered each question correctly and compared that percentage growth from pretest to posttest between the two groups. The results are shown in Table V along with the questions. The first five questions are focused on general concepts and terminology. The mean of the percentage of growth between the two groups were close, with an average growth of 12.1% for the paper group and a growth of 14.0% for the simulation group. The last five questions are more directly related to the specific case of the Grand Canyon and required students to apply what they had learned to answer the questions, i.e., these questions required higher-level thinking, as described in Bloom's taxonomy (Bloom et al., 1956). The mean of the percentage growth in correct answers between the pretest and the posttest in the simulation group was 15.0%, whereas the growth for the paper group was only 2.6% (see Fig. 4). This suggests that the students in the simulation group developed a deeper understanding of the geological processes involved in landscape evolution than did the students in the paper group.

It is also interesting to note that three questions in the paper group showed negative growth in the percentage of students with correct answers, whereas only one question (no. 4) in the simulation group showed this type of negative growth. Question 4 was about the concept of relief and was not discussed in detail in the WILSIM-GC material. The highest gain for both groups was the question about cross section (paper group, 30%; simulation group, 60%). The students in the paper group were required to construct a cross-section manually from a contour map, whereas the simulation group students were shown a transparent plane cutting through the valley (Figs. 1A and 1B). The visual effect of the model appeared to be noticeably more effective at conveying the concept of cross section to students than was the traditional method of constructing a cross section from a contour map.

For the survey questions at the end of the experiment (Table VI), students were asked to select a number between 1 and 6 to indicate whether they agreed or disagreed (1 being strongly disagree and 6 being strongly agree) with each statement. The mean scores for most questions were above 4, and average of mean scores (excluding questions 9 and 13) was 4.23, indicating that most students generally agree with the statements. Questions 9 and 13 were negative statements and received low scores, which means that they generally agreed with the opposite, i.e., WILSIM-GC was experienced as compatible with students' learning approaches and was easy to use. The three statements with which students agreed most strongly (in descending order) were:

10. *The visualization and animation of landform evolution in WILSIM-GC were informative.*
8. *WILSIM-GC helped me to think about "how did the Grand Canyon form."*
12. *It was easy for me to visualize and compare simulated results to real-world landforms when using WILSIM-GC.*

The last four questions (Table VI) specifically asked students to compare WILSIM-GC and paper-based methods, and the average of the mean scores of those four

TABLE V: Pretest and posttest questions and the growth in the percentage of questions answered correctly.[1]

Question No.	Question	Paper Growth	Simulation Growth
1	Which process is primarily responsible for the formation of V-shaped valleys?	13.04	10.00
	A) Downcutting erosion		
	B) Headward erosion		
	C) Knickpoint migration		
	D) Weathering		
2	The process in which a stream lengthens upslope by eroding towards its source is __________.	13.04	0.00
	A) Lateral erosion		
	B) Stream rejuvenation		
	C) Headward erosion		
	D) Downcutting		
3	The process in which flowing water cuts a channel or trough into the land surface to create a stream is ___________.	−4.35	5.00
	A) Lateral erosion		
	B) Stream rejuvenation		
	C) Headward erosion		
	D) Downcutting		
4	What is topographic relief?	8.70	−5.00
	A) The difference in elevation between two points, divided by the distance between those points.		
	B) The difference in elevation between two points.		
	C) The height above sea level of a particular point.		
	D) The difference in elevation between the highest and lowest points on a map.		
5	A graph that represents a two-dimensional slice of a river valley along its width, showing how elevation changes with horizontal distance is called _________.	30.43	60.00
	A) An aspect		
	B) A cross section		
	C) A contour		
	D) A long profile		
6	The Grand Canyon formed directly through geological processes related to __________.	−4.35	0.00
	A) Weathering of bedrock		
	B) Erosion by running water of the Colorado River		
	C) Transportation of eroded sediments downstream		
	D) All of the above		
7	It took _______ for the Grand Canyon to evolve into its present form.	4.35	5.00
	A) Hundreds of years		
	B) Thousands of years		
	C) Hundreds of thousands of years		
	D) Millions of years		

TABLE V: continued.

Question No.	Question	Paper Growth	Simulation Growth
8	How does the strength of rock layers affect the slope of resulting topography?	8.70	0.00
	A) Harder rocks tend to form steep slopes		
	B) Harder rocks tend to form gentler slopes		
	C) Softer rocks tend to form gentler slopes		
	D) Softer rocks tend to form steeper slopes		
	E) Both (A) and (C)		
9	If the rock layers in the Grand Canyon were harder than they really are, you would see ___________.	−4.35	35.00
	A) Increased width of the main canyon, decreased depth of the main canyon, and an increased length of the tributary canyons		
	B) Decreased width of the main canyon, increased depth of the main canyon, and an increased length of the tributary canyons		
	C) Decreased width of the main canyon, decreased depth of the main canyon, and a decreased length of the tributary canyons		
	D) Increased width of the main canyon, increased depth of the main canyon, and an increased length of the tributary canyons		
10	If the rock layers of the Grand Canyon were composed of the same type of hard rock (i.e., if there was no contrast in rock strength between the layers and the erosion resistance was the same), you would see ___________.	8.70	35.00
	A) A decreased length of the tributary canyons		
	B) An increased length of the tributary canyons		
	C) An increased width of the tributary canyons		
	D) No change in length of the tributary canyons		

[1] Note: The growth was calculated as the percentage of correct answers in the posttest minus its counterpart in the pretest. A negative number means the posttest correct percentage was smaller than it was for the pretest for that question.

questions was 4.39, indicating that students favored the WILSIM-GC approach over the traditional paper-based material.

DISCUSSION AND RECOMMENDATIONS

The results of this study are consistent with some of the findings of previous studies (e.g., Edsall and Wentz, 2007). Both computer-based simulation and traditional paper-based material can be effective at enhancing students' understanding of the processes involved in forming the Grand Canyon, as measured by improvements in scores from the pretest to posttest within each group. However, the effect size as measured by Cohen's *d* of the improvement in scores for the simulation group was large, whereas that for the paper group was small to medium. Two-way mixed-design ANOVA showed that there was marginally significant instruction × time interaction effect on student learning at $\alpha = 0.10$ ($p = 0.085$). The significance level of $\alpha = 0.10$ is sometimes adopted in studies with small sample sizes (e.g., Banerjee et al., 2009), and in such cases, the effect size takes on increased importance. The observed p-value of 0.085 suggests that the probability of making a type I error (i.e., incorrectly rejecting the null hypothesis when in fact there is no difference between the two groups in the population) was 8.5%, only slightly above the conventional cutoff of 5%. In other words, there is a 91.5% probability that students in the simulation group did, in fact, learn more than the students in the paper group, and the larger effect size for the simulation group reinforces that likelihood. Nonetheless, future experiments that address the limitations to our study (see the "Limitations" section) would improve our confidence in the encouraging preliminary findings.

The implications of our findings and our recommendations are the following:

(1) The advantages of using a simulation model, in terms of its potential to promote higher-level thinking as demonstrated in this study with WILSIM-GC, can and should be leveraged in teaching students the difficult-to-master concepts and processes of landform evolution. We believe the benefits of using simulations are worth investing the time and effort to develop the associated curricular materials. This is also supported by the literature (e.g., Smetana and Bell, 2012).

(2) Traditional paper-based approaches should not be discarded because they are similarly effective (albeit with a small to medium effect size) for teaching geoscience concepts, information, and terminology. In fact, as suggested in previous studies, scaffolding using traditional teaching approaches is necessary to help students develop enough background knowledge so that they are ready to explore within

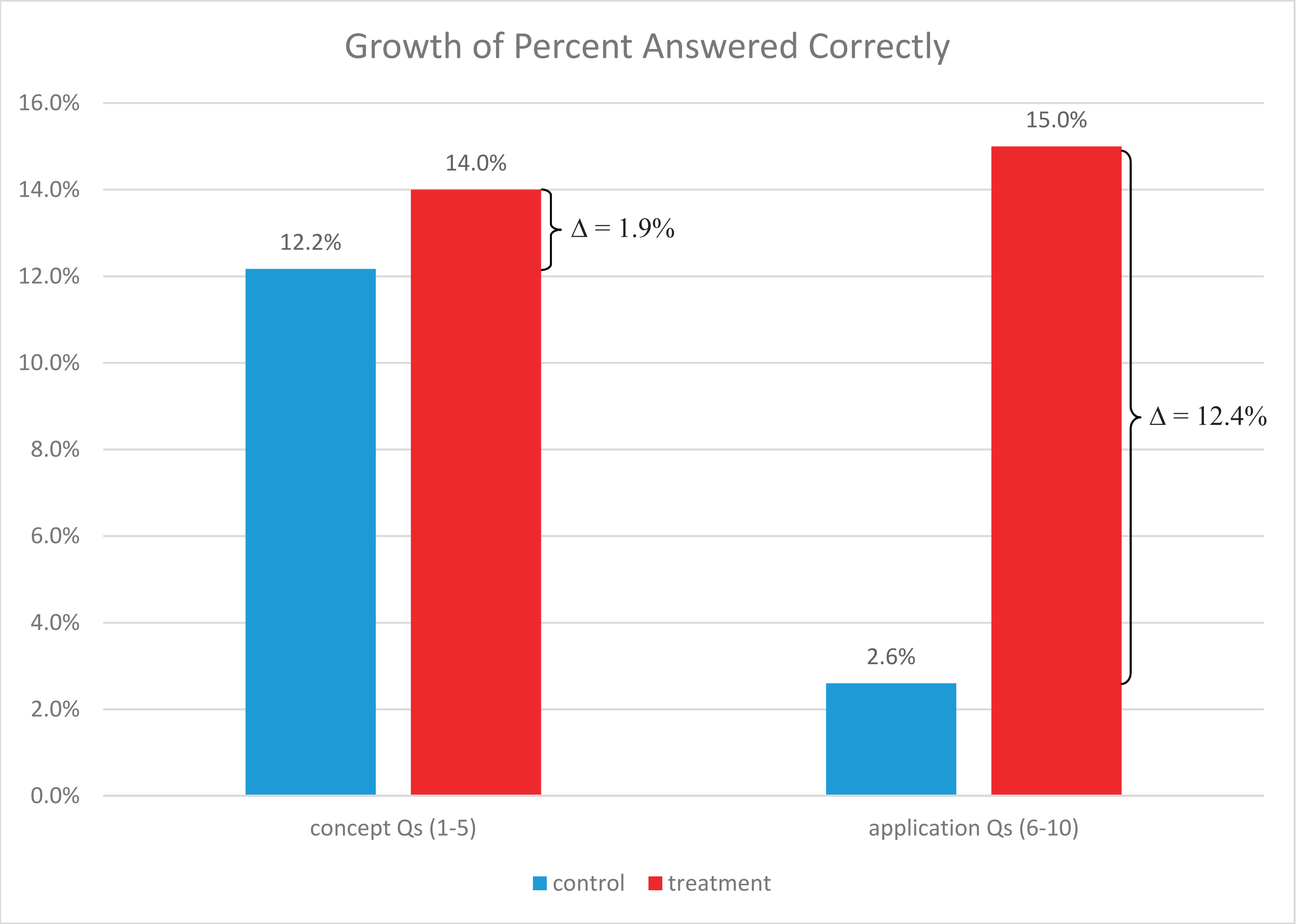


FIGURE 4: Comparison of growth in the percentage of questions answered correctly. The growth between the paper (control) and simulation (treatment) groups for the concept questions (questions 1–5) are small, but for the application questions (questions 6–10) is large (6.5 times larger).

simulations (e.g., Khan, 2011; Schneps et al., 2014). We agree with this suggestion and recommend that traditional approaches be used in curricular materials that provide the basic concepts and foundation for more-advanced exploration and problem solving with computer simulations.

(3) Such integration (of traditional approaches and simulations) may be critical to designing better curricular materials, especially for online courses in which direct interaction with an instructor is not readily available. An online simulation model, such as WILSIM-GC, also lends itself naturally to the increasingly common practice of the "flipped classroom" approach, where traditional lecturing is replaced with interactive activities in the classroom, and online learning is conducted outside of the classroom.

LIMITATIONS

There are a number of limitations of this study. First, the sample size was relatively small. This may limit the power of the statistics to reveal true differences and may explain why the ANOVA interaction was only significant at the $\alpha = 0.10$ level. A larger sample size would increase the power and would more-reliably indicate group differences. Second, the intervention time was short: approximately 45 min, allowing students to complete the intervention, the posttest, and the other laboratory activities during a single laboratory period of 1 h and 50 min. This limited the number of concepts we could cover in the exercises. Third, we conducted this experiment in a 100-level course (composed of five sections, see Table I) with mostly nonmajor students (Table II). Longer intervention time and testing in advanced level courses with mostly major students would allow us to probe for deeper understanding of more-complex concepts. Fourth, students' written answers to the open-ended questions during the WILSIM-GC and paper-based exercises were not collected, which precluded us from analyzing those written answers and gaining more insights from them. Fifth, multiple-choice questions in the pretest and posttest may introduce some uncertainties: students may have guessed some answers correctly. Last, the assignment of groups was by alphabetic order, which is not truly random, i.e., each student's chance of being assigned to either group was not the same, depending on their last name. Nonran-

TABLE VI: Attitudinal survey summary.[1]

Question No.	Question	$\bar{x}$	SD
1	WILSIM-GC helped me understand how Earth systems work over geologic time scales.	4.13	1.28
2	WILSIM-GC made me feel I could solve the problem based on the information given.	4.02	1.23
3	WILSIM-GC helped me have a clear understanding of how I arrived at my final outcomes.	3.74	1.34
4	WILSIM-GC provided me a better way to analyze landform evolution.	4.39	1.29
5	WILSIM-GC encouraged me to identify the critical features of landform evolution.	4.35	1.25
6	WILSIM-GC helped me apply my understanding of the landform evolution.	4.37	1.34
7	WILSIM-GC was engaging and interesting.	4.33	1.48
8	WILSIM-GC helped me to think about how the Grand Canyon was formed.	4.63	1.34
9	WILSIM-GC was not compatible with my learning approach.	3.09	1.84
10	The visualization and animation of landform evolution in WILSIM-GC were informative.	4.65	1.43
11	It was easy to navigate among the various features of WILSIM-GC.	4.48	1.80
12	It was easy for me to visualize and compare simulated results to real-world landforms when using WILSIM-GC.	4.63	1.74
13	It was difficult to use WILSIM-GC.	2.87	2.31
14	The inquiry activities/problems were about the right length.	4.54	1.77
15	I am confident that I understand how to use WILSIM-GC.	4.56	2.07
16	I put enough effort into learning WILSIM-GC.	4.63	2.19
17	I feel WILSIM-GC provided inadequate guidelines to help me solve the problems.	3.63	2.46
18	I feel WILSIM-GC provided inadequate functions to facilitate discussions.	3.39	2.53
19	I want more training on WILSIM-GC.	3.24	2.75
20	I would like to continue to use WILSIM-GC.	3.78	2.84
21	I would encourage others to use WILSIM-GC.	4.17	2.97
22	Compared to using paper-based self-study material, WILSIM-GC offered me better management of my thinking process toward the inquiry activities.	4.31	3.10
23	Compared to using paper-based self-study material, WILSIM-GC was more time-efficient as a learning activity.	4.30	3.24
24	Compared to using paper-based self-study material, WILSIM-GC was more convenient to use.	4.41	3.47
25	Compared to using paper-based self-study material, WILSIM-GC was more fun to use.	4.54	3.49

[1] Note: A value of 1 means students strongly disagreed, whereas a value of 6 means students strongly agreed with a statement.

dom grouping may inadvertently introduce extraneous factors, confounding the true effect of intervention (Ary et al., 2010, p. 268).

FUTURE WORK

To address the limitations identified above, several strategies have been developed or planned for future work. Some are currently underway. We intend to use a larger sample size and assign treatment and control groups randomly, which would increase our ability to more-reliably reveal group differences and to reduce the chance of encountering imbalances in terms of gender or other factors. To increase our sample size, we plan to run this study again in coming semesters here at Northern Illinois University and at other collaborating institutions. In addition, we plan to conduct student interviews and focus groups and to collect and analyze written answers to open-ended questions in each group or incorporate such questions into the pretest and posttest questions in future experiments to acquire new insights into students' understanding of concepts that require higher-level thinking. As we gain more experience and the model becomes more mature, we will test it in higher-level major courses.

Alliger and Horowitz (1989) and Barge (2007) addressed the issue of guessing in tests with multiple-choice questions by asking students to identify whether they knew the answer or they guessed the answer for each question. In contrast to the traditional scoring method of counting all correct answers (''guessed'' and ''knew'') as correct, they only scored those correct answers that students indicated they knew as correct answers. They found a 10% (Barge, 2007) to 15% (Alliger and Horowitz, 1989) increase in knowledge gained (measured by pretest and posttest) with the new scoring method as compared with the knowledge gained using the traditional scoring method. Future experiments may adopt this approach to address the uncertainty caused by guessing. Additionally, it will also be important to examine with greater detail the psychometric properties of the instrument as well as to confirm the difficulty levels of the items through the Rasch model or other such analysis based on item response theory (Cavanaugh and Waugh, 2011).

To keep the simulation computation time to within a few minutes, we have coarsened the model resolution. Unfortunately, this prevents the cross-section data from showing the finer details of the stair-step morphology created by the hard/soft contrast in the Grand Canyon stratigraphy. We are currently working on providing a higher-resolution model for a section of the Grand Canyon to address this shortcoming.

Despite limitations, our initial results show great promise for WILSIM-GC as an instruction tool to help students learn challenging concepts in landform evolution. Our long-term goal is to expand the WILSIM modeling framework to include other landform processes, such as glacial and eolian processes (Pelletier, 2008). Based on the initial success with our step-wise approach of building, testing, and improving, we are confident that we will make the WILSIM model an easy-to-use, next-generation landform simulator that can be leveraged to enhance student learning in various modes of delivery.

Acknowledgments

This project is funded through NSF-TUES program (award DUE-1140375) and further information can be found at http://serc.carleton.edu/landform/index.html. We thank Dr. Tom Smith, Department of Educational Technology, Research and Assessment, Northern Illinois University, for his helpful input and discussion on statistical methods. We thank the GEOG 102 course teaching assistants Steven Spradling and Sarah Smith for their assistance with the experiment and Raye Chiang for compiling the attitudinal survey data.

REFERENCES

Adams, W.K., Paulson, A., and Wieman, C.E. 2008. What levels of guidance promote engaged exploration with interactive simulations? *AIP Conference Proceedings,* 1064(1):59–62. doi:10.1063/1.3021273.

Akpan, J.P. (2001). Issues associated with inserting computer simulations into biology instruction. *Electronic Journal of Science Education,* 5(3). Available at http://ejse.southwestern.edu/article/view/7656/5423 (accessed 9 January 2015).

Alliger, G.M., and Horowitz, H.M. 1989. IBM takes the guessing out of testing. *Training and Development Journal,* 43(4):69–73.

Anglin, G.J., Vaez, H., Cunningham, K.L. 2004. The role of static and animated graphics. *In* Jonassen, D.H., ed., Handbook of research for educational communications and technology, 2nd ed. Bloomington, IN: Association for Educational Communication & Technology, p. 865–916.

Ary, D., Jacobs, L.C., Sorensen, C.K. 2010. Introduction to Research in Education. Belmont, CA: Wadsworth Cengage Learning, p. 669.

Baker, V.R., 2014. Terrestrial analogs, planetary geology, and the nature of geological reasoning. *Planetary and Space Science,* 95:5–10.

Banerjee, A., Chitnis, U.B., Jadhav, S.L., Bhawalkar, J.S., and Chaudhury, S. 2009. Hypothesis testing, type I and type II errors. *Industrial Psychiatry Journal,* 18(2):127–131. doi: 10.4103/0972–6748.62274.

Barge, G.L. 2007. Pre- and posttesting with more impact, *Journal of Extension,* 45(6). Available at http://www.joe.org/joe/2007december/iw1.php. (accessed 14 January 2016).

Bell, R.L., and Trundle, K.C. 2008. The use of a computer simulation to promote scientific conceptions of moon phases. *Journal of Research in Science Teaching,* 45, 346–372.

Bloom, B.S., Engelhart, M.D., Furst, E.J., Hill, W.H., and Krathwohl, D.R. 1956. Taxonomy of educational objectives: The classification of educational goals handbook I—Cognitive domain. New York: David McKay Company.

Cavanagh, R.F., and Waugh, R.F. (2011). Applications of Rasch measurement in learning environments research. Rotterdam, the Netherlands: Sense Publishers.

Chase, C.G. 1992. Fluvial landsculpting and the fractal dimension of topography. *Geomorphology,* 5:39–57.

Chen, S. 2010. The view of scientific inquiry conveyed by simulation-based virtual laboratories. *Computers and Education,* 55(3):1123–1130. doi:10.1016/j.compedu.2010.05.009.

Cohen, J. 1988. Statistical power analysis for the behavioral sciences, 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates.

Cohen, Jacob. 1992. A power primer. *Psychological Bulletin,* 112(1):155–159. doi: 10.1037/0033-2909.112.1.155.

Day, T. 2012. Undergraduate teaching and learning in physical geography. *Progress in Physical Geography,* doi: 10.1177/0309133312442521.

de Jong, T. 2006. Technological advances in inquiry learning. *Science,* 312:532–533.

de Jong, T., and van Joolingen, W. 1998. Scientific discovery learning with computer simulations of conceptual domains. *Review of Educational Research,* 68(2):179–201.

Edsall, R., and Wentz, E. 2007. Comparing strategies for presenting concepts in introductory physical geography: Physical models vs. computer visualization, *Journal of Geography in Higher Education,* 31(3):427–444.

Kastens, K., Manduca, C.A., Cervato, C., Frodeman, R., Goodwin, C., Lieben, L.S., Mogk, D.W., Spangler, T.C., Stillings, N.A., and Titus, S. 2009. How geoscientists think and learn. *EOS Transactions American Geophysical Union,* 90(31):265.

Khan, S. 2011. New pedagogies on teaching science with computer simulations. *Journal of Science Education and Technology,* 20(3):215–232. doi:10.1007/s10956-010-9247-2.

Krumboltz, J.D., and Christal, R.E. 1960. Short-term practice effects on tests of spatial aptitude. *Personnel and Guidance Journal,* 38:385–391.

Luo, W., Duffin, K.L., Peronja, E., Stravers, J.A., and Henry, G.M. 2004. A Web-based interactive landform simulation model (WILSIM). *Computers and Geosciences,* 30(3):215–220.

Luo, W., and Konen, M. 2007. New results from Using a Web-based interactive landform simulation model (WILSIM) in a general education physical geography course. *Journal of Geoscience Education,* 55(5):423–425.

Luo, W., Peronja, E., Duffin, K.L., and Stravers, J.A. 2006. Incorporating nonlinear rules in a Web-based interactive landform simulation model (WILSIM). *Computers and Geosciences,* 32(9):1512–1518. doi: 10.1016/j.cageo.2005.12.012.

Luo, W., Stravers, J.A., and Duffin, K.L. 2005. Lessons learned from using a Web-based interactive landform simulation model (WILSIM) in a general education physical geography course. *Journal of Geoscience Education,* 53(5):489–493.

Pelletier, J.D. 2008. Quantitative modeling of earth surface processes. Cambridge, UK: Cambridge University Press, p. 2995.

Pelletier, J.D. 2010. Numerical modeling of the late Cenozoic geomorphic evolution of Grand Canyon, Arizona. *Geological Society of America Bulletin,* 122:595–608.

Perkins, K., Moore, E., Podolefsky, N., Lancaster, K., and Denison, C. 2012. Towards research-based strategies for using PhET simulations in middle school physical science classes. *AIP Conference Proceedings,* 1413(1):295–298. doi:10.1063/1.3680053.

Podolefsky, N.S., Adams, W.K., Lancaster, K., and Perkins, K.K. 2010a. Characterizing complexity of computer simulations and implications for student learning. *AIP Conference Proceedings,* 1289(1):257–260. doi:10.1063/1.3515215.

Podolefsky, N.S., Moore, E.B., Perkins, K.K. 2013. Implicit scaffolding in interactive simulations: Design strategies to support multiple educational goals. Available at http://arxiv.org/ftp/arxiv/papers/1306/1306.6544.pdf. (accessed 13 January 2016).

Podolefsky, N.S., Perkins, K.K., and Adams, W.K. 2010b. Factors promoting engaged exploration with computer simulations. *Physical Review Special Topics—Physics Education Research,* 6(2):020117.

Ramasundarm, V., Grunwald, S., Mangeot, A., Comerford, N.B., and Bliss, C.M. 2005. Development of an environmental virtual field laboratory. *Computers,* 45(1):21–34.

Randy, L.B., and Trundle, K.C. 2008. The use of a computer simulation to promote scientific conceptions of moon phases, *Journal of Research in Science Teaching,* 45(3):346–372.

Scalise, K., Timms, M., Moorjani, A., Clark, L., Holtermann, K., and Irvin, P.S. 2011. Student learning in science simulations: Design features that promote learning gains, *Journal of Research in Science Teaching,* 48(9):1050–1078.

Schneps, M.H., Ruel, J., Sonnert, G., Dussault, M., Griffin, M., and Sadler, P.M. 2014. Conceptualizing astronomical scale: Virtual simulations on handheld tablet computers reverse misconceptions. *Computers and Education,* 70:269–280. doi:10.1016/j.compedu.2013.09.001.

Smetana, L.K., and Bell, R.L. 2012. Computer simulations to support science instruction and learning: A critical review of the literature. *International Journal of Science Education,* 34(9):1337–1370. doi: 10.1080/09500693.2011.605182.

Stumpf, R.J., Douglass, J., and Dorn, R.I. 2008. Learning desert geomorphology virtually versus in the field. *Journal of Geography in Higher Education,* 32(3):387–399.

Tversky, B., Morrison, J., and Betrancourt, M. 2002. Animation: Can it facilitate? *International Journal of Human Computer Studies*, 57:247–262.

Windschitl, M., and Andre, T. 1998. Using computer simulations to enhance conceptual change: The roles of constructivist instruction and student epistemological beliefs. *Journal of Research in Science Teaching,* 35(2):145–160.