Assessing the effect of systems simulations on systems understanding in undergraduate environmental science courses

Heather Skaza and Krystyna A. Stave
University of Nevada Las Vegas
Department of Environmental Studies
4505 Maryland Parkway Box 454030
Las Vegas, Nevada 89154-4030
702-895-4771 (office) and 702-895-4436 (Fax)
hjskaza@hotmail.com, krystyna.stave@unlv.edu

Abstract

This paper describes the results of a paired experiment testing the effect of system dynamics simulations on systems understanding in undergraduate environmental science courses. The performance of 298 students in four sections was measured at several points during the semester. Half the students used system dynamics simulations in their assignments; the other half did not. Results of regression analysis show that performance on systems questions immediately following the intervention was significantly better for the experimental group than the control. The study also highlighted some problems in the assessment framework we used and led to suggestions for improving both the systems interventions and the assessment tools.

Introduction

Assessing the effect of systems interventions on learning is challenging. Some of the challenges include clearly defining learning objectives, designing systems interventions targeted to specific systems thinking skills, specifying performance objectives, and designing rigorous and repeatable ways to assess their effects.

As Hopper and Stave (2008) reported based on a meta-analysis of systems intervention studies, very few studies provide comparable data. Most of the information we have about the effects of systems interventions in the classroom is anecdotal. Most other studies that attempt to measure a change in student understanding use a systems intervention, but without a control group (Korfiatis, Papatheodorou, and Stamou, 1999, Hogan, 2000, Evagorou, et. al., 2007).

The experimental studies that have been conducted have different purposes and assessment techniques. Some have tested student ability to control dynamic systems in a simulation environment (Cavaleri, Rapheal and Filetti, 2002, Jenson and Brehmer, 2003). In these cases, systemic understanding was measured by the student's success in achieving the best outcome for the system, as defined by the researchers. Others measure student ability to predict behavior, given other variable parameters in textual and/or graphical form (Sterman and Booth

Sweeney, 2002). In these studies systemic understanding was measured by the student's identification of the correct dynamic behavior for the given conditions.

A few studies have used controlled experiments. Fisher (2009) conducted a controlled experiment with some students building systems simulations in Vensim, while others used their traditional tool, the graphing calculator. She reported a significant increase in understanding for the simulation-builders compared to the control group. Wheat (2008) tested economics students' understanding of macroeconomics principles using systems thinking tools and reported a preference for them among the students who used them and an increased conceptual understanding for the students that used them, over the students that did not. Doyle, Radzicki and Trees(1998) and Vennix (1990) both report on studies that tested the relative effectiveness of using systems simulations on undergraduate students' systemic understanding of economic systems. Pala and Vennix (2005) conducted a controlled experiment testing the effect of a systems thinking course on students ability to correctly identify the level of a stock for given flow conditions.

The goal of this study was to conduct a controlled experiment on several systems interventions with repeatable assessment measures.

Description of the study

The current study tests the relative effectiveness of using systems simulations to increase students' systemic understanding of environmental issues in an introductory environmental science course.

Problem Statement

In designing and implementing the experiment, we addressed one main question: Does the use of systems simulations in an introductory environmental science course increase students' systemic understanding of environmental issues?

From that question and through the experiment design process another question developed: How do we best assess a change in systemic understanding?

Hypotheses

We believed that we would see a greater systemic understanding of environmental issues for the group of students using the systems simulations than for the students who did not. This general hypothesis was broken down in to several subhypotheses.

1) Simulation users would have a better overall course performance.
2) Simulation users would perform better on assessments that evaluated overall systems knowledge.

3) Simulation users would demonstrate a greater systemic understanding of the environmental issues the simulations focused on, form the beginning of the course to the end of the course.
4) Simulation users would show a greater systemic understanding of environmental issues on assessments following interventions that use the simulations.

Method

The study subjects were 304 students enrolled in four sections of Introduction to Environmental Studies at the University of Nevada-Las Vegas during the fall semester of 2009. Table 1 shows meeting days and times and class sizes. One small class and one large class were randomly selected to be the experimental groups. The other two sections were the control groups. The two large sections met in a lecture hall, while the smaller classes met in smaller classrooms.

Table 1. *Group Information for Introduction to Environmental Science Classes*

Section number	**Group**	**N**	**Meeting Day**	**Meeting Time**
001	Control	50	Mon., Wed.	10:00-11:15AM
002	Control	105	Mon., Wed.	11:30 AM-12:45 PM
003	Experimental	56	Tues., Thurs.	10:00 -11:15AM
004	Experimental	93	Tues., Thurs.	11:30 AM-12:45 PM

Course Design

The class had five educational components: assigned text book readings, in class lecture, six assessments, an activity that encouraged students to tie course concepts to their day to day experiences, and five assignments based on the readings and lecture. We used the same text, conducted the same lectures and assessments and expected students to complete the same activities for all sections. The only difference between the classes was that the experimental sections used systems simulations to complete three of five assignments. The control sections completed the same assignments, but with only a text description of the environmental issue the assignment focused on. Figure 1 is a timeline of assignment and quiz completion.

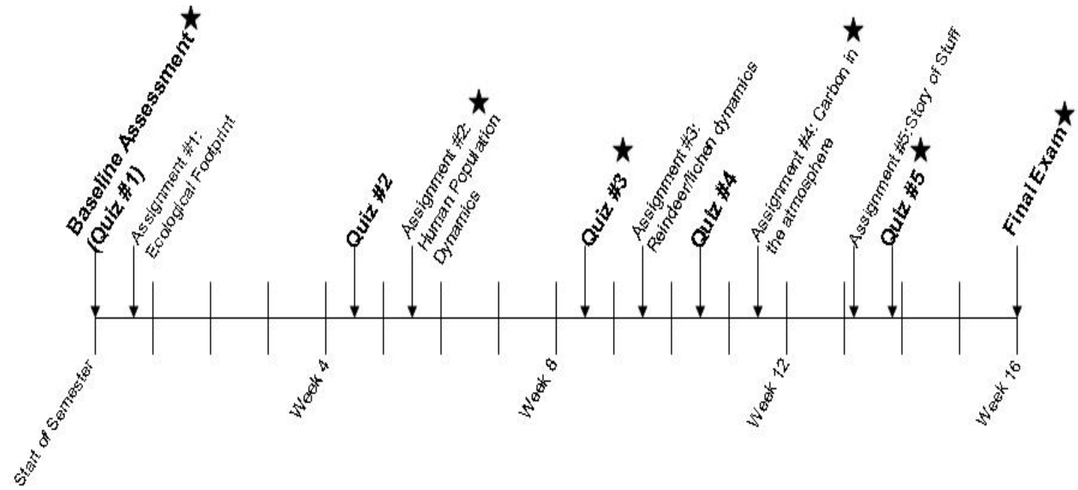

Figure 1. Timeline of assignments and assessments.

Description of Interventions

We used the course assignments to administer systems simulation interventions. Five assignments guided students to examine causal relationships in the environmental issues presented in class. Three of these included a systems simulation for the experimental sections. We gave the control sections an equivalent text description of the system the simulations were based on. All students answered questions in an online assessment with their assignment. The assessment questions asked students about the system they studied in their homework assignment, whether it was through simulation use or text description. Students completed assignments individually and on-line. There was no live guidance from an instructor. Table 2 describes the five assignments.

Table 2. *Homework Assignments and Descriptions*

Assignment	Description
1: Ecological Footprint	SIMULATION: Global Footprint Network ecological footprint calculator Students used an ecological footprint calculator to calculate their ecological footprints. They answered questions about their eco-footprint and how it might compare to someone living in a developing country.
2: Human Population Dynamics	**SYSTEMS SIMULATION**: Original model with total population as the stock, birth rate as the inflow and death rate as the outflow. Students were asked to describe the effect on total population when the number of births and number of deaths in a population are increased or decreased.
3: Reindeer/lichen relationships	**SYSTEMS SIMULATION**: Model of reindeer herd/lichen dynamics (Tabacaru et al., 2009) gives student a tutorial on how to manage the reindeer herd and instructs them to decide on herd size every year for fifteen years to maintain lichen growth at an optimum for their survival. Students were asked to manage a herd of reindeer so that the lichen that is their primary food source is not overgrazed.
4: Carbon in the atmosphere	**SYSTEMS SIMULATION**: Sterman's (2006) bathtub model allows students to increase and decrease carbon dioxide

	emissions. Students were asked to test out carbon emissions levels and note the effect on CO2 in the atmosphere. Assessment questions asked them to relate the stock and flows in the system.
5: The Story of Stuff	No simulation, but the students watch an online video to explain the way that the “stuff” we use moves around Earth’s system. They answered questions asking them to reflect on their role in the consumer cycle.

For this study, we used systems simulations to address three environmental issues: human population dynamics, reindeer and lichen population dynamics (Tabacaru et al., 2009) and carbon accumulation in the atmosphere. We did not analyze the data gathered from the reindeer/lichen exercise. We analyzed data related to the human population dynamics and the carbon accumulation in the atmosphere simulation. These assignments are described below.

Assignment #2: Human Population Dynamics

Assignment #2: Human Population Dynamics was the first assignment in which students used a systems simulation to help students understand an environmental problem. The reading material and simulation for this assignment described global population change as the difference between the number of births and the number of deaths. It was a very simple, one-stock, two-flow system. We broke the assignment into three parts, though the control group only completed the first part.

For Part 1 of the assignment students read a chapter in their textbook on human population change. The book describes world population change as the net difference between the number of births and the number of deaths. All students answered the same question set after reading the book. Assessment questions asked students to describe 1) how global population changes when either birth rate or death rate change and all other variables stay the same or 2) how global population changes when birth rate and death rate are equal, 3) how the birth rate and death rate are related to total population change. The variables were never explicitly described in terms of stocks and flows.

For Part 2 of the assignment, students in the experimental sections used a systems simulation created using Stella software (2010) and made available on the internet by the isee NetSim server. There were two slider bars and two buttons on the simulation’s interface. The total population output graph on the interface had a time horizon of three hundred years and was modeled after the total population change graph used in the course textbook. Under baseline conditions, the graph showed population growing exponentially until it reached about 10 billion people around the year 2050. Students could manipulate birth rate and death rate using two slider bars. Two actions buttons allowed students to run the simulation by clicking “GO” and restart the simulation by clicking “CLEAR.” We kept the simulation structure and interface as simple as possible.

We assumed that students in the introductory course had no experience with population dynamics or a simulation environment. The model interface is shown in Figure 2.

We gave the students a set of instructions for using the simulation. The instructions directed them to investigate the population dynamics they were asked to describe in Part 1 of the assignment. We instructed them to:

1) Run the simulation with current birth rate and death rate to note exponential growth pattern. The trend the simulation produced was identical to the one in their textbook.
2) Decrease the number of births by about one quarter using the slider bar on the interface. Birth rate was still greater than death rate, so population grew exponentially, but at a slower rate.
3) After returning to the initial condition, increase death rate by about one quarter. Again, birth rate remained above death rate and population grew at a slower rate that in the initial condition.
4) Make the number of births and number of deaths equal. Population stayed the same for the duration of the time horizon.

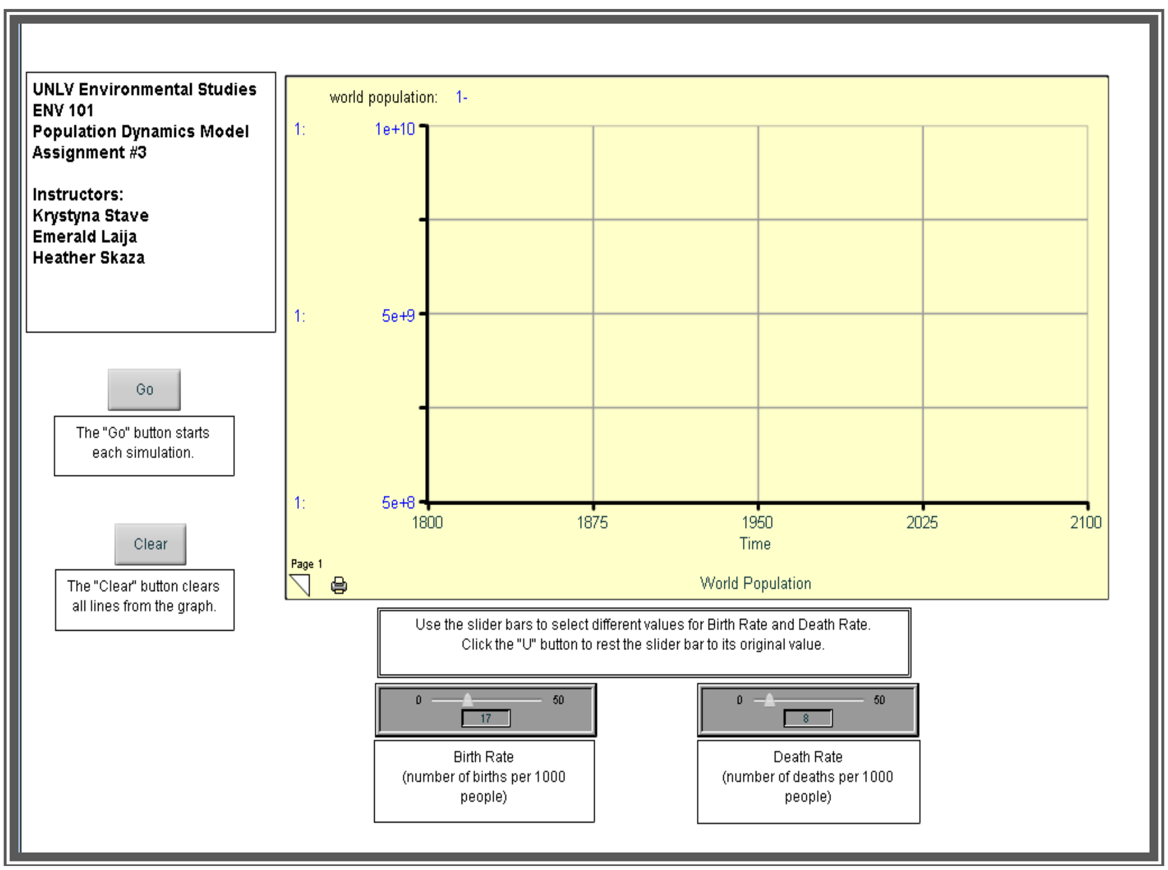

Figure 2. Simulation Interface for Population Dynamics Assignment

For Part 3 of the assignment, experimental group students answered another question set. We asked them to describe the model outputs under each set of conditions and compare them to their hypotheses in Part 1 of the assignment. We asked them to describe each output, whether this trend was surprising to them and why they thought total population changed the way that it did. The question set paralleled the questions asked on Part 1 of the assignment.

We debriefed the assignment for all sections the day after it was submitted on-line. During lecture, we reviewed the question set from Part 1 as a group. We prompted students to tell us what graphs they chose for each question and why they believed total population would change the way that they did. Instructors discussed each question and explained the correct answer if the class did not come to it. For the experimental sections, we also discussed what happened when they ran the simulation in each of the birth rate/death rate conditions.

Assignment #4: Carbon in the Atmosphere

The second systems simulation intervention we tested was Assignment #4: Carbon in the Atmosphere. Again, the assignment was divided into three parts. This time both groups completed all three parts.

For Part 1 of the assignment, all students read John Sterman's "Risk Communication on Climate: Mental Models and Mass Balance" (2008). The article summarizes the findings of Sterman's previous work, describing a general inability for people to understand carbon accumulation in the atmosphere as the net difference between carbon emissions and carbon absorption.

Experimental group and control group students completed different activities for Part 2 of the assignment. The experimental sections read a description of carbon accumulation in the atmosphere on line and the used the Bathtub Dynamics and Climate Change simulation developed by the MIT System Dynamics Group. The simulation introduced students to the stock and flow dynamics associated with carbon accumulation in the atmosphere and then directed them to control carbon emissions under a variety of conditions. For the first part of the simulation, the student's goal was to adjust carbon emissions relative to absorption to produce a trend for carbon in the atmosphere identical to one already displayed on the simulation screen. For the next part of the simulation, students tried to control emissions to keep carbon in the atmosphere at a particular level under conditions of sink saturation and delay. For Part 2, the control sections only read the system description that accompanied the simulation.

For Part 3 of the assignment, both groups answered the same set of questions. The question set asked students to relate carbon emissions and carbon absorption to carbon in the atmosphere in a number of ways. The question set is shown in Figure 6.

Assessments

Students completed a baseline quiz, four periodic quizzes throughout the semester, and a final exam. We also used the questions that students answered on Assignment #2 and Assignment #4 for analysis.

Baseline Quiz and Final Exam Questions

The baseline quiz was a pretest measure. Students took the baseline quiz online. On the first day of class, we instructed them on how to access the baseline quiz on the course website. They completed the assessment by the second class meeting, prior to any instruction. We graded the baseline quiz for completion, not correctness. Students received full credit for any answer. The baseline quiz contained five sections: general knowledge, systems knowledge, the New Ecological Paradigm assessment (Dunlap et al., 2000) (which assessed their attitude and opinions toward the environment), environmental practices, and demographic information. We evaluated the general knowledge and systems knowledge portions of the baseline quiz for this study.

The general knowledge portion of the baseline quiz contained twelve questions that covered a variety of environmental topics. The questions came from Wright's Environmental Literacy Instrument (2007), though we edited some for clarity. We chose these questions because they had already been tested for validity and they tested knowledge that would be discussed in the course.

The systems knowledge portion of the baseline quiz consisted of ten original questions designed to assess students systems thinking abilities and ability to read graphs that related to systems concepts. Five of these were evaluated for systemic understanding. One, short-answer question tested students' systemic knowledge of population dynamics. Four questions tested students' systemic knowledge of carbon accumulation in the atmosphere: three multiple-choice questions and one short-answer question.

The final exam was comprehensive. We administered the final exam in class, on paper, on the last day of class. The final exam included all of the questions on the baseline assessment, except for the demographic information questions. We included these questions on the final exam as a post-test measure.

Quizzes

We administered all other quizzes during the semester in class and on paper. Each quiz between the baseline quiz and the final exam contained about twenty questions that were either multiple choice or short answer. Multiple-choice prompts were either questions to be answered or statements to be completed. There were five answer options with one clear, correct answer. Short answer questions asked the students to describe a concept in a few sentences. Quizzes that followed systems simulation interventions contained at least one question that tested students' systemic knowledge about the topic addressed by the simulation. We analyzed data from systems-related questions on Quiz #3 and Quiz #5 only, as these were the quizzes that followed Assignment #2: Human Population Dynamics and Assignment #4: Carbon in the Atmosphere.

Quiz #3 contained four multiple-choice questions, asking students to identify the correct population trend over time, given a birth rate-death rate relationship. Quiz #5 contained one multiple-choice question asking students to identify the correct trend for carbon emissions that would produce and immediate decrease in carbon in the atmosphere if carbon absorption remained constant. We took this question directly from Booth Sweeney and Sterman's study on student misconceptions about climate change (2002).

Assessment questions on assignments

All of the questions included on Assignments #2 and Assignment #4 were evaluated for systemic understanding. Assignment #2, Part 1 questions were evaluated as a pretest measure for all students. We assumed that control section students pretest and posttest scores were identical, since they had no intervention to change their understanding. We only evaluated questions 3-8 on Part 1 of the assignment. These questions had parallel questions on Part 3, so we could compare student understanding before simulation use to their understanding after. Assignment #2, Part 3 questions were evaluated as a posttest measure immediately following the intervention. Assignment #4, Part 3 questions were evaluated for all students as a post-test measure immediately following the intervention. Table 3 shows the number and types of questions were included on each assessment and how many points they were worth.

Table 3. *Assessment Questions, What They Assessed, Point Value*

Assessment	**What the question(s) assessed**	**No. of questions**	**Points**
Baseline Quiz and Final Exam	General knowledge	12	12
	Systems understanding of population dynamics and carbon in the atmosphere	5	13
Assignment #2 Part 1	Systemic understanding of population dynamics	1	1
Assignment #2 Part 3	Systemic understanding of population dynamics	6	18
Quiz #3	Systemic understanding of carbon in the atmosphere	1	1

Assignment #4 Part 3	Systemic understanding of carbon in the atmosphere	5	13
Quiz #5	Systemic understanding of carbon in the atmosphere	1	1

Evaluation

Stave and Hopper (2008) proposed a hierarchy of systems thinking skills based upon Bloom's taxonomy of learning domains. Using this hierarchy, Skaza and Stave (2009) devised a coding scheme aimed at evaluating the systems thinking abilities in the lower levels of this taxonomy. Specifically, our assessment questions attempted to surface students' ability to recognize interconnections, understand stock and flow variables, and understand how these variable interact to produce an increase or decrease in the systems stock.

From our previous study we learned that students express systemic relationships in a number of ways, even when evaluating the most basic systems skills. The current study codes student responses from zero to five. Table 4 shows a complete list of codes, the systems think ability represented and an example of answer that might receive that score.

Short answer questions on all assessments were evaluated using this coding scheme. Multiple-choice questions were given a score of one for a correct answer and a score of zero for an incorrect answer.

Table 4. *Coding scheme for short answer questions*

Code	**Systems thinking skill represented**	**Example answer for the question "If birth rate is decreasing then why is total population increasing?"**
0	No systems thinking skill demonstrated	"Because grownups are more industrialized than babies."
1	Recognizes interconnections based on lecture or text material, but without mention of any system variables	"Because women are educated more."
2	Recognizes interconnections between system variables, but misunderstands variable relationships	"Because you are starting at a higher total population."
3	Demonstrates understanding of one flow connected to the system's stock	"Because death rate has gone down."
4	Demonstrates understanding of both flows connected to the system's stock, but not to each other	"Birth rate is increasing, but death rate is decreasing."
5	Demonstrates understanding of flow relationships to produce an increase or decrease in the stock	"Birth rate is still higher than death rate. When more people are added to the population than taken away, total population increases."

Initial Analysis and Results

Combining the Sections

Mean scores and standard deviations for class assessments show that class size did not affect a student's success. Therefore, we combined the large and small experimental sections and the large and small control sections for analysis. Table 5 shows the means scores and standard deviations for all quizzes and the final exam. The baseline quiz was not included, since we graded it for completion, not correctness.

Table 5. *Mean Scores and Standard Deviations for Quizzes and Final Exam*

Section Number	**N**	**Quiz #2**		**Quiz #3**		**Quiz #4**		**Quiz #5**		**Final Exam**	
		M	SD	M	SD	M	SD	M	SD	M	SD
001	32	76.0	27.8	78.0	14.9	74.6	22.0	81.4	12.2	103.6	11.3
002	65	70.8	24.7	72.9	17.7	70.1	23.6	78.1	13.0	101.4	10.1
003	34	74.7	19.2	75.7	14.7	78.2	19.2	76.4	12.3	99.7	10.3
004	58	78.0	16.6	74.3	18.7	67.2	26.2	79.5	13.0	102.4	8.7

Initial Analyses

For our first analyses, we calculated mean scores and standard deviations for assessments that tested each one of our hypotheses. We expected to that the experimental group would have significantly higher mean scores on each assessment, supporting each subhypothesis.

We calculated mean scores and standard deviations for baseline quiz questions that assessed baseline general knowledge, systemic knowledge, systemic population knowledge and system carbon in the atmosphere knowledge. We assumed that all students were starting the class with the same baseline general knowledge level, systemic knowledge level, systemic understanding of population dynamics and systemic understanding of carbon in the atmosphere. This was important to establish so that all subsequent analyses would be comparable.

We assumed that all sections would demonstrate the same systemic knowledge level on Part 1 of Assignment #2: Human Population Dynamics. Again, it was important to verify a common baseline knowledge level. To verify this assumption, we calculated mean scores and standard deviations for question set on Part 1 of the assignment, prior to simulation use.

We calculated mean scores and standard deviations for final exam questions that assessed baseline general knowledge, systemic knowledge, systemic population knowledge and system knowledge on carbon accumulation in the atmosphere. We expected to see significantly higher scores on each set of assessments for the experimental group.

We also calculated scores for Assignment 2, Part 3, Quiz #3, Assignment 4, Part 3 and Quiz #5 to test students' systemic knowledge of population dynamics and carbon in the atmosphere during the semester. We expected to see significantly higher scores on each assessment for the experimental group.

Table 6 shows the assessments that tested knowledge for each subhypothesis.
Table 6. *Hypothesis, Intervention, Measure and Analysis*

Hypothesis	Intervention	Measure	Analysis
No hypothesis tested; Necessary to establish common baseline knowledge level	None	BGK, BSK, BPop, BCO2	Mean scores and standard deviations
No hypothesis tested; Necessary to establish common baseline knowledge level prior to Assignment #2	None	A2pre	Mean scores and standard deviations
1) Simulation users would perform better on assessments that tested their general knowledge of environmental issues by the end of the course.	Assignment #2: Population Dynamics, Assignment #4: Carbon in the Atmosphere	FGK	Mean scores and standard deviation
2) Simulation users would perform better on assessments that evaluated systems knowledge by the end of the course.	Assignment #2: Population Dynamics, Assignment #4: Carbon in the Atmosphere	FSK	Mean scores and standard deviation
3) At the end of the course, simulation users would demonstrate a greater systemic understanding of the environmental issues addressed by the systems simulations.	Assignment #2: Population Dynamics	FPop	Mean scores and standard deviation
	Assignment #4: Carbon in the Atmosphere	FCO2	

4) Simulation users would show a greater systemic understanding of the environmental issues addressed by the simulations on assessments following the interventions.	Assignment #2: Population Dynamics	A2Post, Q3	Mean scores and standard deviation
	Assignment #4: Carbon in the Atmosphere	A4, Q5	

BGK=Baseline general knowledge, FGK=Final General Knowledge, BSK=Baseline systems knowledge, FSK=Final Systems Knowledge, BPop=Baseline population knowledge, FPop=Final population knowledge, BCO2=Baseline knowledge on carbon in the atmosphere, FCO2=Final knowledge on carbon in the atmosphere, A2Pre=Assignment 2, pre-simulation questions, A2Post=Assignment 2, post simulation questions, A4=Assignment 4, Q3=Quiz 3, Q5=Quiz 5

Initial Analysis Results

Our first subhypothesis stated that simulation users would perform better on assessments that tested their general knowledge of environmental issues by the end of the course. We expected to find that that the experimental group would have significantly higher mean scores for the general knowledge portion of the final. This hypothesis was not supported. The experimental group's scores (M=12.41, SD=2.32) were not significantly higher than the control group's scores (M=12.81, SD=2.25), $t(189)$=1.21, p=.23.

Our second hypothesis stated that simulation users would perform better on assessments that evaluated systems knowledge by the end of the course. We expected to find that the experimental group would have significantly higher scores on the portion of the final exam that tested systemic knowledge. This hypothesis was not supported. The experimental group's scores (M=10.51, SD=2.74) were not significantly higher than the control group's scores (M=10.70, SD=2.51), $t(189)$=.50, p=.62.

Our third subhypothesis stated that, at the end of the course, simulation users would demonstrate a greater systemic understanding of the environmental issues addressed by the systems simulations. We expected to see significantly higher scores for the experimental group on final exam questions that tested both systemic knowledge of population dynamics and carbon accumulation in the atmosphere. This hypothesis was not supported. The experimental group's scores on the population dynamics questions (M=4.25, SD=1.11) were not significantly higher than the control group's scores (M=4.12, SD=1.14), $t(189)$=.77, p=.44. The experimental group's scores on the questions that tested knowledge on carbon accumulation in the atmosphere (M=6.26, SD=2.12) were not significantly higher than the control group's scores (M=6.58, SD=1.79), $t(189)$=1.11, p=.27.

Our fourth subhypothesis stated that simulation users would show a greater systemic understanding of the environmental issues addressed by the simulations on assessments following the interventions. We expected that the experimental group would demonstrate significantly higher scores on Assignment 3, Part 3, Quiz #3, Assignment #4, Part 3, and Quiz #5. There were mixed results for this hypothesis. The experimental group's scores on Assignment #2, Part 3 (M=14.30, SD=2.94) were significantly higher than the control group's scores (M=12.40, SD=3.45), $t(189)$=4.09, p<.01. This result supports our hypothesis. The experimental group's scores on the Quiz #3 questions (M=3.70, SD=.72) were not significantly higher than the control group's scores (M=3.48, SD=.89), $t\ (189)$=1.78, p=.08. This does not support the hypothesis. The experimental group's scores on Assignment #4, Part 3 (M=10.54, SD=2.42) were not significantly higher than the control group's scores (M=10.11, SD=2.89), $t(189)$=1.03, p=.30. This did not support the hypothesis. The experimental group's scores on the carbon in the atmosphere question on Quiz #5 (M=.49, SD=.50) were not significantly higher than the control group's scores (M=.41, SD=.50), $t\ (189)$=1.06, p=.29.

Table 7 shows mean scores, standard deviations, t-values and p-values for all assessments.

Table 7. *Mean Scores, Standard Deviations, t-values, and p-values*

Assessment	Experimental Group		Control Group			
N	92		97			
	M (points)	SD (points)	M (points)	SD (points)	t	p
BGK	9.04	3.09	9.78	2.83	1.72	.09
BSK	6.61	3.30	7.21	3.26	1.25	.21
BPop	2.34	1.70	2.70	1.67	1.49	.14
BCO2	4.27	2.45	4.51	2.45	.65	.51
FGK	12.41	2.32	12.81	2.25	1.21	.23
FSK	10.51	2.74	10.70	2.51	.50	.62
FPop	4.25	1.11	4.12	1.14	.77	.44
FCO2	6.26	2.12	6.58	1.79	1.11	.27
A2pre	12.12	3.21	12.40	3.42	.56	.58
A2post	14.30	2.94	12.40	3.45	4.09	p<.01
Q3	3.70	.72	3.48	.89	1.78	.08
A4	10.54	2.42	10.11	2.89	1.03	.30
Q5	.49	.50	.41	.50	1.06	.29

BGK=Baseline general knowledge, FGK=Final General Knowledge, BSK=Baseline systems knowledge, FSK=Final Systems Knowledge, BPop=Baseline population knowledge, FPop=Final population knowledge, BCO2=Baseline knowledge on carbon in the atmosphere, FCO2=Final knowledge on carbon in the atmosphere, A2pre=Assignment 2, pre simulation questions, A2Post=Assignment 2, post simulation questions, A4=Assignment 4, Q3=Quiz 3, Q5=Quiz 5

Regression Analysis and Results

Multiple Regression Analyses

When we didn't find significant differences in the experimental and control group's knowledge levels on the final exam assessments, we used multiple regression analyses to take a more detailed look at the relationship between students' performance on assessments and simulation use. Multiple regression analysis allows us to consider the effect of more than one variable on a dependent variable, enabling better explanations for the value of the dependent variable (Anderson, Sweeney & Williams, 2007). We used this method to test for a relationship between general and systemic understanding on a number of analyses and simulation use.

We formed new hypotheses, based on what we expected to see in the regression results:

1) Regression results would show a positive relationship between performance on questions that tested students' general knowledge of environmental issues by the end of the course and simulation use.
2) Regression results would show a positive relationship between performance on questions that tested that evaluated systems knowledge by the end of the course and simulation use.
3) Regression results would show a positive relationship between performance on questions that tested students' knowledge on the subjects addressed by the systems simulations and simulation use.
4) Regression results would show a positive relationship between performance on questions that tested students' knowledge on the subjects addressed by the simulations on assessments following the interventions and simulation use.

Our first subhypothesis was that regression results would show a positive relationship between performance on questions that tested students' general knowledge of environmental issues by the end of the course and simulation use. We tested this hypothesis by using a multiple regression model to model Final General Knowledge as a function of Baseline General Knowledge and simulation use.

$$FGK = b_0 + bBGK + bSIM$$

Our second hypothesis stated that regression results would show a positive relationship between performance on questions that tested that evaluated systems knowledge by the end of the course and simulation use. We tested this hypothesis by using a multiple regression model to model Final Systemic Knowledge as a function of Baseline Systemic Knowledge and simulation use.

$$FSK = b_0 + bBSK + bSIM$$

Our third hypothesis stated that regression results would show a positive relationship between performance on questions that tested students' knowledge on the subjects addressed by the systems simulations and simulation use. We tested this hypothesis by using a multiple regression model to model:

1) Final Population Knowledge as a function of Baseline Systemic Knowledge and simulation use.

$$FPop = b_0 + bBPop + bSIM$$

2) Final Carbon Accumulation Knowledge as a function of Baseline Carbon Accumulation Knowledge and simulation use.

$$FCO2 = b_0 + bBCO2 + bSIM$$

Our fourth hypothesis stated that regression results would show a positive relationship between performance on questions that tested students' knowledge on the subjects addressed by the simulations on assessments following the interventions and simulation use. We tested this hypothesis by using a multiple regression model to model:

1) Assignment #2, Part 3 scores as a function of baseline population knowledge and simulation use.

$$A2Post = b_0 + bBPop + bSIM$$

Control group scores on Part 3 were assumed to be the same as their scores on Part 1 of the assignment, as they had no intervention to cause a change in understanding. Since pre-simulation scores and post-simulation scores were the same for this group, we did not use the pre-intervention score as a variable for baseline knowledge in the regression analysis.

2) Student performance on Quiz #3 as a function of baseline population knowledge and simulation use.

$$Q3 = b_0 + bBPop + bSIM$$

3) Assignment #4, Part 3 scores (A4post) as a function of baseline systemic knowledge about carbon in the atmosphere (BCO2) and simulation use (SIM).

$$A4 = b_0 + bCO2 + bSIM$$

4) Quiz #5 performance as a function of baseline knowledge and simulation use.

$$Q5 = b_0 + bCO2 + bSIM$$

Table 8. *Hypothesis, Intervention, Measure and Regression Model*

Hypothesis	Intervention	Measure	Model
1) Regression results would show a positive relationship between performance on questions that tested students' general knowledge of environmental issues by the end of the course and simulation use.	Assignment #2: Population Dynamics, Assignment #4: Carbon in the Atmosphere	FGK	FGK = b_0 + bBGK + bSIM
2) Regression results would show a positive relationship between performance on questions that tested that evaluated systems knowledge by the end of the course and simulation use.	Assignment #2: Population Dynamics, Assignment #4: Carbon in the Atmosphere	FSK	FSK = b_0 + bBSK + bSIM
3) Regression results would show a positive relationship between performance on questions that tested students' knowledge on the subjects addressed by the systems simulations and simulation use.	Assignment #2: Population Dynamics	FPop	FPop = b_0 + bBPop + bSIM
	Assignment #4: Carbon in the Atmosphere	FCO2	FCO2 = b_0 + bBCO2 + bSIM
4) Regression results would show a positive relationship between performance on questions that tested students' knowledge on the subjects addressed by the simulations on assessments following the interventions and simulation use.	Assignment #2: Population Dynamics	A2Post Q3	A2Post = b_0 + bBPop + bSIM Q3 = b_0 + bBPop + bSIM
	Assignment #4: Carbon in the Atmosphere	A4, Q5	A4 = b_0 + bCO2 + bSIM Q5 = b_0 + bCO2 + bSIM

BGK=Baseline general knowledge, FGK=Final General Knowledge, BSK=Baseline systems knowledge, FSK=Final Systems Knowledge, BPop=Baseline population knowledge, FPop=Final population knowledge, BCO2=Baseline knowledge on carbon in the atmosphere, FCO2=Final knowledge on carbon in the atmosphere, A2Post=Assignment 2, post simulation questions, A4=Assignment 4, Q3=Quiz 3, Q5=Quiz 5

Multiple Regression Analysis Results

Our first hypothesis was not supported. There was no significant relationship between performance on the final general knowledge questions and simulation use, $\beta=-0.17$, $t(189)=-0.56$, $p=0.58$. In this case, baseline general knowledge was the predictor of final general knowledge, $\beta=0.32$, $t(189)=6.18$, $p<0.01$.

Our second hypothesis was not supported. There was no significant relationship between students' final systemic knowledge level at the end of the course and simulation use, $\beta=-0.04$, $t(189)=-0.12$, $p=0.91$. Again, baseline systemic knowledge was the most significant predictor of final systemic performance, $\beta=0.25$, $t(189)=4.49$, $p<0.01$.

Our third hypothesis was not supported. There was no significant relationship between systemic understanding of population dynamics at the end of the course and simulation use, $\beta=0.18$, $t(189)=1.14$, $p=0.25$. In this case, baseline systemic knowledge had a significant impact on final systemic understanding of population dynamics, $\beta=0.14$, $t(189)=3.02$, $p<0.01$.

There was no significant relationship between systemic understanding of carbon accumulation at the end of the course and simulation use, $\beta=-0.27$, $t(189)=-01.00$, $p=0.32$. In this case, baseline systemic knowledge of carbon accumulation was the main predictor of final exam performance on the carbon accumulation questions, $\beta=0.22$, $t(189)=3.89$, $p<0.01$.

Analyses that tested our third hypothesis showed mixed results. Multiple regression model results showed a significant positive relationship between scores on Assignment #2, Part 3, post-intervention and simulation use, $\beta=2.08$, $t(189)=4.64$, $p<0.01$. There was also a significant positive relationship between post-intervention assessment scores and baseline systems knowledge on population dynamics, $\beta=0.50$, $t(189)=3.74$, $p<0.01$.

There was a significant positive relationship between students performance on Quiz #3 and simulation use, $\beta=0.24$, $t(189)=2.01$, $p<0.05$. Performance on Quiz #3 was also significantly correlated with baseline population knowledge, $\beta=0.08$, $t(189)=2.24$, $p<0.05$.

Performance on Assignment 4, Part 3 was not significantly related to simulation use, $\beta=0.63$, $t(189)=1.56$, $p=0.12$. In this case, baseline systemic knowledge of carbon accumulation was a significant predictor of success on the assessment questions, $\beta=0.39$, $t(189)=4.62$, $p<0.01$.

Students performance on Quiz 5 was not significantly related to simulation use, $\beta=0.09$, $t(189)=1.33$, $p=0.19$. Performance on Quiz 5 was significantly related to baseline systemic knowledge about carbon accumulation (BCO2), $\beta=0.05$, $t(189)=3.35$, $p<0.01$.

Multiple regression results an all analyses are shown in Table 9.

Table 9. *Multiple Regression Results*

		Final general knowledge	Final systems knowledge	Final population knowledge	Final CO2 knowledge	Assignment 2 post intervention	Quiz 3:Populaiton Dynamics questions	Assignment 4 post intervention	Quiz 5: CO2 Question
Variable									
SIM		-0.17	-0.04	0.18	-0.27	2.09*	0.24*	0.53	0.09
BGK		0.32*	-	-	-	-	-	-	-
BSK		-	0.25*	-	-	-	-	-	-
Bpop		-	-	0.14*	-	0.50*	0.08	-	-
BCO2		-	-	-	0.22*	-	-	0.41*	0.05*
Bgraph		-	-	-	-	-	-	-	-
A2Pre		-	-	-	-	-	-	-	-
A2Post		-	-		-	-		-	-
A4Post		-	-	-	-	-	-	-	-
Fgraph		-	-	-	-	-	-	-	-
					-	-	-	-	-
Intercept		9.71	8.91	3.73	5.61	11.05	3.28	8.27	0.19
r square		0.18	0.10	0.05	0.08	0.15	0.04	0.11	0.06
Adjusted r square		0.17	0.09	0.04	0.07	0.14	0.03	0.10	0.05
	*p<.01								

Discussion

We found that scores were significantly better for simulation users immediately after using the simulations, but not later on in the semester. Two possible explanations are:

1) Students in the experimental sections may have lost the systemic knowledge that they gained through simulation use and that they displayed on assessment immediately following simulation use.
2) Students in the control sections increased their systemic understanding through a number of other class activities.

If the experimental group lost the systemic knowledge that they demonstrated on the Assignment #2 and on Quiz #3, then we would expect to see lower scores on the final systemic knowledge assessment than on Assignment #2 or Quiz #3. To test this, we compared scores on short answer population questions on the baseline quiz (BPop), Assignment #2 (A2) and the final exam (FPop). Figure 3 shows both groups' change in systemic understanding over the course of the semester. The experimental group showed an increase in systemic understanding between Assignment #2 and the final exam, confirming that students in the experimental sections retained the systemic knowledge they gained through simulation use.

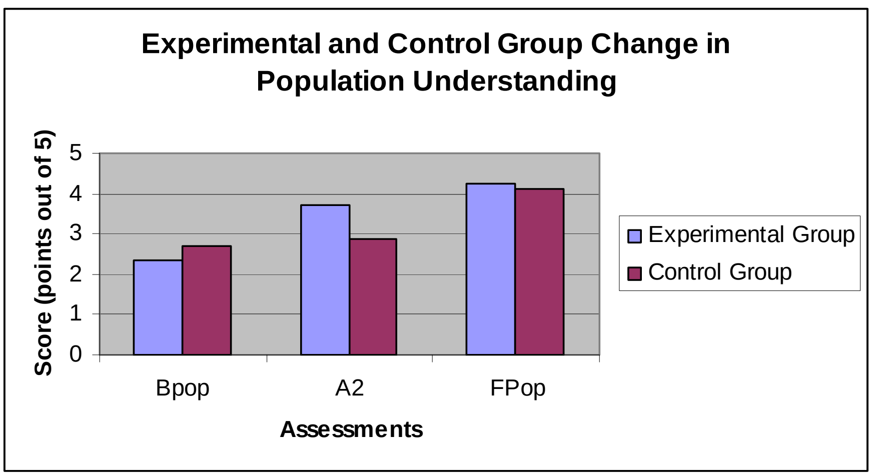

Figure 3. Change in Systemic Understanding of Population Dynamics

Figure 3 also shows that the control group's scores increased between Assignment #2 and the end of the semester. Another possible explanation for the lack of difference in systemic understanding by the end of the course is that the control group's systemic understanding could have increased. We believe that the control group showed an increase in their systemic knowledge due to an emphasis on systems principle throughout the course.

Course material was presented in lecture with systems thinking principles in mind. The course textbook emphasized interconnections between the human and natural world. Each day, class lectures began with the graphic shown in Figure 4a, which was intended to reinforce the idea that the human/environment relationship is one of reciprocal feedback. Lecturers then highlighted how that relationship was present in the topic they were lecturing on that day. Figures 4b shows how the graphic was presented for the fossil fuels lecture.

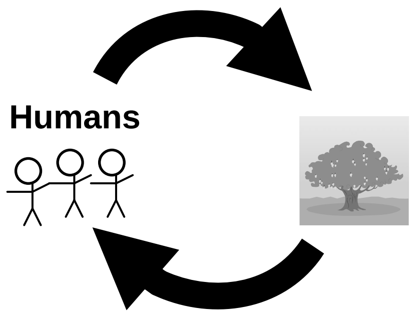

Figure 4a. Human/Environment Relationship

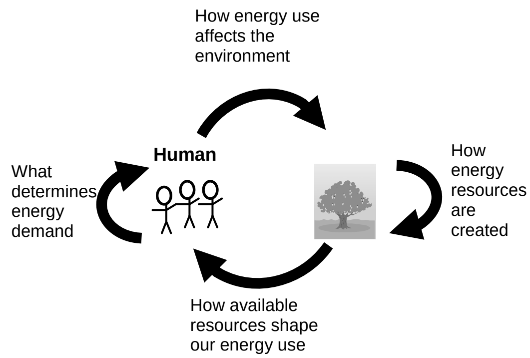

Figure 4b. Human/Environment Relationship within Fossil Fuels

Course lectures and the textbook emphasized system connections, feedback and dynamic behavior. We believe the reason we did not see a greater difference in the performance of the two groups was largely due to the overall emphasis on systems principles throughout the course for both groups of students. We delivered this message to both the experimental sections and control sections consistently throughout the course.

Another important part of class assignments was the debriefing that we conducted for all students after they turned the assignment in. All students who came to lecture that day would have heard the debriefing. Instructors read each question in the assignment and asked students to respond. Students called out answers. If no answer was called out, instructors encouraged student response with hints. If no answer was given, instructors gave students the correct answer to the question. Although student response drove each debriefing session, any misconceptions about the systems principles involved in the assignment were corrected. Both the experimental and the control sections received the debriefing.

Although we are pleased that the systemic understanding of all students appears to have increased, we did expect that the simulations would have had a greater effect. Why don't we see more of an increase in systemic understanding for the experimental group over the control group? Why didn't the systems simulations have more of an impact on student understanding than the other course materials?

Intervention issues

We had several restrictions for this study. Students would complete the assignments on their own, without guidance from an instructor. We assumed that they had no previous environmental science education. We assumed they had no experience in a simulation environment. Part of our challenge was being able to design effective systems simulation interventions for a large, lecture-based course.

Lack of Guidance

Students worked with the simulations without live instruction or guidance, because we did not have classroom computers available for the number of students that we had. Students were given written instructions and descriptions of the system the simulation was modeled after. Students using the simulation had only a surface interaction with it. There was no instructor present to encourage them to think about what sort of interactions were taking place within the system to produce the trend they saw on the screen. As a result, student often explained stock and flow interactions in terms of variables that were not represented in the system they were working with. For example, if a student was asked, "What causes carbon to accumulate in the atmosphere?" a student might answer, "Too much industry." If they were asked, "Why is total population increasing even though birth rate is decreasing?" they might answer, "Because this population has more medicine available." Because there was not enough support during the exercise, students tended to rely heavily on knowledge that had acquired from other course materials.

Assignment Design

Sawicka (2005) discusses the role of a learner's cognitive capacity in using a systems simulation. She argues that when the working memory is primarily devoted to extraneous information in simulation design (i.e. interface operation), the less 'surplus' working memory there is to develop an understanding of the underlying system. Each simulation that we used was different in presentation and what it asked the user to do. Every time the student interacted with a new simulation, they had to interpret a new interface, understand new subject matter, understand the task and come up with a problem-solving strategy. This decreased the potential for students to 'get better' at simulation use and focus on the lesson it was trying to teach. During each debriefing session, students expressed frustrations about accessing the simulation, interaction with the interface and understanding the goal of simulation use. If the simulations we used were more similar in these areas, the students could have been better able to understand the subject matter within the simulation (i.e. human population dynamics or carbon accumulation in the atmosphere).

Assessment issues

If the interventions had been perfectly designed to facilitate students' systemic understanding, we still may not have seen the difference that we expected to see between the experimental and the control groups. Carefully designed assessments allow students to demonstrate their change in understanding. To improve our understanding of students' change in systemic knowledge, more assessment techniques should be tested.

More Assessment Methods

We saw the most significant relationship between simulation use and systemic understanding on the assessment questions for Assignment #2. These questions asked students to identify a trend over time for a given birth rate/death rate condition and explain why they chose the trend that they did. We asked students to express their understanding in more than one way. We should have done this for other assessments as well.

Part of what we wanted to test was how to best assess systemic understanding. However, we only used two assessment techniques: multiple choice questions and short-answer explanations of system characteristics. Systems dynamicists use causal maps and stock and flow diagrams to express stock and flow relationships. Future studies should use these representations to assess systemic understanding. While students may have been unable to create a stock and flow or causal loop diagram, it is reasonable to assume that they could have completed a partially-created diagram in with the appropriate variables. In our next steps, we will test more assessment techniques and use several when assessing understanding of even one interconnection to get a sense for what a student really knows.

The need for more rigorous ways to test students' understanding is emphasized here. What we are really striving for is a way to get an accurate picture of what the student is thinking of when they are initially presented with these complex environmental systems, how that changes as the student uses the simulation and what is looks like after they have completed the simulation activity. The best way to do this may be to be with the student as they complete the activity to ask things like, "Why did you choose what you did?" or "What do you think is producing that behavior?" One shortcoming in the research setup is that there was not access to the students as they were completing the assignments. Evaluation was restricted to the words they put on the page. A student may understand fully that population decline is the result of a lower number of births than deaths, but states its cause as being a higher education level for women, they really have a more systemic understanding than they have demonstrated, but the evaluator has to rate the response that is given. If the researchers were with the student they could probe for the understanding that doesn't immediately come out of paper. The authors are currently considering how interview techniques may be incorporated into future data collection.

Cheek (1992) discusses the need for the advancement of assessment tools parallel to the advancement in instructional techniques in science education. One method he describes for evaluating student understanding is evaluating student performance of the task. This would involve observing the student as they complete the task. While this may not have been possible in the context of this study, we could have incorporated assessment questions that asked the student what they did when working with the simulation. This would have given us more data on the students' experience with the simulation. Combining this information with their performance on systemic understanding questions would have led to a better understanding about what parts of the simulation were effective in increasing systemic understanding.

Assessment as a Teaching Tool

A qualitative review of student responses in Assignment #2 and Assignment #4 showed that students' answers improved from the beginning of the assignment to the end. For both assignments we started with simple questions that asked students to describe the relationship between two variables in the system. The last questions of the question set asked them to relate both flows and the stock in the system. It is possible that students learned how to put the variables together by working their way through the questions. This is problematic if we are trying to assess their change in understanding as a result of simulation use only, although it does present an interesting way to increase the effectiveness of simulation use.

Conclusions and Recommendation for Further Study

This study furthers Stave and Hopper's (2008) work by implementing interventions and assessment based on the Taxonomy of Systems Thinking Characteristics. It begins the work of revising and verifying the taxonomy through controlled, experimental research. Future studies should address the assessment and intervention deficiencies described in this paper. Interventions need to be revised to include a higher level of interaction with simulation. We should expect the students to learn more about the system underlying the simulation to have a richer understanding of what the simulation is designed to teach.

We need to devise new ways for assessing student systemic understanding. New assessment methods should ask students to express their mental models in a number of ways: verbally, graphically, in a diagram, etc. Future studies should test assessment techniques for their effectiveness in making student thinking visible, while they are testing the effectiveness of the systems simulation intervention.

We asked the question "Does the use of systems simulations in an introductory environmental science course increase students' systemic understanding of environmental issues?" We found support for the use of systems simulations in the environmental science classroom. We also found the need for more rigorous assessment methods and better interventions design. Large, introductory courses like the one in this study present several challenges in designing and implementing a systems simulation lesson, but hey also provide a great opportunity for increasing systemic understanding of environmental issues.

References

Booth Sweeney,L. & Sterman, J.D. (2000). Bathtub dynamics: initial results of a systems thinking inventory. *System Dynamics Review*, 16(4):249–286.

Booth Sweeney,L. & Sterman, J.D. (2007). Thinking about systems: student and teacher conceptions of natural and social systems. *System Dynamics Review*, 23(4):285–312.

Doyle, J.K., Radzicki, M.J. & and Trees, W.S. (July, 1998). *Measuring Changes in Mental Models of Dynamic Systems: An Exploratory Study*. Paper presented at the 16th International Conference of the System Dynamics Society, Quebec City, Canada.

Evagorou, M., Korfiatis, K., Nicolau, C., & Constantinou, C. (2008). An Investigation of the Potential of Interactive Simulations for Developing System Thinking Skills in Elementary School: A case study with fifth-graders and sixth-graders. *International Journal of Science Education*, 2008,1-20.

Fisher, D.M. (2008, July). *Building Slightly More Complex Models: Calculators vs. STELLA*. Paper presented at the 26th International Conference of the System Dynamics Society, Athens, Greece.

Grant, W.E. (1998). Ecology and natural resource management: reflections from a systems perspective. *Ecological Modelling*: 108: 67–76.

Korfiatis, K., E. Papatheodorou, E., Stamou, G. P. (1999). An investigation of the effectiveness of computer simulation programs as tutorial tools for teaching population ecology at university. *International Journal of Science Education*, 21(12), 1269-1280.

Hogan, Kathleen. (2000). Assessing students systems reasoning in ecology. Journal of Biological Education, 35(1), 22-28.

Hopper, M. & Stave, K. A. (2008, July). Assessing the Effectiveness of Systems Thinking Interventions in the Classroom. Paper presented at the 26th International Conference of the System Dynamics Society, Athens, Greece.

Maani, K.E. & Maharaj, V. (2004). Links between systems thinking and complex decision making. *System Dynamics Review,* 20(1): 21-48.

Sawicka, A. & Kopainsky, B. (2008). Reindeer herd management model.

Stave, K. &Hopper, M. (2007). *What Constitutes Systems Thinking? A review of practitioner views.* Paper presented at the 25th International Conference of the System Dynamics Society, Boston, MA.

Sterman, J.D & Booth Sweeney,L. (2002). Cloudy skies: assessing public understanding of global warming. *System Dynamics Review*, 18(2): 207–240.

Wheat Jr., I.D. (2007). The feedback method of teaching macroeconomics: is it effective? *System Dynamics Review*, 23(4), 391-413.