Assessing System Dynamics Curricula: Past, Present, and Future

Susan L. Ganter, James K. Doyle, and Michael J. Radzicki
Worcester Polytechnic Institute
100 Institute Road
Worcester, MA 01609-2280 USA
email: sganter@wpi.edu

Abstract

Many claims have been made for the efficacy of the growing number of curricula utilizing system dynamics. However, although there is educational research that will likely support these claims, systematic efforts to document them thus far have been inadequate or non-existent. This paper will review and critique one such effort to assess the effectiveness of system dynamics in the classroom. The compelling need for improved SD assessment procedures, as well as recommendations for future projects in SD assessment, will then be discussed.

Introduction

Until the mid-1970s, system dynamics was used mostly as a research tool and was taught primarily in graduate-level management and systems analysis courses. In 1975, however, Roberts (1978) began to experiment with teaching system dynamics to fifth and sixth graders. In collaboration with the MIT System Dynamics group, Roberts (1981) proceeded to develop and test a high school curriculum in system dynamics.

Since these initial efforts, interest in employing system dynamics as a pedagogical tool in reforming curricula in grades K-12 has increased substantially. In 1994, for example, more than 260 people from throughout the United States and other countries attended a conference on Systems Thinking/Dynamic Modeling for K-12 Education (Forrester 1994). The purpose of this paper is to describe and critique the current state of efforts to assess the efficacy of the growing number of system dynamics curricula; to outline methods for improving assessment of these curricula; and to suggest directions for future research.

The Typical System Dynamics Curriculum

System dynamics (SD) curricula focus on modeling the behavior of complex systems over time in order to solve problems and improve understanding. Typically, students are first introduced to important system dynamics concepts such as stocks and flows, feedback loops, and system nonlinearities. Then, students receive instruction in using a dynamic computer simulation software package. Finally, students are given problems to solve and are guided through the stages of building and running a dynamic simulation model of the problem. Ideally, the standard course material is not presented entirely in lecture form, but is accessed by the students as necessary during model building and analysis.

The main difference between SD and more traditional curricula is the emphasis on the time dimension – SD programs focus on dynamic models. However, it is important to note that this innovative curriculum differs from traditional curricula in several other important ways:

1. SD curricula focus on problem solving rather than accumulating factual knowledge;

2. The main activity in SD classes is interaction with computers and simulation modeling software;

3. SD curricula emphasize graphics and visual displays of information rather than verbal descriptions;

4. Many SD teachers have students work together on problems in small groups rather than individually;

5. Many SD curricula incorporate "learner-directed learning", in which students share responsibility for the pace and direction of the educational process; and,

6. Having students focus on building models means that SD curricula introduce the process of "synthesizing" knowledge at a much earlier stage than traditional programs.

Obviously, SD curricula represent a dramatic departure from traditional programs, revamping several key features of the educational process simultaneously.

Claims for the Efficacy of System Dynamics Curricula

Given the novelty and complexity of implementing SD curricula, it is no surprise that little formal assessment of their effectiveness has been carried out to date. However, there is no shortage of claims that have been made for the superiority of SD programs, based largely on anecdotal information and the informal reports of teachers. The following claims, for example, have been reported by Forrester (1993; 1994):

1. SD motivates students to work harder. Draper (1989), for example, claims that SD students learn more, learn more quickly, do extra work, and exhibit fewer disciplinary problems.

2. The work environment of SD programs is superior because it creates a more exciting, challenging, and collaborative atmosphere that takes advantage of students' intrinsic curiosity and increases students' confidence when dealing with complex problems.

3. The SD approach increases the retention of knowledge by providing students an organizing framework that makes the material more meaningful and shows how seemingly disparate academic disciplines are related.

4. The focus of SD curricula on solving problems through modeling forces students to think in a structured way, making them better "critical thinkers".

5. Learning SD will facilitate the transfer of knowledge and insights across academic boundaries. In addition, insights gained during computer modeling of systems will transfer to everyday problem solving and decision making.

6. An SD-based education will produce more flexible, general problem solvers who can more easily move from field to field.

Apparently many teachers and SD professionals feel that the benefits of an SD education are obvious, the results of current SD curricula have been dramatic and positive, and there is no need to reserve judgment while waiting for the results from a formal assessment. However, as we will show, the need for assessment is great, as assessment efforts to date have been inadequate or non-existent.

A Review of Classroom Dynamics: A Report on the Assessment of SD

The only large-scale effort to assess SD curricula to date is a study by Mandinach and Cline (1994) of the Educational Testing Service (ETS) that follows the Systems Thinking and Curriculum Innovation (STACI) project from 1986 to 1992. The evaluation of the project's effectiveness over this time period is restricted to issues of implementation, with the exception of limited analyses of classroom observations, participant interviews, and student exams. The authors emphasize that it is difficult, if not impossible, to evaluate the effect of a new teaching strategy on learning without first examining the process and success of the implementation of the project's main objectives. Although the original intent of the authors was to observe the change in student learning and organization of school systems initiated by the project, the study was reduced to simply analyzing the process of change and its perceived effect on classroom dynamics.

The Evolution of the STACI Project

The STACI project began as a two-year research effort in 1986 headed by the ETS with funding from the Office of Educational Research and Improvement (OERI). The initial purpose of the project was to examine the cognitive demands and consequences of learning with system dynamics and STELLA at the secondary level as observed at one high school in Vermont. Topics included in the study were biology, chemistry, physics, history, and general physical science. Data to be used in the analysis were collected from course exams and grades, standardized tests, observations, and interviews. In addition, a special instrument developed by the ETS staff called the Systems Thinking Instrument (STI) was administered to the students in the courses using system dynamics.

It is evident that the information to be obtained from such a research design was not clearly established prior to the implementation of the study. For example, the authors fail to recognize the inherent weaknesses and built-in biases that accompany the use of course exams and grades to measure student learning. No standards of what it means to understand system dynamics or the content specific to the courses were imposed by the study. The STI represents an attempt to incorporate such a standardization; however, it was only administered to the students in the courses using system dynamics at the end of these courses, with no basis for comparison to either a control group or a similar test administered before the system dynamics experience. Methods for collecting data through observations and interviews are also not outlined, including frequency and content of these efforts. In addition, the authors realized too late that the implementation of the fundamental ideas of system dynamics in the courses studied was very limited. Principles of SD were only being used sporadically on a few topics. Therefore, the outcome of this two-year project was a brief outline of the effects of system dynamics in one secondary school based on very unreliable tests and last minute attempts to salvage some information through anecdotal summaries of the project team's observations.

The result of this initial research breakdown was a complete restructuring of the project, with the new focus on the issues associated with the implementation of system dynamics at the K-12 level. To better understand the transferability of SD curricula, the project created a network of eight schools around the country consisting of seven secondary schools and one middle school with a wide variety of characteristics. Donation of computer equipment for the schools was secured from Apple Computer and this second phase of the project began with a two-week workshop in the summer of 1988 for the 32 participating teachers.

Phase 2: The STACI^N Project

The STACI^N Project (N for "network" of schools formed in 1988) proposed to focus on two major components: teacher support and research activities. Teacher support consisted of workshops, an electronic mail address for networking between campuses, meetings within disciplines and schools, and stipends for the participating teachers ($3,000 per year). Unfortunately, the project did not use any funds to secure consultation from experts in curriculum development or system dynamics. The participating teachers, most of whom were novices in SD and had no experience developing curriculum for alternative teaching methods that utilized technology, were expected to develop (mostly from scratch) all SD materials to be used in their classes. They were then expected to implement these materials with support only from the other participating teachers. Needless to say, the quality of the SD materials developed and the degree of implementation varied immensely from class to class, as it had in Phase 1 of the project. In addition, no provisions were made for the upgrade of the technological hardware and software at the schools. Consequently, many machines were outdated or out of service after the first two years of the "new and improved" STACI^N project. No commitments for continued funding or support of the project had been secured from the participating schools; therefore, the maintenance of the technology and, ultimately, the funds needed to continue the project were non-existent.

Research for the STACI^N portion of the project intended to encompass three areas: student learning, impact on teacher behavior, and impact on schools. Student learning was to be assessed through interviews and longitudinal testing; impact on teacher behavior, including course organization and classroom activity, was to be measured through classroom observations and interviews; and impact on the schools was to be assessed through repeated interviews with students, teachers, administrators, parents, and school board members. Interviews in all three areas of the research utilized the same questions in order to obtain different perspectives on the same issues. The research design took the form of a *formative experiment* as developed by Newman (1990); the authors ruled out the use of a quasi-experimental design, as they felt it would be too difficult to incorporate and adequately maintain control and experimental groups that would yield reliable results. They supported their decision with work by Salomon (1990; 1991) which cautions against any attempt to make conclusions about the effect of school-based curriculum innovation projects due to the complex nature of educational organizations.

Although the development of a quasi-experimental design that is reliable in a school setting is difficult at best, much information can be obtained through a well-controlled effort that collects data of this type, especially over a period as long as that of the STACI project. Rather than attempt to do this, however, the authors used this difficulty as an excuse for sloppily collected data and poor administration of the project. Even after discussing the impossibility of creating such an experiment, the authors still collected data on student learning based on a quasi-experimental methodology and then blamed the unreliable results (which they fail to even report) on the claim that such a design was not appropriate. However, the student learning data collected were not comprehensive even for an ideal environment. For example, performance on course content was measured with quizzes, tests, laboratory reports, and examinations that were administered by the individual teachers and were not controlled in any systematic manner by the project team. Such data are biased and cannot be used to measure learning within disciplines or schools because of inherent differences between teachers. Interviews, observations, and classroom videotapes were also not systematically monitored, with no formal mechanisms used to control and record this quantitative data. Again, no results are reported. Finally, students in the classes using system dynamics were administered a standardized test on systems thinking. However, results could only be compared to a general, non-SD standardized test previously administered, since no provisions were made for a pre-test/post-test design or comparisons with a control group.

The project did attempt to involve the participating teachers in the collection of information for research purposes. This information included a paper-and-pencil student test of SD concepts, student essays requiring SD, peer teacher observations, and teacher-designed alternative assessment techniques. However, as with the previously described data collection, it is unclear how the collection of this information was monitored by the project team or even whether all classes involved in the project were participating. Again, it is critical in this type of research design that all data collection be closely monitored and standardized and that any schools to be observed, including a control group of schools not using SD, are required to fully participate and cooperate in such data collection.

Outcomes of the STACI Project

Because of the non-standardized format of the data collection on student learning and organizational changes, the only information that could realistically be reported by the authors was Observations, Issues, Cautions, and Lessons Learned (see Mandinach and Cline 1994, Chapter 4). The issues encountered in the study fell into several categories: school and district administration, physical resources, curriculum, teachers, professional support, project management, and research design. Although this type of information is critical to the further development of any educational innovation, it is disappointing that there are not more objective results. The six-year (and, ultimately, eight-year) span of the project could have produced a rich array of data on student learning, curriculum development, and organizational change that would be of great value to the SD community.

Evidence from the Psychological and Educational Literature

Given the poor quality of previous efforts to assess SD curricula, it is necessary to examine the more general literature on learning and education in order to evaluate claims for its effectiveness. According to the literature, the focus of SD curricula on increasing motivation is appropriate: increasing classroom morale is one of the most effective ways to increase learning (Walberg 1986). In addition, teaching strategies that result in relatively relaxed learning environments similar to the SD approach have been shown to improve academic performance (Good 1983; Walberg 1986). There is also substantial evidence that cooperative group learning promotes academic success (Walberg 1986). However, working effectively in groups does not come naturally to most students; they need to be taught such skills as how to share responsibility and how to encourage their group partners (Gallagher 1994).

Evidence concerning the effectiveness of learner-directed learning over and above the effects of cooperative learning is more difficult to find. At least one study, however, has noted that gifted students are more effective practitioners of the monitoring skills necessary when learning is self-regulated (e.g., organizing information, seeking peer assistance; see Zimmerman and Martinez-Pons 1990). SD curricula may be well-advised to include direct instruction in self-regulation techniques.

The focus of SD programs on building dynamic computer models to solve problems is precisely the type of task which is known to increase student motivation. However, there are some issues of concern for teaching problem solving skills through system dynamics. First, SD curricula tend to emphasize the creation of new mental models for problems. Research suggests, however, that if students hold pre-conceived mental models that are incorrect, these misconceptions can be very persistent and may interfere with new learning unless they are addressed directly (Brown 1992). Second, by introducing synthesis at an early stage during the learning process, SD curricula are essentially asking students to solve problems using a limited knowledge base. Research on problem solving, however, suggests that many novice errors are due simply to the lack of important information (Chi et al. 1988).

The claims of SD practitioners that employing a dynamic model to help organize information will improve retention and understanding is well-established in psychological theory (Bruner 1963). Of course, almost any meaningful way of organizing information will improve retention, and it is not known whether stock-flow diagrams, for example, are more effective than other representations. Mayer (1989), however, has suggested that diagrams designed to help learners build conceptual models of systems may be an especially promising technique for increasing understanding.

The psychological literature on the transfer of knowledge to new problems in different academic disciplines is not encouraging. Beginning with Gick and Holyoak (1980), many studies have demonstrated that surprisingly few people are able to transfer problem solving strategies to analogous situations, even under optimal conditions; they simply fail in most cases to perceive the analogy. There is also a substantial danger that students may take analogies too far or mistakenly apply them when it is not appropriate (Spiro et al. 1989). Perkins and Salomon (1989), however, are more optimistic; although they acknowledge that transfer is difficult, they report that transfer can be improved by teaching students general heuristic strategies.

The claim that SD curricula will teach students concepts that are transferable to an intuitive understanding of problems in everyday life outside the classroom seems overly optimistic. Certainly students may think about applying SD to new problems later in life, but if computer simulation is necessary to understand complex systems in the classroom, it will be necessary to understand them outside the classroom as well. Training in SD will not alter the cognitive limitations that prevent people from thinking clearly about complex systems. Sterman (1989; 1994), for example, reports that people have great difficulty understanding even the "simplest" of dynamic feedback systems without the aid of computer simulation.

Overall, there is strong support in the literature for the potential effectiveness of many elements of SD curricula (with the notable exception of knowledge transfer). However, there is no direct evidence that the SD "perspective" of dynamic modeling has beneficial effects over and above the proven beneficial effects of other aspects of SD curricula, such as cooperative learning and problem solving orientation.

The Need for Formal Assessment of System Dynamics Curricula

Given the strength of opinions held by SD proponents, the large amount of anecdotal evidence, and the generally promising support in the educational literature, it is tempting to say that formal assessment of SD curricula should not be a priority. However, even when results seem dramatic and obvious, there are good reasons to conduct formal assessments:

- The lack of evidence from carefully controlled assessment studies may limit the rate at which SD curricula become accepted by new schools. In particular, a willingness to subject ideas to rigorous testing is an important element in gaining the trust of parents, teachers, and administrators.

- In order to build a convincing case for the introduction of SD curricula in place of other proposed innovations, it is necessary to demonstrate that the concepts of SD and the analytic technique of dynamic simulation provide benefits over and above the known benefits of the elements that SD programs share with other types of innovations. Currently, there is a tendency for SD practitioners to compare their curricula with the worst examples of traditional teaching. However, a more convincing test would compare SD curricula with the best available alternative: a cooperative, learner-centered curriculum that focuses on problem solving, brings computers into the classroom, and emphasizes synthesis of knowledge, but employs a traditional modeling technique in place of SD.

- Only systematic experimentation can lead to the improvement of SD curricula. Currently teachers of SD are studying many different topics, and each teacher employs a slightly different procedure. Thus, when one class is more successful than another, it is not possible to isolate the cause. In order for the field to accumulate knowledge and move forward, controlled experimentation in which potentially causal variables are isolated one at a time can no longer be avoided.

- The absence of rigorous assessment leaves the field of SD education vulnerable to charges of bias. There are several well-known biases that occur when experiments are not adequately controlled. For example, Wason (1960) describes a "confirmation bias" in testing hypotheses; that is, people tend to focus on evidence that could support their hypothesis and to ignore evidence that could disconfirm their hypothesis. It is at least possible that, for every anecdote in favor of SD curricula, there is a less flattering anecdote that goes unreported. Untrained experimenters may also introduce expectation effects (Rosenthal 1960). Currently, SD is taught only by volunteers who are highly enthusiastic about SD and have strong expectations about its chances of success. It is possible that some of the gains made by current SD programs occur simply because teachers expect them to occur.

Priorities for Future Research

In summary, we believe the following areas represent the most pressing needs for future SD assessment efforts:

1. Increased consistency and standardization in SD programs at different schools;

2. Increased quality control and more intensive teacher training;

3. Better experimental control, including controlled comparisons with the "best" alternative programs; and

4. Basic psychological experimentation to document unsubstantiated claims for improved problem-solving processes.

Another area in which rigorous SD assessment techniques have the potential to enhance current practice is in learning organizations (Senge 1990). Indeed, if corporations spend time and money in an effort to become learning organizations, it is in their best interest to continually assess whether they are moving toward this goal.

According to Senge (1990), a learning organization is built on a foundation of five disciplines: systems thinking, personal mastery, mental models, shared vision, and team learning. Assessment instruments and techniques developed for K-university SD education can, in principle, be adapted to corporate settings in an effort to ascertain whether systems thinking goals are being achieved. Moreover, other assessment instruments and techniques can be developed to assess the progress of a corporation in the other four disciplines.

Conclusions

Some have argued that SD curricula cannot or should not be tested against more traditional methods because they are qualitatively different, or because the benefits of SD are difficult to define and quantify, or because important elements of SD classes will have to be sacrificed in order to provide a controlled comparison with traditional classes. It is true that assessing SD is likely to be difficult, complicated, and time consuming. But it is not impossible. Many of the hypothesized benefits of SD, in fact, do seem to be quantifiable and testable using currently available methods (e.g., increased knowledge retention, increased likelihood of transfer). Given the compelling need for assessment, the SD community must begin the process of resolving these problems.

References

Brown, D. F. 1992. Using examples and analogies to remediate misconceptions in physics: Factors influencing conceptual change. *Journal of Research on Science and Teaching.* 29: 17-34.

Bruner, J. 1963. *The Process of Education.* New York: Vintage Books.

Chi, M., R. Glaser, and M. Farr, eds. 1988. *The Nature of Expertise.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Draper, F. 1989. Letter to J. W. Forrester. May 2.

Forrester, J. W. 1993. System dynamics as an organizing framework for pre-college education. *System Dynamics Review.* 9(2): 183-194.

Forrester, J. W. 1994. Learning through system dynamics as preparation for the 21st century. MIT: D-4434-1.

Gallagher, J. J. 1994. Teaching and learning: New models. *Annual Review of Psychology.* 45: 171-195.

Gick, M. L. and K. J. Holyoak. 1980. Analogical problem solving. *Cognitive Psychology.* 12: 306-355.

Good, T. L. 1983. Recent classroom research: Implications for teacher education. In *Essential knowledge for beginning educators*, ed. D. C. Smith, 55-65. Washington, D. C.: American Association of Colleges of Teacher Education.

Mandinach, E. B. and H. F. Cline. 1993. Systems, science, and schools. *System Dynamics Review.* 9(2): 195-206.

Mandinach, E. B. and H. F. Cline. 1994. *Classroom Dynamics: Implementing a Technology-Based Learning Environment.* New Jersey: Lawrence Erlbaum Associates, Inc.

Meyer, R. E. 1989. Models for understanding. *Review of Educational Research.* 59: 43-64.

Newman, D. 1990. Opportunities for research on the organizational impact of school computers. *Educational Researcher.* 19(3): 8-13.

Perkins, D. N. and G. Salomon. 1989. Are cognitive skills context bound? *Educational Research.* 18: 16-25.

Roberts, N. 1978. Teaching dynamic feedback systems thinking: An elementary view. *Management Science.* 24(8): 836-843.

Roberts, N. 1981. Developing an introductory curriculum in system modeling and simulation: The system dynamics approach. *Technological Forecasting and Social Change.* 19: 175-186.

Rosenthal, R. 1969. Interpersonal expectations: Effects of experimenter's hypothesis. In *Artifacts in Behavioral Research*, ed. R. Rosenthal and R. L. Rosnow. New York: Academic Press.

Salomon, G. 1990. Studying the flute *and* the orchestra: Controlled vs. classroom research on computers. *International Journal for Educational Research.* 14: 37-47.

Salomon, G. 1991. Transcending the qualitative-quantitative debate: The analytic and systemic approaches to educational research. *Educational Researcher.* 20(6): 10-18.

Senge, P. M. 1990. *The Fifth Discipline: The Art of Practice of the Learning Organization.* New York: Doubleday.

Spiro, R. J., P. J. Feltovich, R. L. Coulson, and D. K. Anderson. 1989. Multiple analogies for complex concepts: Antidotes for analogy-induced misconception in advanced knowledge acquisition. In *Similarity and Analogical Reasoning*, ed. S. Vosniadou and A. Ortony, 498-531. Cambridge: Cambridge Univ. Press.

Sterman, J. D. 1994. Learning in and about complex systems. *System Dynamics Review.* 10: 291-330.

Sterman, J. D. 1989. Misperceptions of feedback in dynamic decision making. *Organizational Behavior and Human Decision Making Processes.* 43: 301-35.

Walberg, H. 1986. Synthesis of research on teaching. In *Handbook of Research on Teaching*, ed. M. C. Wittrock, 214-229. New York: Macmillan.

Wason, P. C. 1960. Reasoning about a rule. *Quarterly Journal of Experimental Psychology.* 12: 129-140.

Zimmerman, B. J. and M. Martinez-Pons. 1990. Student differences in self-regulated learning: relating grade, sex, and giftedness to self-efficacy and strategy use. *Journal of Educational Psychology.* 82: 51-59.