Does a Model Facilitate Learning?

Some preliminary experimental findings

David Wheat, Virginia Western Community College[1]
Larry Weathers, Harvard Public Schools
Robin Goldstein, Harvard Public Schools

Abstract

The purpose of the experiment described in this paper is to compare the learning that takes place with different methods of delivering essentially the same information about Gross Domestic Product to student groups. The main delivery methods discussed are (1) simple narrative only, and (2) the same narrative, accompanied by a diagram revealed in stages, using STELLA's "story" feature. This experiment was administered to secondary students in the Harvard Public Schools in Massachusetts, and to community college and secondary students in Virginia. Tentative results suggest that students having access to the model structure learn more than students receiving only narrative instruction.

1. The Problem

Economics instruction is under a microscope, and the picture is cloudy. Experiments by economists suggest that the value added by economics courses is "minimal" (Walstad and Allgood, 1999), economics literacy rates are only "modestly" encouraging (Wood and Doyle, 2002), and comparative statics instructional methods may be no more effective than mere verbal instruction (Cohn et al., 2001). Students often find it difficult to see how the proliferating textbook graphs relate to one another, and how the variables on different graphs interact to influence economic system performance. The elusive dynamics of interaction is almost certainly missed by students just struggling to keep track of the *ceteris paribus* assumptions behind each new graph. Heavy reliance on "chalk-and-talk" lecture methods (Becker and Watts, 1998) may compound the problem. What could explain the apparent weakness in economics instructional methods, and what solution is available?

[1] Also, a PhD student at the University of Bergen, in Norway.

Our hypothesis is that the traditional macroeconomics instructional method—graphical comparative statics presented in a lecture format—provides a weakly structured learning pattern for most economics students. Over forty years ago, educational psychologist Jerome Bruner (1963) concluded that "…the most basic thing that can be said about human memory…is that unless detail is placed into a structured pattern, it is rapidly forgotten." Bruner's *structured pattern* is Jay W. Forrester's (1994) system dynamics "…*framework* into which facts can be placed [so that] learning becomes more relevant and meaningful."

We share Forrester's vision. Two of the authors are actively engaged in implementing systems thinking and system dynamics modeling in a Massachusetts public school. The third is developing and using a system dynamics model and interactive learning environment (ILE) in a macroeconomics distance-learning course at a community college in Virginia. Also under development is a set of experiments to test the effectiveness of that ILE as an instructional tool. One of those experiments is the subject of this paper.

2. *MacroLab*

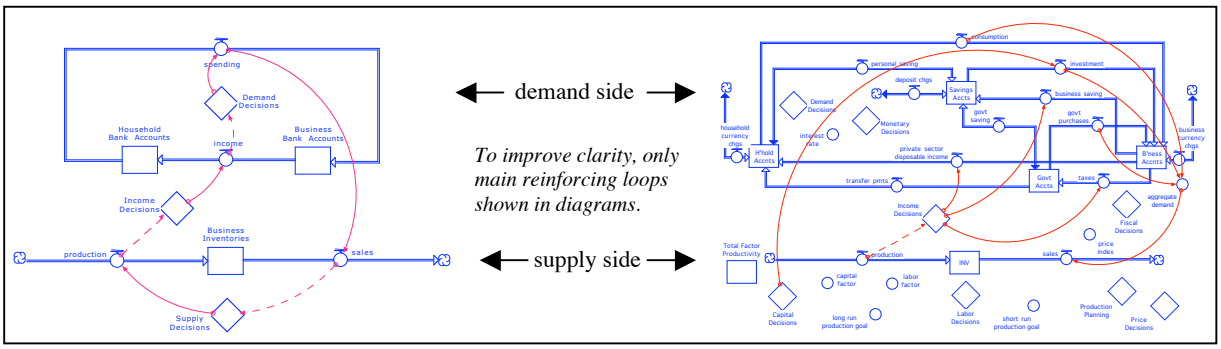
The ILE is called *MacroLab*, and it consists of six major learning units designed for a one-semester course in introductory macroeconomics (Circular Flow, Supply Side, Demand Side, Government, Money & Banking, and Exports & Imports).

Underlying System Dynamics Model. The model, built with STELLA[2] software, makes use of a traditional macroeconomics concept—a circular flow of income and spending—recast in stock and flow terms as the nominal "demand side" of an economic system. The "supply side"

[2] STELLA is a registered trademark of iseeSystems Inc. (www.iseesystems.com)

features a classic system dynamics representation of aggregate flows of real goods and services--production and sales--buffered by an inventory stock. Sales trends and inventory conditions provide information feedback that affects production goals and employment of factors of production. With real sales driven by nominal spending, and with nominal income driven by real production, the loop is closed between the demand and supply sides.

The basic structure is clear on the left of Figure 1, while admittedly less so on the right. The difference reflects the stage of development. The left-side version is studied by the students in the second week of the course, while the right-side is the end product of a full semester's



Starter Model (left) in Week 2 Figure 1. End-of-course Model (right)

construction project.[3]

User Interface. It should be emphasized that *MacroLab* is not a management flight simulator in the tradition of *People Express* (Sterman1988), nor does it resemble natural resource management games like *Fish Banks* (Meadows *et al*., 1993). It is an instructional tool in a

[3] Most link-forming arrows have been omitted to improve clarity. Those shown are in the main reinforcing feedback loop for economic growth (or decline). STELLA's diamond-shaped icons contain sub-models of various decision processes, within which most of the counteracting feedback loops operate. Figure 1 displays a closed economy, but an international sector will be added during the fall semester, 2004.

distance-learning course where students receive minimal guidance from an instructor.[4] With each new unit, *MacroLab* uses STELLA's "story" feature to display and explain the structure of sectors being added to the model, and suggest new behavior that can be expected. After reading the story (at home), students participate (at home) in simulation activities that compare behaviors of evolving structures while exploring traditional macroeconomics topics. They answer questions designed to assess understanding of model structure revealed in the story, model behavior observed during simulation experiments, and the connection between structure and behavior. Later, the students post their answers online, using the *Blackboard* distance learning technology. Some suggested solutions are available to aid the students during the activities, and instructor feedback is emailed soon after completion.

Designing a user interface—and simulation activities—for the distance learner with neither an economics nor system dynamics background has been challenging, but it has also imposed (and continues to impose) requirements for clarity and simplicity that have probably made *MacroLab* a more useful tool than it otherwise might be.[5] For the experiment discussed in this paper, the standard interface was simplified even more to minimize distraction for the participating subjects.

3. The GDP Experiment

Purpose. The purpose of the experiment is to compare the learning that takes place with three different methods of delivering essentially the same information about gross domestic

[4] MacroLab will be used in a regular classroom setting for the first time during the Fall 2004 semester. During the Spring 2005 semester, a microeconomics version (MicroLab) will take shape. At Worcester Polytechnic Institute, Jim Lyneis (jmlyneis@WPI.EDU) already uses system dynamics to teach microeconomics.

[5] Contact David Wheat (dwheat@wheatresources.com) for a current description or a demo copy of *MacroLab*.

product (GDP) to three groups of subjects. The three delivery methods are (1) simple narrative only, (2) the same narrative, accompanied by a diagram revealed in stages, and (3) the same narrative and diagram, accompanied by a simulation activity.

Subjects. Three variations of the GDP experiment involving 147 subjects were conducted in Massachusetts and Virginia over a six-month period from November 2003 to May 2004.[6] The November experiment was conducted in Massachusetts at Harvard Public Schools ("HPS" experiment), with participation by 68 junior and senior economics students in three different classes. The April experiment involved 58 students (ranging in age from 17 to 54) enrolled in four different classes of political science and macroeconomics at Virginia Western Community College in Roanoke, Virginia ("VWCC" experiment). The May experiment was joined by 21 junior and senior computer science students at Hidden Valley High School, also in Roanoke ("HVH" experiment).

Procedure. Subjects were randomly assigned to groups and pre-tested. In the HPS and VWCC experiments, subjects were assigned to the three groups (Group 1, Group 2, and Group 3). For reasons explained in the appendix, the HVH subjects were divided into just two groups (Groups 2 and 3).

The research design called for pre-testing to be completed at least a day earlier than the actual experiment, but that did not always happen, with the result that many subjects took the pre-test immediately prior to the experiment, and then took the post-test immediately after the experiment. It is possible that those subjects may have been more likely to give post-test answers that were the same as the pre-test answers. If so, then learning gain would be underestimated in those cases. To the extent that those subjects were randomly distributed

[6] In this paper, the focus is on the features common to each experiment.

among the groups, however, any "pre-test" effects should not affect differences observed between groups.[7]

The experiment for each group involved use of STELLA's story feature. Thus, all subjects were required to navigate to an instructions page and read their learning objectives. All had to click on a button that activated a story about GDP, and all had to "press the spacebar" to progress through their particular story.

Each page in the story read by Group 1 subjects contained text-only information about the meaning of GDP, its measurement, and how it fits in an overall economic system as a concept of "production." Group 2 subjects read a story with the same textual information as Group 1, but the Group 2 story was accompanied by an unfolding stock-and-flow diagram that revealed the structure of a simple economy in a manner to complement the narrative.

Group 3 read the same story as Group 2; i.e., the participants got both the narrative and the unfolding stock-and-flow diagram. However, when Group 3 participants finished the story, they activated a simple simulation exercise designed to reinforce a key point in the story about the relationship between production (GDP), inventories, and final sales. (The validity of the Group 3 results is questionable, however, and those results are discussed only in the appendix.) After each group finished its version of the experiment, post-tests were taken. The "learning gain" for each group was measured as the difference between average pre- and post-tests scores for that group.

Variations in Content and Procedure. Of necessity, most experiments go through a research and development phase in order to "iron out" the wrinkles in both procedure and content, so as to enhance the reliability and validity of the results. The GDP experiment is still in

[7] The results have not yet been reviewed to determine if such randomness can be assumed.

that R&D stage. After the HPS experiment debriefing, it was decided to simplify the experiment in two ways. First, the pre- and post-test questions were reduced in number, an effort was made to clarify them substantially, and a few questions that were problematic or less useful were deleted or replaced. As a result of those changes, the three questions below (with correct answers underlined) were only questions that were the same in all three experiments:

1. The definition of GDP is the total value of…
   (a) final gross profits produced within a nation during a year.
   (b) final profits, less depreciation, produced in a nation during a year.
   (c) <u>final goods and services produced in a nation during a year</u>.

2. The relationship between GDP, sales, and inventories is most like the one between…
   (a) profits, taxes, and depreciation.
   (b) <u>deposits, withdrawals, and bank balances</u>.
   (c) imports, exports, and exchange rates.

3. A rising GDP tends to raise income and then sales, which increases GDP again.
   (a) <u>True</u> (b) False

These questions are obviously simple, but they are intended to probe three different types of learning. Question 1 is a straightforward factual question, where merely recalling information contained in the "story" would produce a correct answer. Question 2 seeks awareness of an analogous relationship, and it requires a higher-order level of thinking than mere recall. No such analogy was stated in the narrative information read by participants in Groups 1. Likewise, there was no *explicit* analogy accompanying the "unfolding model structure" observed by Group 2 participants. However, among the elements in that structure are two stock-and-flow relationships that have been extracted from the model and reproduced below in Figure 2.

The top relationship (A) refers to the flow of dollars from business bank accounts to household bank accounts when income payments are made to the factors of production, and the return flow of dollars when those households purchase goods and services from businesses. The bottom relationship (B) refers to the flow of goods and services in the production process into an inventory stock that is reduced by subsequent sales. The textual narrative (available to both groups) mentions both types of stock and flow relationships, but not in the same context and not in a way that would imply an analogous relationship. However, this *visual* analogy is present for participants in Group 2.

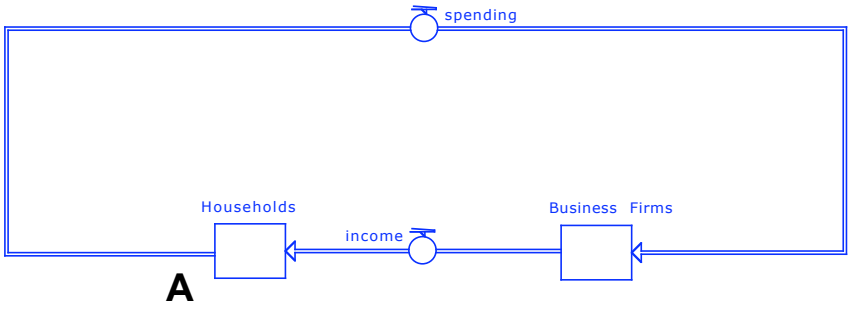


figure 2

Question 3 attempts to probe the participants' sense of dynamics—the behavior of GDP over time. In the narrative available to both Groups 1 and 2, it was emphasized that GDP (gross domestic product) is the *production* of final goods and services, so that information was the same for both groups. Also exactly the same for both groups was a series of statements that traced the impact of production on income (with a delay), the impact of income on sales (with a delay), and the effect of sales on production (delay), with the conclusion that "…production, income, and sales are part of a mutually-reinforcing process." However, Group 2 had access to an unfolding loop that accompanied that narrative (figure 3).

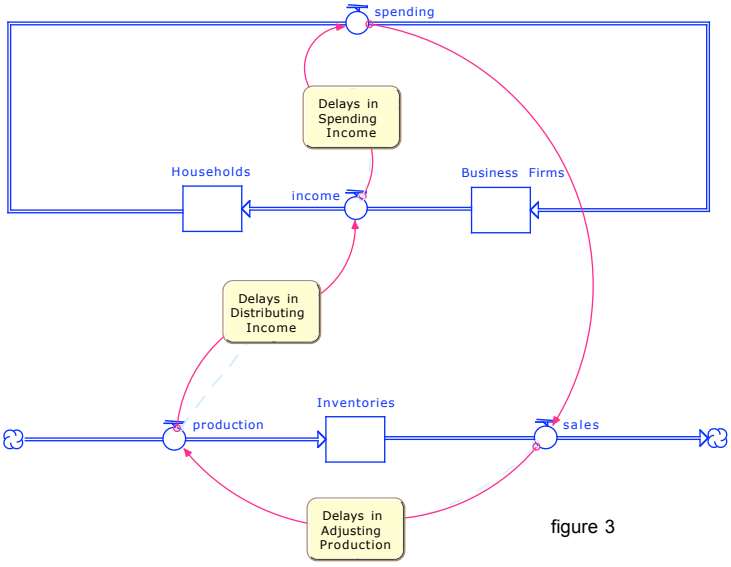


figure 3

4. Results of the Experiment

Recall the different experimental treatments for the two groups. Group 1 read a "STELLA story" containing a text-only narration. The Group 2 story had the same narrative, but was accompanied by unfolding stock-and-flow diagrams.

Hypothesis:

> *Group 2 would demonstrate more learning gain (as measured by the difference in pre- and post-test scores) than Group 1.*

The general results of the experiment are consistent with the hypothesis.

In the HPS experiment in Massachusetts, participants in Group 2 gained an average of 30 points between pre- and post-tests, compared to an average gain of 15 points by Group 1 participants.[8] Use of a one-tail t-test for statistical significance suggests that the probability of this result occurring by chance is 8 percent.

[8] "Points" refers to the percentage of correct answers. Thus, in the HPS experiment, the average Group 1 pre-test score was 57 percent (57 points out of 100), while the average post-test score was 72 (72 points out of 100). The learning gain, then, was 15 points.

A similar result was found in the VWCC experiment, but the Group 2 margin over Group 1 was smaller, and not statistically significant. The average Group 2 score also improved 30 points, but Group 1increased by 24 points ($p < .28$; one-tail t-test).

Combining the results of the HPS and VWCC experiments enables a larger sample, but raises a number of experimental and statistical issues regarding the comparability of the two groups. Thus, at this stage in our analysis, such a combination is presented as merely suggestive of statistical significance ($p < .07$) rather than a confirmation. The learning gains are displayed graphically in Figure 4.

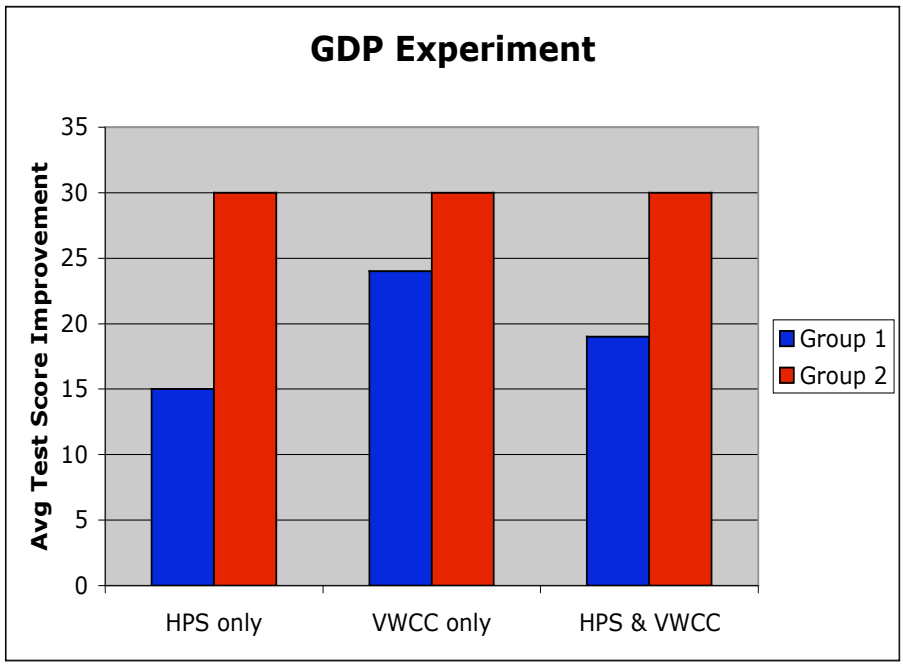


figure 4

* one-tail t-tests

5. Conclusion

The results of this experiment, while tentative, suggest that a model—a simplified representation of reality—can facilitate learning. In this case, the facilitation appears to be provided by simple diagrams that reinforce narrative text. Both Groups 1 and 2 received behaviorial descriptions, but only Group 2 was permitted to see diagrams of structure that could

infer such behavior. Other experiments are planned for *MacroLab*, and one will be aimed at more explicit testing of the connection between structural understanding and behavioral understanding, with the latter measured in terms of explanations and/or predictions of simulation patterns.

Interested readers are encouraged to contact the authors with comments on this experiment, as well as suggestions for the design of future experiments.

Appendix

The validity of Group 3 results is questionable. The results are provided here, however, along with a discussion of the issues surrounding that portion of the experiment.

Group 3 participants read the same story as Group 2 (i.e., they got both the narrative and the unfolding stock-and-flow diagram). However, when Group 3 subjects finished the story, they activated a simple simulation exercise designed to reinforce a key point in the story about the relationship between production, inventories, and sales.

After the HPS experiment, the simulation activity for Group 3 was simplified to reduce the subjects' required manipulation of control buttons, switches, etc. Those requirements appeared to be a serious distraction and source of frustration for the HPS subjects in Group 3, according to those who administered and monitored the experiments. Moreover, since Group 3 demonstrated *less* learning gain than the other two groups, we concluded that the excessive "management" of the simulation activity might have interfered with the learning process for Group 3. These content changes were put into effect for the VWCC and HVH experiments.

Despite the content changes, the VWCC subjects in Group 3 still fared poorly, compared to Groups 1 and 2. Further consultations with specialists in cognitive psychology, led to another revision—elimination of an "extra" task for Group 3.

In the HPS and VWCC experiments, Group 3 subjects had a pencil-and-paper assignment during the simulation activity. The pencil-and-paper exercise had been conceived as a means to optimize the benefits of that activity. However, we saw that the exercise added considerably to the total time that Group 3 was spending on the overall experiment, and we were concerned about tiring the participants and exceeding their attention spans. We also suspected that the

exercise itself might be causing more confusion than clarity. So, we eliminated the pencil-and-paper exercise during the Group 3 simulation activity in the HVH experiment. Due to the small number of subjects in the HVH experiment, only two instructional methods were tested—those for Group 2 and Group 3 (i.e., there was no Group 1 in the HVH experiment).

In the HVH experiment, for the first time, Group 3 showed a strong learning gain. Both Groups 2 and 3 improved averaged scores by 27 points (about the same as the 30 point gains by Group 2 in each of the HPS and VWCC experiments).

However, the hypothesis that Group 3 would learn *more* than Group 2 was not supported. It may be that the simulation exercise is just too simple to produce any measurable effects, when coupled with the simplicity of the three questions. Moreover, a potential design flaw in simulation display could be causing some problems. The graph displays production and sales—two flows that are easily understood. However, the graph also shows a flow that is fairly easy to grasp in narrative form but may be more of a challenge when presented graphically: a *change* in inventories. When inventories are rising at a decreasing rate (e.g., when production is slowing down to match a fall in sales), the graph of "change in inventories" shows a downward trend even though the stock of inventories is still rising (albeit more slowly). We think that is confusing the Group 3 participants, who have had no training in stocks and flows prior to exposure to such concepts during the experiment.

Thus, the current Group 3 results are probably not meaningful. Future experiments will attempt to improve the design so as to achieve valid results.

References

Becker, W., & Watts, M. (1998). *Teaching Economics to Undergraduates: An Alternative to Chalk and Talk*. Northampton, MA: Edward Elgar.

Bruner, J. S. (1960). *The Process of Education*. Cambridge, MA: Harvard University Press.

Cohn, E., Cohn, S., Balch, D., & Bradley, J. (2001). Do graphs promote learning in principles of economics? *Journal of Economic Education*, *Fall*, 299-310.

Forrester, J. W. Policies, decisions, and information sources for modeling. Morecroft and Sterman (1994).

Meadows, D.L., Fiddaman, T., and Shannon, D. (1993). *Fish Banks, Ltd. A Micro-computer Assisted Group Simulation That Teaches Principles of Sustainable Management of Renewable Natural Resources, 3rd ed*. Laboratory for Interactive Learning. University of New Hampshire, Durham, NH.

Morecroft, J. D. W., & Sterman, J. D. (eds.). (1994). *Modeling for Learning Organizations*. Portland, OR: Productivity Press.

Sterman, J. D. (1988). *People express management flight simulator.* Cambridge, MA: Sloan School of Management, MIT.

Walstad, W. B., & Allgood, S. (1999). What do college seniors know about economics? *American Economic Review*, *89:2,* 350–54.

Wood, W. C., & Doyle, J. M. (2002). Economic literacy among corporate employees. *Journal of Economic Education*, *Summer*, 195-205.