Papers

4. Do Stock and Flow Feedback Diagrams Promote Learning in Macroeconomics?

Do Stock-and-Flow Feedback Diagrams Promote Learning in Macroeconomics?

I. David Wheat, Jr.
Senior Lecturer in System Dynamics
University of Bergen

Abstract

This article describes the value added by a stock-and-flow feedback diagram to text-only instruction in macroeconomics. The experiment was motivated by a prior study in which the use of graphs to teach macroeconomics was no more effective than verbal instruction alone. Here, in contrast, students using the feedback diagram did show more improvement in post-test scores than those who received only narrative instruction.

JEL: A22, C91, E32
Key words: conceptual change, dynamics, education, experiment, mental model, meta-analysis, value added

Six years have passed since Cohn et al. (2001) provided an empirical basis for doubts about the pedagogical value added by graphs in the macroeconomics principles course. Motivated by that outcome, this article reports on a similar experiment. As in the Cohn study, two groups of students received narrative instruction on a macroeconomics topic, but one group also received a supplemental visual aid. Instead of a graph, however, the visual aid in this experiment was a stock-and-flow feedback diagram (Figure 4). Section 1 explores the issue of effective visual aids in economics, and section 2 summarizes the concepts of the feedback method. The experiment is described in section 3. Section 4 presents the results, which are discussed in section 5.

1. Visuals that Aid Learning

Can a stock-and-flow feedback diagram add pedagogical value to text-only instruction about GDP? That is the question addressed in this article. Less formally, we could ask, "Would students improve their understanding of GDP if their reading material included an annotated stock-and-flow feedback diagram?"

The answer might seem self-evident. The popular belief that a picture is worth a thousand words is supported by education research (Standing 1973), and the human brain has a remarkable capacity for long-term storage and retrieval of visual images (Bahrick et al. 1976). In addition, there is evidence that illustrations and diagrams facilitate learning at levels deeper than mere retention and recall. For students aptly described as "visual learners," *comprehension* of information is fostered by visualization of that information (Wolfe 2001, Shaw 2000). Less obvious, however, is the instructional value added by supplementary illustrations in economics classrooms and textbooks. There, the standard visual aid is the comparative static graph (Kennedy 2000, Cohn et al. 2001), and the Cohn study indicates that graphs may add little or nothing to mere verbal instruction. In fact, one experiment found that students in a graph-supplemented lecture actually showed *less* improvement than those in a lecture-only session. Thus, not every macroeconomics picture appears to be worth its proverbial opportunity cost. This article reports on an experiment to assess the value added by a different type of visual aid for macroeconomics—stock-and-flow feedback diagrams based on system dynamics methodology.

System dynamics is a method for studying and managing complex information feedback systems (Forrester 1961, Sterman 2000). When studying the structure and behavior of a complex system, it is necessary to engage in an iterative process of developing and refining a computer model and, at the same time, updating one's mental image of the system under study. If the primary purpose for the study is pedagogical, then properly transforming students' understanding of the system's structure and behavior is the ultimate goal. The computer modeling and all the associated tasks, including feedback loop diagramming, are merely means to an end—to improve students' ability to correctly envision, describe, and explain a process about which they know very little initially. If, on the other hand, the system dynamics project is managerial in nature, there will almost always be a client who is quite familiar with the "system" that needs better management. However, even then, the service is less likely to be development of a *computer* model than assistance in renovating the client's *mental* model of the systemic problem. The policy design task for the modeler is helping the client envision additional feedback structure that can be added to the system to improve the client's managerial effectiveness. The role of mental models in shaping perceptions of complex systems has been emphasized by system dynamics computer modelers for decades:

> *The mental image of the world around us that we carry in our heads is a model. One does not have a city or a government, or a country in his head. He has only selected concepts and relationships, which he uses to represent the real system.* (Forrester 1971, p. 3)

> *Each person carries in his head a mental model, an abstraction of all his perceptions and experiences in the world, which he uses to guide his decisions.* (Meadows et al. 1974, p. 3)

> *Mental models are deeply ingrained assumptions, generalizations, or even pictures or images that influence how we understand the world.* (Senge 1990, p. 8)

> *A mental model is a network of facts and concepts...that contains our understanding of social and physical phenomena.* (Morecroft 1994, p. 7)

> *A mental model of a dynamic system is a relatively enduring and accessible, but limited, internal conceptual representation of an external system ... whose structure is analogous to the perceived structure of that system.* (Doyle et al. 2000, p. 5)

The mental model concept is rooted in cognitive psychology and stems from Craik (1943). According to Rouse and Morris (1986), "mental models are the mechanisms whereby humans are able to generate descriptions of system purpose and form, explanations of system functioning and observed system states, and predictions of future system states." Seel (2001) views mental models as "inventions of the mind that represent, organize, and restructure domain-specific knowledge." More than forty years ago, education psychologist

Jerome Bruner (1960) 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, 1960) This suggests that when teaching about a complex system, an *effective* visual aid is, by definition, one that facilitates the structuring of a student's mental model of that system.

Given that system dynamics projects often have this "improving the mental model" requirement—whether the emphasis is on studying or managing—there is always a need for tools that facilitate communication between the modeler and the audience for the model. One such tool is the simple feedback diagram, either in stock-and-flow format (such as the GDP model used in this experiment) or even simple word-and-arrow diagrams that illustrate feedback loops.

At any point in time, students' mental models are imperfect approximations of their perceptions of reality. Moreover, a change in that perception is a prerequisite for a change in the mental model. Indeed, one could *define* teaching about real-world systems as an attempt to change students' mental models by improving their perception of how those systems work. Since a mental model is "relatively enduring" (Doyle et al. 2000, p. 5), the requisite perceptual change occurs gradually, by way of a self-adjusting learning cycle or counteracting feedback learning loop (Kolb 1984, Sterman 2000) that gradually updates prior perceptions and the associated mental model. However, Doyle and others emphasize the limited capacity of persons to form accurate perceptions of the structure of external *dynamic* systems and make accurate predictions of the behavior of such systems.[1] *In the context of teaching about complex dynamic systems, therefore, visual aids that clarify processes of change over time may facilitate desired mental model renovation.*

The pedagogical potential of stock-and-flow diagrams was suggested by Forrester's (1994) description of system dynamics as a "*framework* into which facts can be placed [so that] learning becomes more relevant and meaningful." Forrester's framework is Bruner's structured pattern.

2. Feedback Method

Over the past five years, an alternative approach to macroeconomics instruction has been developed to address these issues.[2] Called the *feedback method*, it draws on the mental model concept developed by psychologists and applied by system dynamicists, the tradition of feedback thinking in economics (Richardson, 1991), and the methodology of system dynamics computer modeling. Its purpose is to make the study of dynamic behavior in the economy accessible to a much broader population of students. For example, since students using the feedback method do not manipulate equations or rely on static graphs for visualization of dynamics, those without an aptitude for math could still appreciate the dynamics of market economies. Students who are gifted in math could understand macroeconomic dynamics before they learn calculus.

---

[1] On the issue of misperception of feedback, see Sterman (1989) and Moxnes (1998).

[2] The author uses this approach to teach a macroeconomics principles course delivered via the Internet to students enrolled at Virginia Western Community College in Roanoke, Virginia.

The conceptual building blocks for the feedback method are *stocks*, *flows*, and *feedback loops*, generically illustrated in Figure 1. These are also the central concepts in system dynamics, a method for studying and managing problems in complex feedback systems, based on Forrester (1961) and Sterman (2000). A *stock* is an accumulation of material (e.g., inventories) or information (e.g., sales data that are collected and analyzed prior to decisions about future production). A net *flow* is the rate of change in a stock. Mathematically, the stock integrates the net flow. The *feedback loop* transmits information about the state of the system from the stock to the decision rules that govern the flow. The flow, in turn, updates the stock. In this experiment, the visual aid was a simple system of material and information stocks and their associated flows. Figure 4 displays the completed version of the stock-and-flow diagram, and its development in stages is described in Appendix I.[3]

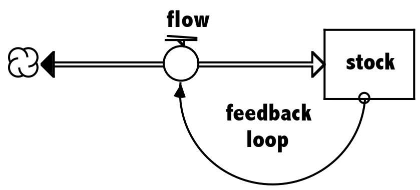


Figure 1. Generic Representation of Central Concepts in System Dynamics

3. The Experimental Design

Students were randomly assigned to a control group or experimental group. In both groups, students worked alone at individual computers, using the story-telling feature of the *STELLA* software.[4] Turning the "pages" in the GDP story was similar to advancing slides in presentation software. The first page listed the students' learning objectives:

Successful completion of this learning activity should enable you to…

- *define GDP and clarify its meaning,*
- *describe how GDP can be measured,*
- *explain how GDP fits in the "bigger picture" of a national economy.*

The story read by the control group contained only textual information about the meaning of GDP, its measurement, and its placement in an overall macroeconomy. Students in the experimental group read the same textual information, but their story was accompanied by an unfolding stock-and-flow feedback diagram that revealed the structure of a simple economy in a manner designed to complement the narrative. Appendix I contains the instructional content for both methods. Pre- and post-tests measured students knowledge and understanding. After a debriefing of those who administered the experiment at site 1, the test instrument for site 2 was clarified by reducing the number of questions and re-wording most of them. Figures 2 and 3 list the questions at sites 1 and 2, respectively.

---

[3] The diagrams in Figures 1 and 4 are merely the iconic structures of the underlying systems of equations that generate behavior during a simulation run.

[4] *STELLA* is a registered trademark of isee systems, inc. (http://www.iseesystems.com).

1. GDP stands for (a) gross depreciation & profits. (b) gross domestic product.* (c) gross domestic profits.
2. The definition of GDP is the total value of (a) goods & services produced in a nation during a year.* (b) profits, less depreciation, produced in a nation during a year. (c) gross profits produced within a nation during a year.
3. Using “dollars/year” in the measurement of GDP enables the combining of (a) profits from industries with different depreciation rates. (b) profits from domestic production and foreign-owned production. (c) the sales value of diverse goods & services.*
4. If rapidly rising production were followed by slowly rising incomes, that would cause producers’ inventories to (a) rise.* (b) fall. (c) remain the same.
5. If rapidly rising incomes were followed by slowly rising sales, that would cause producers’ inventories to (a) rise.* (b) fall. (c) remain the same.
6. Think about the relationship between GDP and inventories. That relationship is most like the one between (a) profits and depreciation. (b) deposits and bank balances.* (c) imports and exports.
7. Think about the relationship between GDP, sales & inventories. That relationship is most like the one between (a) profits, taxes, & depreciation. (b) deposits, withdrawals, & bank balances.* (c) imports, exports, & exchange rates.
8. “GDP minus sales equals inventory changes.” (a) That is true, because GDP is production.* (b) That is false, because GDP is production.
9. “Rising inventories is a sign of an imbalance between supply and demand.” (a) That is true, because that means sales have been lower than production.* (b) That is false, because supply and demand must be equal.
10. “Rising production tends to raise income and then sales. If so, GDP rises.” (a) That is true, because an increase in sales causes production to increase, and production is GDP.* (b) That is false, because production is GDP, and “cause” cannot become the “effect.”

Figure 2. Site 1 Test Questions
*correct answers

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) final goods and services produced in a nation during a year.*
2. The relationship between GDP, sales, and inventories is most like the one between (a) profits, taxes, & depreciation. (b) deposits, withdrawals, & bank balances.* (c) imports, exports, & exchange rates
3. If inventory levels are falling, then GDP and sales are equal. (a) True (b) False*
4. A rising GDP tends to raise income and then sales, which increases GDP again. (a) True * (b) False
5. Which is the most likely production trend after the sudden drop in sales? (a) Production would also drop quickly, but then remain constant at the new lower rate. (b) Production would decline slowly.* (c) Production would increase slowly.
6. Which is the most likely sales trend after the sudden drop in sales? (a) Sales would remain constant at the new lower rate. (b) Sales would continue to decline, but slowly.* (c) Sales would increase slowly.
7. If prices did change after the sudden drop in sales, then (a) sales would remain constant at the new lower rate. (b) sales would continue to decline slowly. (c) sales would increase slowly.*

Figure 3. Site 2 Test Questions
*correct answers

The effectiveness of the instructional methods was defined in terms of test score *improvement*, and two measures were utilized: *higher score* and *greater learning gain*. The first was the percentage of students in each group with post-test scores higher than pre-test scores. The second was the average normalized percentage learning gain for students in each group. The two measures suggested two distinct hypotheses.

Higher Score Hypothesis. Let P_{TS} and P_{FS} refer to the proportion of students at each Site (1 or 2) who raised their test scores after receiving the Text and Feedback instruction, respectively. The null hypothesis was

$$H_0: \; P_{TS} = P_{FS} \,,$$

implying no difference in the effects of the two instructional methods.

Greater Learning Gain Hypothesis. The two methods were also compared in terms of student learning gain, defined as each student's normalized percentage gain (NPG) in test scores. The NPG is the difference between individual pre- and post-test scores, expressed as a percentage of the maximum possible improvement in the pre-test score.

$$\text{NPG} = \frac{100^* \text{(posttest score - pretest score)}}{\text{(100 - pretest score)}}$$

The denominator in the NPG equation is the learning "gap" that needs to be closed, based on the pre-test. The quantity in parentheses in the numerator is the absolute gain. Thus, NPG is the *percentage of the gap* that is *closed* after the instruction.[5] Let G_{TS} and G_{FS} refer to the mean normalized percentage gain at each Site for the T and F instructional methods, respectively. The second null hypothesis, then, was H_{00}: $G_{TS} = G_{FS}$.

Sample Selection and Characteristics. At site 1, the forty-six participants were junior and senior economics students at Harvard Public Schools near Boston, Massachusetts.[6] Near Roanoke, Virginia, the site 2 participants included twenty-seven economics students from Dabney Lancaster Community College and twelve political science students from Virginia Western Community College. None of the participants at either site had prior experience with system dynamics stock-and-flow concepts or diagrams. In contrast, *at least* 85 percent of all the students—100 percent at site 1 and 69 percent at site 2—had completed almost a full semester of economics that included at least some macro instruction. The percentage could be even higher because the number of political science students with prior economics training was unknown and assumed to be zero.

---

[5] The "gap closing method" follows Soper (1973, 1976), supported by Highsmith (1976) notwithstanding Becker's criticism (1976).

[6] I am grateful to Larry Weathers for administering the experiment at Harvard Public Schools.

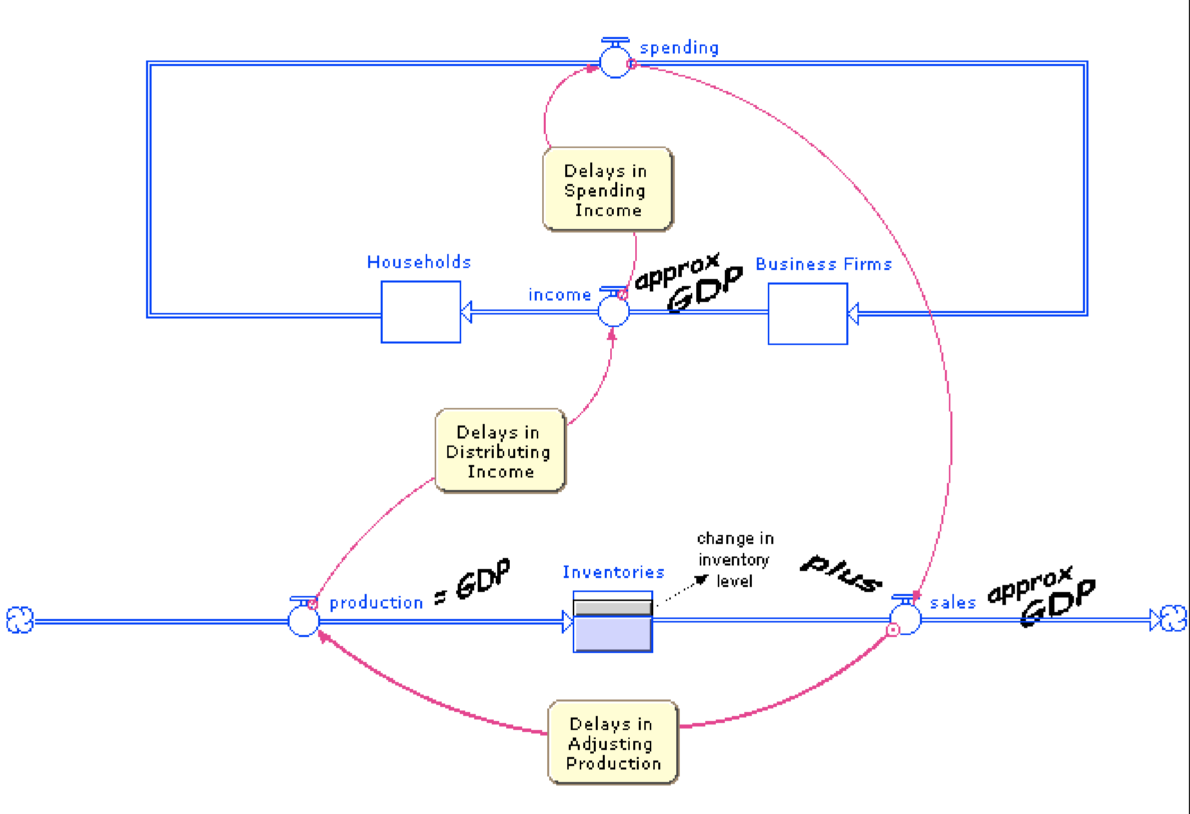


Figure 4. This Completed Stock-and-Flow Feedback Diagram Was Unfolded in Stages during the Instructional Phase of the Experiment

4. Results

Despite all the similarities in the experiments at the two sites, the slightly different test instruments seemed to preclude traditional pooling of the data for analysis. A more conservative approach involves meta-analysis, which produced the final results in this study. The pre-meta analysis (or "preliminary") results are summarized in Figures 5 and 6, and Appendix II contains the raw data. For both measures of improvement at both sites, the students using the feedback method to supplement the textual explanation of GDP outperformed those relying on text alone. The percentage differences ranged from 25-30 percent for the "higher score" measure of improvement, to nearly 100 percent for the "learning gain" measure. However, the samples were small and the variance wide at both sites, and statistical significance was elusive. Only at site 2 on the learning gain measure (Figure 6) did the results even break the .10 threshold in a traditional two-tail test.[7]

	Instructional Method		
	Text-only	Feedback & Text	p
Site 1 n = 46	P_{T1} = 58.3%	P_{F1} = 72.7%	0.314
Site 2 n = 39	P_{T2} = 61.1%	P_{F2} = 81.0%	0.174

Figure 5. Proportion Scoring Higher Pre-Meta Analysis Results

[7] For the tests in Figure 5, the t values were 1.01 (df = 44) and 1.37 (df = 37) for sites 1 and 2, respectively. For the tests in Figure 6, the t values were 0.99 (df = 44) and 1.71 (df = 36) for sites 1 and 2, respectively.

	Instructional Method		p
	Text-only	Feedback & Text	
Site 1	G_{T1} = 16.1%	G_{F1} = 32.2%	0.325
Site 2	G_{T2} = 25.3%	G_{F2} = 48.8%	0.094

Figure 6. Mean Normalized Pct. Gain
Pre-Meta Analysis Results

A common research strategy in similar situations is to boost the sample size by pooling the data from the two sites. Here, however, differences in the test instruments militated against standard pooling techniques. Only three of the questions were identical at both sites, the number of test items was different, and there was reason to believe that site 1 students found some questions hard to understand. Although the experiments at the two sites were variations rather than replications, both test instruments were designed to elicit the same information, and the results at both sites were consistent. A method of aggregating the results was needed, and the selection was *meta-analysis*, described by Stanley (2001, pp. 131-133):

> *Meta-analysis is a body of statistical methods that have been found useful in reviewing and evaluating research results. If a number of independent studies have been conducted on a particular subject, using different data sets and methods, then combining their results can furnish more insight and greater explanatory power than the mere listing of individual results. ... High returns in the advancement of empirical understanding await economic researchers willing to develop and apply meta-analysis.*

Stanley documents the contribution that meta-regression analysis is making to empirical research in economics. He also provides a summary of widespread applications in other research fields where, instead of estimating regression parameters, the research goal is often the evaluation of various studies of experimental medical treatments in health care or instructional methods in education. Finally, he illustrates the general methodology of meta-analysis and discusses its strengths and weaknesses. A vast literature exists in this field, and Stanley's references provide an introductory reading list.[8]

As typically practiced, meta-analysis involves statistical analysis of a large number of individual studies of a particular research question over a long time period, with the purpose of systematically and objectively aggregating and evaluating the weight of the cumulative evidence. Stanley's (1998) meta-analysis of empirical evidence of Ricardian equivalence, for example, examined twenty-eight separate studies published between 1980 and 1995. The meta-analysis of the minimum wage impact on employment by Card and Krueger (1995) focused on fifteen published studies over a twenty-year period. In a 1994 meta-analysis of the effect of education funding on student outcomes, Hedges et al. reviewed thirty-eight studies spanning a decade. In contrast, the meta-analysis performed here involves just two studies spanning less than six months. Unlike conventional meta-analyses, our purpose was

[8] Meta-analysis as practiced today traces its origins to Glass (1976, 1977). Before reading the seminal works, however, see Glass (2000) for a useful and entertaining background.

not to strengthen our understanding of studies done by others. Our more limited goal was to strengthen the interpretation of our own two studies, which were probably more similar in design than most pairs of studies in the meta-analysis research literature. The separate experiments at Sites 1 and 2 had so much in common that pooling would have been used if the test instruments had been identical.

The two experiments actually produced four outcomes—two at each site. At each site, one outcome addressed the "higher score" hypothesis and the other evaluated the "greater gain" hypothesis. Both outcomes measured a particular effect of the diagram-enhanced instructional method, but they were not statistically independent since the same students were used to test both hypotheses. Thus, *two separate meta-analyses* were necessary. Each meta-analysis treated a pair of independent outcomes (e.g., "greater gain" results produced at Sites 1 and 2) as a sample of size two. For that sample, a weighted average of the results was calculated, based on the standardized mean and variance and the sample size at each site. That produced an aggregate point estimate of the standardized difference in means for the feedback method (F) and the text-only method (T). Analysis of variance techniques established probability levels and confidence intervals. The results are summarized in Figures 7 and 8, and Appendix III contains the statistics.

	Standardized Difference in Means: $\mathbf{P_F - P_T}$	p
Site 1	29.9%	.314
Site 2	44.2%	.174
Meta-Analysis	36.4%	.097

Figure 7. Final Higher Score Results

In Figure 7, $\mathbf{P_F - P_T}$ is the standardized difference in the mean proportion of students scoring higher on the post-test for instructional methods F and T. The p values for Sites 1 and 2 are the same as in Figure 5. The bottom row, however, displays the results of the meta-analysis, where the weighted-average of the standardized difference in means is 36.4%, with a posterior probability of less than 10 percent that the null hypothesis was true.

	Standardized Difference in Means: $\mathbf{G_F - G_T}$	p
Site 1	29.2%	.325
Site 2	55.7%	.094
Meta-Analysis	41.0%	.067

Figure 8. Final Greater Gain Results

In Figure 8, $\mathbf{G_F - G_T}$ is the standardized difference in the mean normalized percentage gain (NPG) for instructional methods F and T. The p values for Sites 1 and 2 are the same as in Figure 6, and the bottom row summarizes the meta-analysis. The weighted-average of the standardized difference in means is 41.0%, with an updated probability of a Type I error less than 7 percent.

4. Discussion

Students with access to the stock-and-flow feedback diagram were more likely to show test score improvement. They also showed larger learning gains. The meta-analyses yielded higher statistical confidence levels than the results at each site considered alone. Are those higher levels credible? One way to answer that question is to compare the meta-analysis confidence levels with those that would have been attained if standard pooling of the data had been done. Recall that pooling was *not* done because the test instruments at the two sites, though similar, were not identical. If the results *had been* pooled, the better performance of the feedback method would have shown at .06 and .04 confidence levels for the higher score and greater gain measures, respectively. As expected, the statistical confidence levels associated with the meta-analysis are more conservative than the results that pooling would have produced. Credibility of the meta-analysis is enhanced by the intuitively appealing result that the confidence levels fall in between pooled results and stand-alone results. Nevertheless, even the meta-analysis statistical significance levels are borderline—slightly above the 5 percent cut-off point usually deemed acceptable. We explore issues that are relevant to interpreting the results, with the goal of identifying the potential and direction of any bias. Specifically, we look for potential bias arising from sample selection, instructional content, and time-on-task.

Almost all participants brought prior economics education to the experiment, while none brought system dynamics training. That could affect the outcome in two ways, both of which would dilute the effect of the feedback diagram and push the results in the direction of the null hypothesis. First, sheer familiarity with economic terms and concepts, such as GDP, could have generated pre-test scores for both groups that would be higher than expected from students without any economics education. Higher baseline pre-test scores would reduce somewhat the opportunities for improvement as a result of any instruction, and would diminish the potential for the more effective mode of instruction to display differential effects. Secondly, the use of familiar economic terminology in the text-only instruction could have had a reinforcing effect on students with prior economics education. That would have inflated the impact of the text-only method, making the beneficial impact of the feedback diagram harder to detect. In short, the background of the participants may have biased the results in *favor* of the null hypotheses—contrary to the actual results. In the absence of such bias, the confidence level achieved by the results might have been even higher.

A comparison of the instructional content for the two methods (reproduced in Appendix I) should confirm that the students in both groups received the same textual information about GDP. The only distinction was that the experimental group also received a stock-and-flow feedback diagram (with minimal additional annotation). Moreover, a careful reading of the instructional content reveals that the text-only instruction contained numerous references to stocks, flows, delays, and even the bathtub analogy, a staple of simple system dynamics exposition. In other words, *the students in the control group received textual instruction that was replete with system dynamics concepts*. The similarity of the two methods, therefore, was not limited to information content. Both also had a system dynamics conceptual foundation. The only difference was that the control group lacked the visual reinforcement of the stock-and-flow feedback diagram. The instructional content, therefore, could not be considered biased in favor of the diagram.

A criticism could be directed at the relative time requirements of the two instructional methods. Most students in the control group completed the textual instruction and post-test in about fifteen minutes. The students in the experimental group needed about twice as long to complete their task. Arguably, the additional time devoted to studying the feedback method could be partially responsible for the post-test score improvement. However, whenever diagrams are supplementary to textual information, that constitutes extra instructional material for the students. Studying the extra visual material requires extra time. If the textual material is the same—as the experiment required—the overall time requirements will necessarily increase. An *experimental solution* would be to require students in the control group to *repeat* their study of the textual material (with predictably less enthusiasm). A more practical solution would be to set a *minimum* time limit that would be long enough for control group students (and experimental group students, too, if they finished quickly) to review the material before post-testing. However, such experimental "solutions" raise interesting questions about real-world instructional time requirements and efficiency. Would a stock-and-flow feedback diagram supplement a lecture in such a way that the overall impact would be as efficient as giving the lecture twice without the diagram?

The bottom line issue for the time-on-task question is whether the additional time produced commensurate results. Did the benefits outweigh the costs? Was the extra fifteen minutes worth increasing the proportion of improved scores by 25-30 percent? Was the extra fifteen minutes worth doubling the mean normalized percentage gain? Viewed in percentage terms, the extra fifteen minutes doubled the time requirements, and such questions are not trivial. Viewed in absolute terms, however, the payoff seems clear.

This paper was motivated by Cohn's (2001) comparison of graphs and narrative instruction, and it may be considered an *indirect* comparison of system dynamics-based diagramming methods and conventional graphical instruction. The results here suggest that feedback diagramming adds value to mere narrative instruction by facilitating perception of systemic structure and its attendant behavior. Despite the consistency of results in this experiment, there is clearly a need to upgrade its design in ways that justify higher levels of confidence. A larger sample size is needed, and participants should have no prior exposure to economics or system dynamics. Better field-testing of questions is necessary to reduce the variance in mean response measurements, which is a prerequisite for raising statistical confidence. Beyond these design issues, the scope of the research should be widened to learn more about how students' form perceptions of the structure and behavior of economic systems, how that perception formation process is influenced by instructional methods, and —ultimately—how students' mental models of economic systems can be most effectively and efficiently improved.

References

Bahrick, H. P., Bahrick, P. O., & Wittlinger, R. P. (1976). Fifty Years of Memory for Names and Faces: A Cross-Sectional Approach. *Journal of Experimental Psychology*, *104*, 54-75.

Becker, W. (1976). Programmed Instruction in Large-Lecture Courses: A Technical Comment. *Journal of Economic Education*, *8*(1), 38-40.

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

Card, D. & Krueger, A. B. (1995). Time-Series Minimum Wage Studies: A Meta-Analysis. *American Economic Review*, *85*(2), 238-243.

Cohn, E. S., Cohn, S., Balch, D., & Bradley, J. (2001). Do Graphs Promote Learning in Principles of Economics? *Journal of Economic Education*, *Fall*, 299-310.

Doyle, J. K., Ford, D. N., Radzicki, M. J., & Trees, W. S. (2000). Mental models of dynamic systems. *Working Papers, Department of Social Science and Policy Studies, Worcester Polytechnic Institute*, *27*.

Ford, A. (1999). *Modeling the Environment* (paperback ed.). Washington, D.C.: Island Press.

Forrester, J. W. (1961). *Industrial Dynamics.* Cambridge MA: Wright-Allen Press.

______ (1971). Counterintuitive Behavior of Social Systems. *Collected Papers of J.W. Forrester.* Cambridge, MA: Wright-Allen Press.

______ (1994). Policies, Decisions, and Information Sources for Modeling. In J. D. W. Morecroft & J. D. Sterman (Eds.), *Modeling for Learning Organizations.* Portland, OR: Productivity Press.

Glass, G. V. (1976). Primary, Secondary, and Meta-Analysis of Research. *Educational Researcher*, *5*, 3-8.

______ (1977). Integrating Findings: The Meta-Analysis of Research. *Review of Research in Education*, *5*, 351-379.

______ (2000). Meta-Analysis at 25. *UC Berkeley-Stanford Colloquium on Meta-Analysis, Department of Psychology* Retrieved January 19, 2007, http://glass.ed.asu.edu/gene/papers/meta25.html

Hedges, L. V., Laine, R. D., & Greenwald, R. (1994). Does Money Matter? A Meta-Analysis of Studies of the Effects of Differential School Impacts on Student Outcomes. *Educational Researcher*, *23*(3), 5-14.

Highsmith, R. (1976). Second Generation Research in Economic Education. *Journal of Economic Education*, *8*(1), 48-51.

Kennedy, P. E. (2000). *Macroeconomic Essentials: Understanding Economics in the News.* Cambridge MA: The MIT Press.

Kolb, D. A. (1984). *Experiential Learning: Experience as a Source of Learning and Development.* Upper Saddle River, NJ: Prentice Hall Inc.

Meadows, D. L., Behrens III, W. W., Meadows, D. H., Naill, R. F., & Zahn, E. (1974). *Dynamics of Growth in a Finite World.* Cambridge, MA: Wright-Allen Press.

Morecroft, J. D. W. (1994). Executive Knowledge, Models, and Learning. In J. D. W. Morecroft & J. D. Sterman (Eds.), *Modeling for Learning Organizations.* Portland, OR: Productivity Press.

Moxnes, E. (1998). Overexploitation of Renewable Resources: The Role of Misperceptions. *Journal of Economic Behavior and Organization*, *37*(1), 107-127.

Richardson, G. P. (1991). *Feedback Thought in Social Science and Systems Theory.* Waltham, MA: Pegasus Communications, Inc.

Rouse, W. B. & Morris, N. M. (1986). On Looking into the Black Box: Prospects and Limits in the Search for Mental Models. *Psychological Bulletin*, *100:3*.

Seel, N. (2001). Epistemology, Situated Cognition, and Mental Models: Like a Bridge over Troubled Water. *Instructional Science*, *29*, 403-427.

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

Shaw, G. (2000). *Keeping Mozart in Mind.* San Diego, CA: Academic Press.

Soper, J. C. (1973). Programmed Instruction in Large-Lecture Courses. *Journal of Economic Education*, *Spring*, 125-129.

______ (1976). Second Generation Research in Economic Education: Problems of Specification & Independence. *Journal of Economic Education*, *8*(1), 40-48.

Standing, L. (1973). Learning 10,000 Pictures. *Quarterly Journal of Experimental Psychology*, *25*, 207-222.

Stanley, T. D. (1998). New Wine in Old Bottles: A Meta-Analysis of Ricardian Equivalence. *Southern Economic Journal*, *64*, 713-727.

______ (2001). Wheat from Chaff: Meta-Analysis as Quantitative Literature Review. *The Journal of Economic Perspectives*, *15*(3), 131-150.

Sterman, J. D. (1989). Misperception of Feedback in Dynamic Decision-Making. *Organizational Behavior and Human Decision Processes*, *43*(3), 301-335.

______ (2000). *Business Dynamics: Systems Thinking and Modeling for a Complex World.* Boston, MA: McGraw-Hill Companies.

Wheat, I. D., Jr. (2007a). Adding Feedback Loops to the Macro Instructional Toolkit. *The Feedback Method: A System Dynamics Approach to Teaching Macroeconomics.* Unpublished PhD dissertation, University of Bergen, Bergen, Norway.

______ (2007b). Student Preferences when *Explaining* Dynamics. *The Feedback Method: A System Dynamics Approach to Teaching Macroeconomics.* Unpublished PhD dissertation, University of Bergen, Bergen, Norway.

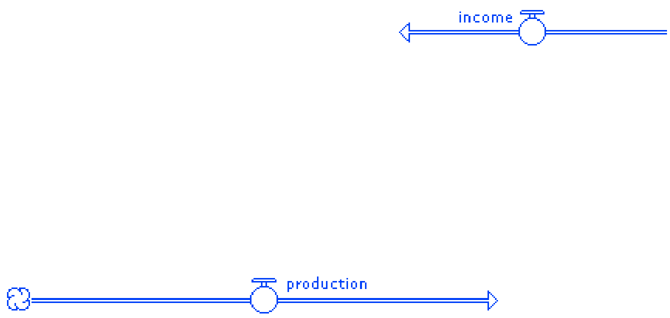
______ (2007c). Student Preferences when *Learning* Dynamics. *The Feedback Method: A System Dynamics Approach to Teaching Macroeconomics.* Unpublished PhD dissertation, University of Bergen, Bergen, Norway.

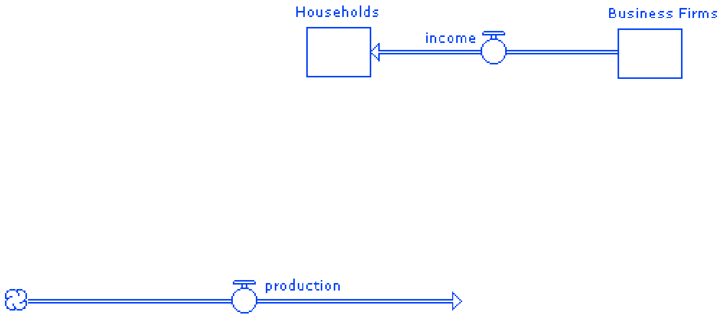
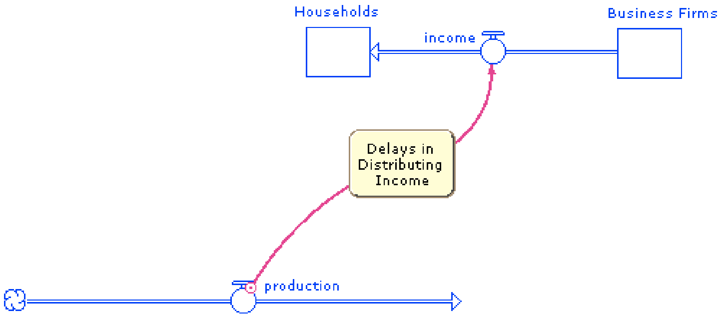
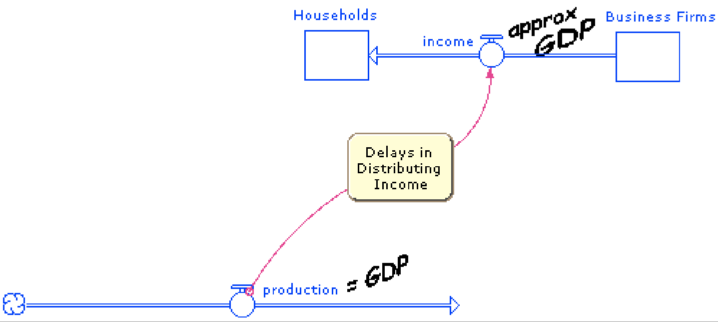
______ (2007d). Teaching Business Cycle Dynamics: A Comparison of Graphs and Loops. *The Feedback Method: A System Dynamics Approach to Teaching Macroeconomics.* Unpublished PhD dissertation, University of Bergen, Bergen, Norway.

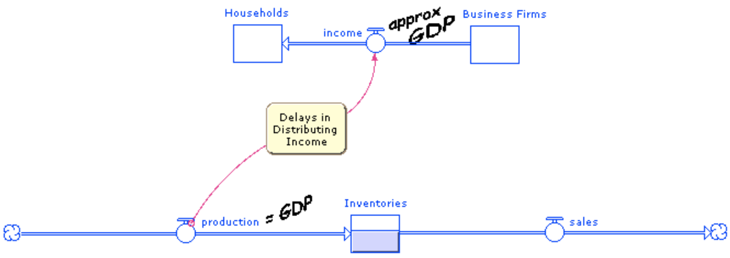
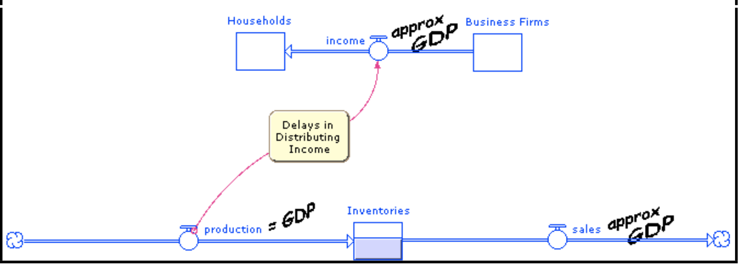
Wolfe, P. (2001). *Brain Matters: Translating Research into Classroom Practice.* Alexandria, VA: Association for Supervision and Curriculum Development.

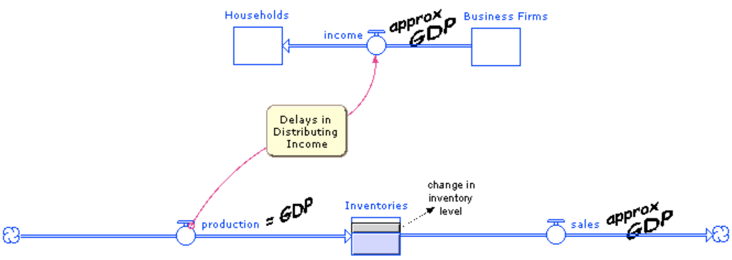
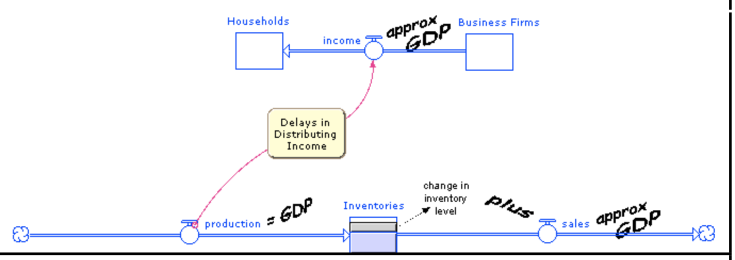
Appendix I. Instructional Content

In the right-hand column below, any textual differences between the two methods are shown in *italics*. When a diagram first appears in the feedback method story, it will be shown in the right-hand column. During the instruction received by the students, all diagrams remained visible all the time, even as more text was added. To conserve space in this table, however, the diagrams will not be repeatedly shown. This table will only show diagrams when they are added or modified.

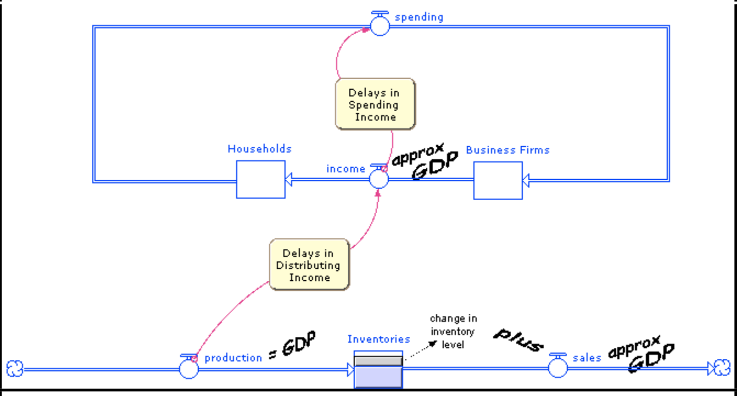
Text-only Method (T)	Feedback Method (F)
1. What is GDP?	1. What is GDP?
2, In a national economy, the annual rate at which final goods & services are produced is called the Gross Domestic Product, or just GDP. GDP is measured in trillions of dollars/year in the U.S.	2. In a national economy, the annual rate at which final goods & services are produced is called the Gross Domestic Product, or just GDP. GDP is measured in trillions of dollars/year in the U.S. *This pipeline icon symbolizes a flow of final goods & services being produced nationally during a year. The annual rate at which that flow moves through the pipeline is the Gross Domestic Product, or just GDP.* 
3. GDP, therefore, is a measure of the value of final goods & services produced nationally during a year. These days, the production of final goods & services in the U.S. is valued at about 10 trillion dollars/ year.	3. GDP, therefore, is a measure of the value of final goods & services produced nationally during a year. These days, the production of final goods & services in the U.S. is valued at about 10 trillion dollars/year.
4. There is a corresponding flow of income (wages, interest, rent, profits) being paid to the factors of production (those responsible for producing a nation's goods & services).	4. *This pipeline icon symbolizes a* flow of income (wages, interest, rent, profits) being paid to the factors of production (those responsible for producing a nation's goods & services). 

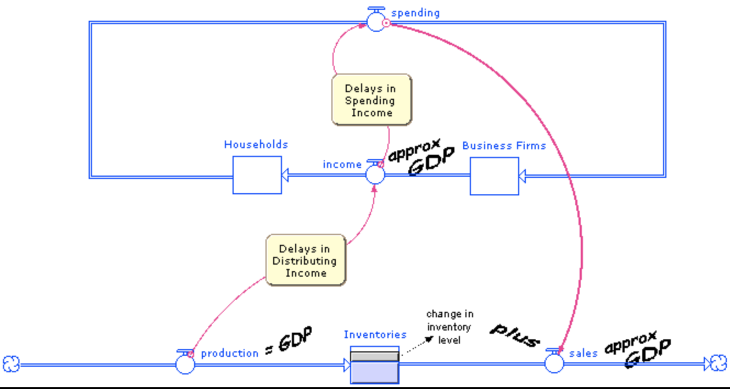
5. The income flows from a stock of money in business firms' bank accounts into another stock of money in households' bank accounts. Income is also measured in trillions of dollars/year.	5. The income flows from a stock of money in Business Firms' bank accounts into another stock of money in Households' bank accounts. *As it passes through the "valve" in the middle of the pipeline,* income is also measured in trillions of dollars/year. 
6. Income is earned as production occurs, but actual distribution occurs sometime thereafter.	6. Income is earned as production occurs, but actual distribution occurs sometime thereafter. *A change in the production rate causes a change in the rate of income distribution.* 
7. Workers typically get paid extra wages in the month following the extra production, and shareholders may wait many months before receiving additional profits. Information has to be collected and income distribution decisions made before the "income flow" responds fully to changes in the "production flow." The production rate is the definition of GDP: the annual rate at which the economy produces all final goods & services.	*7. The change in income is not instantaneous.* Workers typically get paid extra wages in the month following the extra production, and shareholders may wait many months before receiving additional profits. Information has to be collected and income distribution decisions made before the "income flow" responds fully to changes in the "production flow." The production rate is the definition of GDP: the annual rate at which the economy produces all final goods & services. 
8. Over time, income flowing to the factors of production approximates the value of the goods & services being produced--GDP. Thus, the term "income" is sometimes used interchangeably with GDP.	8. Over time, income flowing to the factors of production approximates the value of the goods & services being produced--GDP. Thus, the term "income" is sometimes used interchangeably with GDP. *Even direct income payments are affected by delays (e.g., workers don't receive paychecks at the end of each production day). Other delays are due to roundabout flows (e.g., taxes to government followed by income to households after government spending occurs).*
9. How is GDP measured?	9. How is GDP measured?

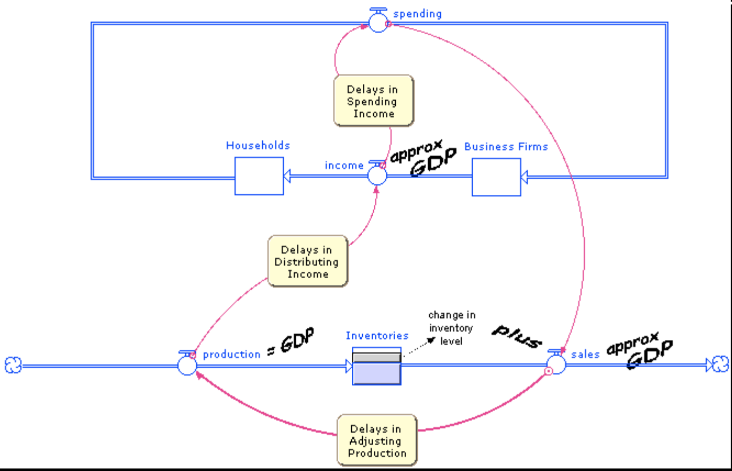
10. GDP cannot be measured directly. An indirect method is needed.	10. GDP cannot be measured directly. An indirect method is needed.
11. Income flows are approximately equal to the value of production flows. Thus, "adding up the income" flows is one method of estimating GDP. It takes time for income to respond fully to changes in production. During such transitions--which are normal--income is only an approximation of GDP.	11. Income flows are approximately equal to the value of production flows. Thus, "adding up the income" flows is one method of estimating GDP. It takes time for income to respond fully to changes in production. During such transitions--which are normal--income is only an approximation of GDP.
12. Another way to estimate GDP is to "add up the sales" flows. Goods that have been produced flow into inventories before being sold. When sales occur, goods flow out of those inventories.	12. Another way to estimate GDP is to "add up the sales" flows. Goods that have been produced flow into inventories before being sold. When sales occur, goods flow out of those inventories. 
13. If the inventories are at the same level at the beginning and end of a month, then the production inflow and the sales outflow must have been equal during that month. Analogy: If the level of water in a bathtub remains the same even though water is flowing in and out, then the water inflow and outflow rates must be equal.	13. If the inventories are at the same level at the beginning and end of a month, then the production inflow and the sales outflow must have been equal during that month. Analogy: If the level of water in a bathtub remains the same even though water is flowing in and out, then the water inflow and outflow rates must be equal.
14. Thus, when there is no change in inventory levels, sales data could be used to estimate the production rate for the goods. In that case, adding up the sales receipts for the entire economy provides an estimate of GDP.	14. Thus, when there is no change in inventory levels, sales data could be used to estimate the production rate for the goods. In that case, adding up the sales receipts for the entire economy provides an estimate of GDP. 

15. Suppose, however, that the inventory level at the end of the month is DIFFERENT from the inventory level at the beginning of the month.	15. Suppose, however, that the inventory level at the end of the month is DIFFERENT from the inventory level at the beginning of the month. 
16. In fact, a change in inventory levels is the rule rather than the exception. Inventories tend to rise and fall from month to month.	16. In fact, a change in inventory levels is the rule rather than the exception. Inventories tend to rise and fall from month to month.
17. Inventory levels rise when the production rate is faster than the sales rate. And inventory levels fall when the production rate is slower than the sales rate.	17. Inventory levels rise when the production rate is faster than the sales rate. And inventory levels fall when the production rate is slower than the sales rate.
18. When inventory levels are changing, sales data ALONE would not provide a good estimate of production. And, adding up the sales receipts for the entire economy would not provide a good estimate of GDP.	18. When inventory levels are changing, sales data ALONE would not provide a good estimate of production. And, adding up the sales receipts for the entire economy would not provide a good estimate of GDP.
19. However, the change in inventories is equal to the difference between production and sales. When production exceeds sales by $100 billion worth of goods during the year, then inventories will rise by that amount during that year. Thus, the adding-up-the-sales method still provides a good estimate of GDP when the amount of change in inventories is added to the sales rate. (Inventories fall when sales exceed production; in that case, the change in inventories is subtracted from sales.)	19. However, the change in inventories is equal to the difference between production and sales. When production exceeds sales by $100 billion worth of goods during the year, then inventories will rise by that amount during that year. Thus, the adding-up-the-sales method still provides a good estimate of GDP when the amount of change in inventories is added to the sales rate. (Inventories fall when sales exceed production; in that case, the change in inventories is subtracted from sales.) 
20. How does GDP fit into the bigger picture of the economy?	20. How does GDP fit into the bigger picture of the economy?
21. The various stocks (e.g., bank accounts and inventories) and flows (e.g., production, sales, and income) mentioned in this story are part of a very simple model of the economy.	21. The various stocks (e.g., bank accounts and inventories) and flows (e.g., production, sales, and income) in this diagram are part of a very simple model of the economy.
22. A change in the production rate causes the income rate to change, but with delays that have been discussed.	22. A change in the production rate causes the income rate to change, but with delays that have been discussed.

<table>
<tr>
<td>23. A change in income causes spending (i.e., sales) to change, but not immediately. Some delays are due to consumers adjusting spending patterns directly; others are due to roundabout effects such as deposits to savings accounts that get spent later by consumers, businesses, or government.</td>
<td>23a. A change in income causes spending to change, but not immediately. Some delays are due to consumers adjusting spending patterns directly; others are due to roundabout effects such as deposits to savings accounts that get spent later by consumers, businesses, or government.

Spending causes funds to flow from Households' bank accounts to Business Firms' bank accounts.


</td>
</tr>
<tr>
<td></td>
<td>*23b. Spending is the same activity as sales; i.e., spending equals sales.*


</td>
</tr>
<tr>
<td>24. A change in sales causes production rates to change, as producers seek to provide supplies that match demand.</td>
<td>24. A change in sales causes production rates to change, as producers seek to provide supplies that match demand.


</td>
</tr>
</table>

25. Production adjustments take time, however. First, producers must decide whether a change in sales is just temporary. It also takes time to make adjustments in factors of production (e.g., labor and capital) if production goals do change.	25. Production adjustments take time, however. First, producers must decide whether a change in sales is just temporary. It also takes time to make adjustments in factors of production (e.g., labor and capital) if production goals do change.
26. Therefore, production, income, and sales are part of a mutually-reinforcing process. The process gains momentum over time, but that momentum is slowed by the decision-making delays along the way.	26. Therefore, production, income, and sales are part of a mutually-reinforcing process. The process gains momentum over time, but that momentum is slowed by the decision-making delays along the way.

Appendix II. Data

Site 1 questions:		1	2	3	4	5	6	7	8	9	10
answers:		b	a	c	a	a	b	b	a	a	a
T students											
1.1.T	pre	b	a	x	a	b	a	a	a	a	b
	post	b	a	c	a	b	a	a	a	a	a
2.1.T	pre	b	c	a	a	a	a	a	a	a	b
	post	b	c	c	a	a	a	a	b	a	b
3.1.T	pre	b	a	x	a	c	b	b	b	a	b
	post	b	a	c	a	b	b	b	a	b	a
4.1.T	pre	b	c	c	a	b	b	b	a	a	b
	post	b	a	c	a	b	b	b	a	a	b
5.1.T	pre	b	a	b	a	c	b	b	a	a	a
	post	b	a	c	a	a	b	b	b	a	b
6.1.T	pre	b	a	c	a	b	a	a	a	a	a
	post	b	a	c	a	b	a	b	a	a	a
7.1.T	pre	b	c	a	a	a	a	a	a	a	b
	post	b	a	c	a	b	b	b	a	a	b
8.1.T	pre	b	a	c	a	b	b	b	b	a	a
	post	b	a	x	a	a	b	b	a	b	b
9.1.T	pre	b	a	c	a	b	c	c	b	a	b
	post	b	a	c	a	b	b	b	a	a	a
10.1.T	pre	b	a	c	a	a	a	c	a	a	b
	post	b	a	c	a	a	b	b	a	a	b
11.1.T	pre	b	a	a	a	b	b	a	a	a	a
	post	b	a	c	a	b	b	b	a	a	a
12.1.T	pre	b	a	a	b	a	a	c	b	a	a
	post	b	a	b	a	b	a	a	a	b	a
13.1.T	pre	b	a	c	a	c	b	b	a	a	a
	post	b	a	c	a	b	c	c	a	a	a
14.1.T	pre	b	a	c	a	b	b	b	b	a	a
	post	b	a	c	a	b	b	b	a	a	b
15.1.T	pre	b	a	c	b	c	a	a	b	a	a
	post	b	a	c	a	b	b	b	a	b	b
16.1.T	pre	b	c	b	b	a	a	c	a	b	a
	post	b	a	c	a	b	a	b	a	a	b
17.1.T	pre	b	c	c	b	b	b	b	b	a	a
	post	b	a	c	a	a	a	a	a	b	a
18.1.T	pre	b	c	c	a	a	b	b	b	a	b
	post	b	c	c	a	b	b	b	a	a	b
19.1.T	pre	a	a	a	a	c	a	b	b	a	b
	post	b	a	c	a	b	b	b	b	a	a
20.1.T	pre	b	a	b	a	a	b	b	a	a	a
	post	b	a	c	a	b	b	b	a	a	b
21.1.T	pre	b	c	b	c	b	b	a	b	a	b
	post	c	c	a	a	c	a	c	a	b	a
22.1.T	pre	b	a	c	b	a	b	a	a	b	b
	post	b	a	a	a	c	b	b	a	a	a
23.1.T	pre	b	a	b	a	b	a	a	a	a	a
	post	b	a	c	a	a	b	b	b	a	a
24.1.T	pre	b	a	c	a	b	b	b	a	a	a
	post	b	a	c	a	a	b	b	b	a	a

Site 1 questions:		1	2	3	4	5	6	7	8	9	10
answers:		b	a	c	a	a	b	b	a	a	a
F students											
1.1.F	pre	b	a	a	a	a	b	b	a	b	b
	post	b	a	c	a	a	b	b	a	b	b
2.1.F	pre	b	a	a	a	a	a	a	a	a	b
	post	b	a	a	a	b	a	a	a	a	a
3.1.F	pre	c	c	c	a	b	a	a	a	a	a
	post	b	a	c	a	b	b	b	b	a	a
4.1.F	pre	b	c	a	b	a	a	a	a	a	b
	post	b	a	b	a	b	b	b	a	a	b
5.1.F	pre	b	c	c	a	b	a	c	a	a	a
	post	b	a	b	a	b	c	b	a	a	a
6.1.F	pre	b	a	a	c	b	a	c	a	b	b
	post	b	a	c	a	b	c	b	a	a	b
7.1.F	pre	b	c	x	b	a	c	c	a	a	b
	post	b	a	c	a	b	c	c	a	a	b
8.1.F	pre	b	a	c	b	a	b	b	a	a	a
	post	b	a	c	a	a	b	b	a	a	b
9.1.F	pre	b	c	b	a	b	a	a	a	a	b
	post	b	a	c	a	b	b	b	a	a	b
10.1.F	pre	b	c	a	a	b	b	c	a	a	a
	post	b	a	c	a	b	b	b	a	a	a
11.1.F	pre	b	c	c	a	b	a	a	a	a	a
	post	b	a	a	a	b	a	a	b	a	a
12.1.F	pre	b	a	a	a	b	b	b	a	a	a
	post	b	a	c	a	b	b	b	a	a	a
13.1.F	pre	b	c	a	a	b	a	a	b	a	b
	post	b	a	c	a	a	b	b	a	a	b
14.1.F	pre	b	c	a	a	b	b	b	a	a	a
	post	b	a	a	a	b	b	b	a	a	a
15.1.F	pre	b	c	b	b	b	b	b	a	a	b
	post	b	a	c	a	c	b	b	a	a	a
16.1.F	pre	b	a	a	c	a	b	b	a	a	b
	post	b	a	c	a	b	b	b	a	a	b
17.1.F	pre	b	a	c	a	b	b	c	a	a	a
	post	b	a	a	a	b	b	b	a	a	a
18.1.F	pre	b	a	b	a	b	b	a	a	a	b
	post	b	a	c	a	c	b	b	a	a	a
19.1.F	pre	b	a	c	b	b	a	a	a	a	a
	post	b	a	c	a	b	b	b	a	a	a
20.1.F	pre	b	a	a	a	c	b	b	a	a	a
	post	b	a	a	a	a	a	a	a	a	b
21.1.F	pre	b	c	a	b	a	b	b	a	a	b
	post	b	a	c	a	a	b	b	b	a	b
22.1.F	pre	b	a	c	a	b	a	b	a	b	b
	post	b	a	a	a	b	a	a	b	a	b

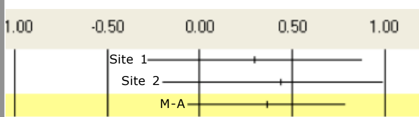
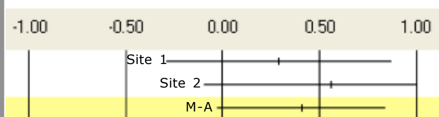
See Figure 2 for wording of questions 1-10 at site 1. A correct answer was scored as “1” and an incorrect answer as “0.”

Appendix II. Data, continued

Site 2 questions:		1	2	3	4	5	6	7		Site 2 questions:		1	2	3	4	5	7	9
answers:		c	b	b	a	b	b	c		answers:		c	b	b	a	b	b	c
T students										F students								
1.2.T	pre	c	c	b	a	b	c	a		1.2.F	pre	c	b	b	a	b	c	b
	post	c	b	a	a	b	a	c			post	c	c	b	a	b	c	b
2.2.T	pre	b	a	b	a	b	b	c		2.2.F	pre	a	b	b	a	b	b	c
	post	c	b	b	b	b	a	c			post	c	b	b	a	b	b	c
3.2.T	pre	a	c	b	a	a	c	b		3.2,F	pre	b	b	b	a	a	c	a
	post	c	b	a	a	c	c	c			post	c	b	b	a	b	a	c
4.2.T	pre	c	c	a	b	a	a	c		4.2.F	pre	c	b	b	a	a	a	a
	post	c	b	a	b	a	b	a			post	c	b	b	a	a	b	c
5.2.T	pre	c	b	b	a	a	b	c		5.2.F	pre	c	a	b	a	b	b	b
	post	c	b	b	a	b	c	c			post	c	b	b	a	b	b	a
6.2.T	pre	c	b	b	a	a	c	c		6.2.F	pre	c	a	b	b	b	a	c
	post	c	b	b	a	a	c	c			post	c	b	b	a	b	b	c
7.2.T	pre	a	a	b	a	b	a	c		7.2.F	pre	c	b	b	b	b	c	a
	post	a	c	b	a	a	c	c			post	c	b	b	a	b	c	a
8.2.T	pre	a	a	b	b	a	c	c		8.2.F	pre	c	c	a	b	b	b	c
	post	c	b	b	a	b	c	c			post	c	b	b	a	b	b	
9.2.T	pre	c	c	a	a	a	b	a		9.2.F	pre	a	b	b	a	b	a	a
	post	c	b	b	b	b	a	b			post	c	b	b	a	a	a	a
10.2.T	pre	c	b	b	a	b	a	b		10.2.F	pre	c	a	b	a	c	a	c
	post	c	x	b	a	b	b	c			post	c	b	b	a	b	b	c
11.2.T	pre	a	a	b	a	c	c	a		11.2.F	pre	c	b	a	a	b	c	b
	post	c	b	b	a	c	c	c			post	c	b	b	a	b	a	c
12.2.T	pre	c	b	b	b	b	c	c		12.2.F	pre	c	a	b	a	b	b	a
	post	c	b	b	a	b	a	c			post	c	b	b	a	b	b	b
13.2.T	pre	b	a	b	a	b	a	b		13.2.F	pre	b	a	b	a	a	a	c
	post	c	b	b	a	b	b	b			post	c	b	b	b	b	a	c
14.2.T	pre	c	b	b	a	b	b	c		14.2.F	pre	b	b	b	a	a	a	c
	post	c	c	b	a	b	b	a			post	c	b	b	a	b	b	a
15.2.T	pre	c	a	b	a	b	b	c		15.2.F	pre	c	b	b	b	b	b	a
	post	c	a	b	a	b	b	a			post	c	b	b	a	b	b	a
16.2.T	pre	b	c	b	a	c	b	b		16.2.F	pre	a	c	b	a	a	a	c
	post	a	a	b	b	a	a	c			post	c	b	b	a	a	a	c
17.2.T	pre	a	a	b	a	b	a	b		17.2.F	pre	a	c	b	a	a	a	c
	post	c	b	b	a	b	b	b			post	c	b	b	a	b	a	a
18.2.T	pre	c	a	a	a	a	a	a		18.2.F	pre	b	c	b	a	b	b	a
	post	c	b	b	a	b	a	c			post	a	b	b	a	b	c	a
										19.2.F	pre	c	b	a	b	a	a	c
											post	c	a	b	a	b	a	c
										20.2.F	pre	c	a	b	a	b	b	b
											post	c	b	b	s	b	b	c
										21.2.F	pre	b	a	b	a	a	b	c
											post	a	b	a	a	b	a	c

See Figure 3 for wording of questions 1-7 at site 2. A correct answer was scored as “1” and an incorrect answer as “0.”

Appendix III. Meta-Analysis Statistics

	Meta-Analysis 1 Higher Score Hypothesis			Meta-Analysis 2 Greater Gain Hypothesis		
	Studies		Meta-Analysis	Studies		Meta-Analysis
	Site 1	Site 2		Site 1	Site 2	
Relative Weight	0.545	0.455		0.557	0.443	
Text-only Mean	0.583	0.611		0.161	0.253	
Text-only Std Dev	0.504	0.502		0.621	0.482	
Text-only Sample	24	18		24	17	
Feedback Mean	0.727	0.810		0.322	0.488	
Feedback Std Dev	0.456	0.402		0.462	0.367	
Feedback Sample	22	21		22	21	
Std Diff in Means	0.299	0.442	0.364	0.292	0.557	0.410
Standard Error	0.297	0.325	0.219	0.297	0.332	0.221
Variance	0.088	0.106	0.048	0.088	0.111	0.049
lower limit	-0.283	-0.196	-0.066	-0.289	-0.095	-0.024
upper limit	0.881	1.079	0.793	0.874	1.208	0.843
Z-value	1.007	1.358	1.660	0.985	1.675	1.850
p-value	0.314	0.174	0.097	0.325	0.094	0.064
	95% confidence intervals Std Difference in Means  			95% confidence intervals Std Difference in Means  		
Q-test for heterogeneity	null hypothesis: Q = 0 Q = 0.105 p = 0.746 ==> sites 1 & 2 homogenous			null hypothesis: Q = 0 Q = 0.353 p = 0.553 ==> sites 1 & 2 homogenous		

Statistics generated by Borenstein M, Hedges L, Higgins J, Rothstein H. *Comprehensive Meta-analysis Version 2*, Biostat, Englewood NJ (2005). (http://www.meta-analysis.com/)