Table of Contents

Go Back

Learning Bathtub Dynamics: A Follow-up

Moderator: John Sterman
Sloan School of Management - MIT
30 Wadsworth Street, E53-351
Cambridge, MA 02142
Phone: 617-253-1951
E-mail: jsterman@MIT.EDU

Coordinator and Convener: Lees Stuntz,
Director Creative Learning Exchange
1 Keefe Rd.
Acton, MA 01720
Phone:978-287-0070
E-mail: stuntzln@clexchange.org

Panel:

Linda Booth Sweeney
Harvard Graduate School Of Education
75 Reservoir St.
Cambridge, MA 02138
Phone: 617-354-1390
E-mail:
Linda_Booth_Sweeney@harvard.Edu

Diana Fisher
Wilson High School
1151 SW Vermont St
Portland, OR 97219
Phone: 503 916-5280
E-mail: dmfisher@aracnet.com

John Heinbokel
Vermont Commons School
75 Green Mountain Dr.
South Burlington, VT 05403
Phone: 802.865.8084
E-mail:
heinbokel@vtcommonsschool.org

Gordon Kubanek
Ottawa Carleton District Schoolboard
621 Southmore Dr. W.
Ottawa, ON K1V7A4
Phone: 613 733 0610
E-mail: kubanek@enoreo.on.ca

Deb Lyneis
Creative Learning Exchange
PO Box 121, 215 Landgrove RD
Weston, VT 05161
Phone: 802 824-4219
E-mail: lyneisd@clexchange.org

Jim Lyneis
Worcester Polytechnic Institute
PO Box121, 215 Landgrove RD
Weston, VT 05161
Phone: 802 824-4219
E-mail: jmlyneis@alum.mit.edu

Jeff Potash
Vermont Commons School
75 Green Mountain Dr.
South Burlington, VT 05403
Phone: 802.865.8084
E-mail:
jeffpotash@vtcommonsschool.org

Rob Quaden
Carlisle Public Schools
83 School Street
Carlisle, MA 01741
Phone: 978 369-6550
E-mail: quaden@carlisle.mec.edu

Alan Ticotsky
Carlisle Public Schools
83 School ST
Carlisle, MA 01741
Phone: 978-369-6550
E-mail: ticotsky@rcn.com

Ron Zaraza
Portland Public Schools
2544 SE 16th Ave.
Portland, OR 97202
Phone: 503-916-5280
E-mail: rzaraza@jps.net

Abstract

At the Bergen ISD Conference in 2001, John Sterman and Linda Booth-Sweeney gave a talk entitled "Bathtub Dynamics". That talk has inspired teachers across the world to repeat Sterman and Booth-Sweeney's experiment to test different populations on the system dynamics concepts inherent in the bathtub exercise. These educators have used the Bathtub Dynamics Protocols from Sterman and Booth-Sweeney in their classrooms ranging from middle school through college. The discussion covers :

1. *Methodology for using this activity and data gathering*
2. *Usability of this activity at various age levels.*
3. *The ability to teach these concepts at various levels.*
4. *The implications for future teaching and learning.*

Bathtub Dynamics at Carlisle Public Schools

Rob Quaden
Carlisle Public Schools
83 School Street
Carlisle, MA 01741
Phone: 978 369-6550
E-mail: rquaden@carlisle.mec.edu

Alan Ticotsky
Carlisle Public Schools
83 School ST
Carlisle, MA 01741
Phone: 978-369-6550
E-mail:ticotsky@rcn.com

Bathtub Dynamics

In May, 2003, eighth grade students in Carlisle Public Schools performed the systems thinking inventory from "*Bathtub Dynamics: Initial results of a systems thinking inventory*" by Linda Booth Sweeney and John Sterman. "Honors" students, roughly the upper 50% in math achievement, were given the tasks in the bathtub dynamics and cash flow problems.

Students performed the task independently in a quiz format and were promised extra credit for correct answers.q

Students are in the ninth month of an Algebra I class and have learned to make simple STELLA models.

Q 1-7 (Bathtub Task 1) and Q 1-8 (Bathtub Task 2) are the evaluation criteria from Booth-Sweeney and Sterman

Bathtub Task 1

	Q 1	**Q 2**	**Q 3**	**Q 4**	**Q 5**	**Q 6**	**Q 7**
Totals	38	39	38	35	36	32	29
Percentages	0.83	0.85	0.83	0.76	0.78	0.7	0.63

Bathtub Task 2

	Q 1	**Q 2**	**Q 3**	**Q 4**	**Q 5**	**Q 6**	**Q 7**	**Q 8**
Totals	30	34	39	22	20	12	9	7
Percentages	0.65	0.74	0.85	0.48	0.43	0.26	0.2	0.15

Department Store Task

In May, 2003, eighth grade students in Carlisle Public Schools performed the systems thinking inventory. All students in the grade level performed the task.

Students performed the task independently in a quiz format and were promised extra credit for correct answers.

Students are in the ninth month of an Algebra I class and have learned to make simple STELLA models. Honors students are roughly the upper 50% in math achievement, Enriched are the other 50%.

Flow Q1 means “During which minute did most people enter the store?”
Flow Q2 means “During which minute did most people leave the store?”
Stock Q1 means “During which minute were the most people in the store?”
Stock Q2 means “During which minute were the fewest people in the store?”

Department Store Data

	Flow Q1	**Flow Q2**	**Stock Q1**	**Stock Q2**
Honors (N=46)	44	44	6	5
	95.65%	95.65%	13.04%	10.87%
Enriched (N=39)	36	37	2	2
	92.31%	94.87%	5.13%	5.13%

Booth Sweeney, L., Sterman J., Bathtub Dynamics: Initial results of a systems thinking inventory, *System Dynamics Review*,**9**(2): 249-286.

Bathtub Dynamics at Vermont Commons School

John Heinbokel
Vermont Commons School
75 Green Mountain Dr.
South Burlington, VT 05403
Phone: 802.865.8084
E-mail:
heinbokel@vtcommonsschool.org

Jeff Potash
Vermont Commons School
75 Green Mountain Dr.
South Burlington, VT 05403
Phone: 802.865.8084
E-mail:
jeffpotash@vtcommonsschool.org

1. We used the simplest of the bathtub tasks, #1 in Booth-Sweeney's and Sterman's presentation. For two of our groups we used the water flow illustration. For two others, the finance illustration, modified with the inclusion of 1) a STELLA stock-flow diagram replicating the picture of B-S & S and 2) with the inclusion of a second task (in addition to the original one, "to draw the behavior of [the stock]…"): "write (on the back of this sheet) a brief prose description of what was happening to the balance and what was causing any predicted changes to occur."

2. In two cases we provided a written sheet and simply stepped back and let the students work individually for about 10-15 minutes. In two other cases we passed out the sheets and then verbally went through the "quiz" making sure that the language (e.g. "receipts" or "expenditures") was clear and that the meaning of the graphs of the two flows was clear for the students. Then they pursued the written task.

3. Students were in four social science course sections:

Group	Grade	Course	Prior Modeling	Water/dollars	Init Verbal Explan.?
1-2(n=7)	7-8	Vt History	Use of BOTGs, Use of Models, Minimal Model Construct	Water	No
3A(n=10)	8-9	Human Geogr	Ditto, Many Had Model Constr. Experience Previous Year	Dollar	No
3B(n=12)	8-9	Human Geogr.	Ditto	Dollars	Yes
4(n=10)	9-10	Global Stud I	Many Had Model Constr Experience Previous Year	Water	Yes

4. Statistical Summary

B-S&S Evaluation #	% Correct 1-2	3A	3B	4
1.	.14 (1)	10 (1)	.75 (9)	80 (8)
2.	.14 (1)	10 (1)	.75 (9)	80 (8)
3.	14 (1)	10 (1)	.75 (9)	80 (8)
4.	29 (2)	10 (1)	.83 (10)	80 (8)
5.	29 (2)	0	.67 (8)	70 (7)
6.	.14 (1)	0	.50 (6)	60 (6)
7.	.14 (1)	0	.50 (6)	60 (6)
Stock mimics in-flow[1]	.43 (3)	.50 (5)	0	10 (1)

5. Discussion: We had initially assumed that the "poor" performance that our youngest students had exhibited, even after fairly extensive exposure to and use of BOTGs during the first half of the year prior to this exercise, might be due to an unfamiliarity with the B-S&S illustration. To correct for that, we added a comparable, but familiar, STELLA map of the same problem; we also asked for a verbal explanation of the process being plotted in order to stimulate a different (non-graphical) sort of processing. Group 3A got that revised task (Lees, BTW, 3A was the group we just brought to DQ)

   On debriefing those first two groups, we learned that many of the students didn't understand the finance terminology and that many had trouble understanding, from the graphical presentation, what the behavior of the flows were. Subsequent groups received a verbal explanation of the terminology and a quick description of what the flow graphs were depicting. All their questions were answered (except "What's the correct answer?") before turning them loose. In that process, might we have inadvertently cued them into the correct responses? Possibly, although we tried hard to maintain a neutral stance.

6. Our sense as educators is that, in addition to a basic comprehension of systemic behaviors, students need a variety of basic, non-SD, skills and insights in order to do well on this task. The young students with whom we worked seemed to lack much of that foundation and it seemed to be more that foundation deficit than SD deficit that prevented them from performing well. The specific form of the "picture" depicting the system, whether the B-S&S one or the STELLA stock-flow, seemed to make little difference. Being quite clear on the meaning of the flow and their values and changes over time, DID, however, make a marked difference in how well the students could work through the problem. For those of us trained in the traditional sciences, or who have pursued quantitative tasks for years, interpreting those graphs may seem to be a trivial exercise and expectation. Our experience suggests otherwise. Where and when in a student's schooling should those weaknesses or deficiencies be addressed? Is this an area where SD in the student's experience can provide that background (our students hadn't gotten there, despite fairly frequent exposures to this sort of

thinking)? Those aren't the questions we were asked to address, but they are the ones that strike us today as the more interesting.

In terms of "truth in advertising," the only group that got a truly "pure" B-S&S test was the Group 1-2 kids; all the others got variously adulterated versions in our effort to isolate the SD comprehension from other elements of intellectual capacity.

[1] These students depicted the BOTG of the Stock as following the same, step-wise, pattern of shifting plateaus as did the BOTG of the in-flow that they were presented.

Bathtub Dynamics – Ottawa

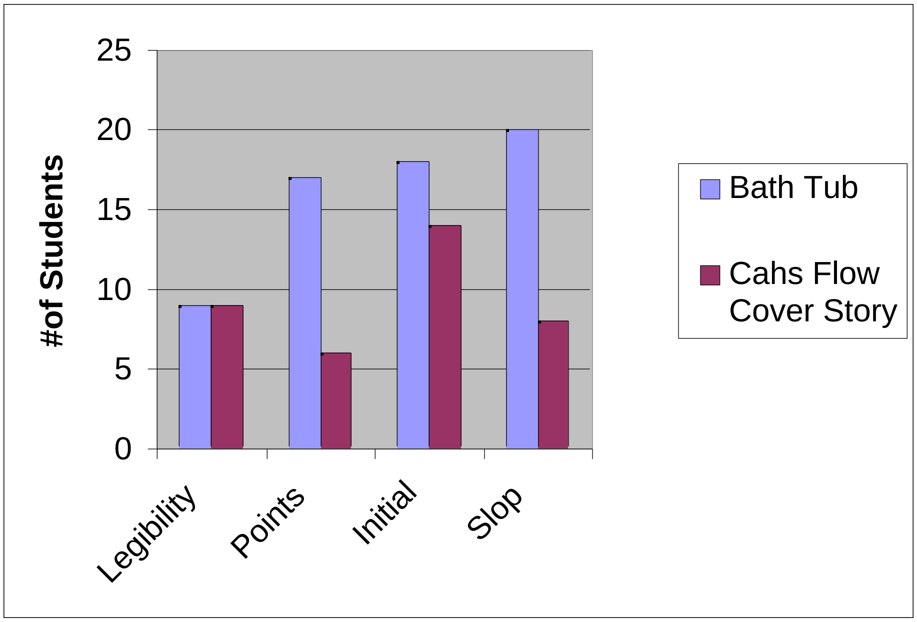
Gordon Kubanek
Ottawa Carleton District Schoolboard
621 Southmore Dr. W.
Ottawa, ON K1V7A4
Phone: 613 733 0610
<kubanek@enoreo.on.ca

Correlation Results to Demographics & Criteria

1. No great difference between male and female
2. No difference between ESL vs English first language [hypothesis: graphical nature meant language not a factor.
3. Student choosing to go into Engineering/Science had better results than those who said they wanted to go into Arts/Social Sciences.
4. Arts/Social Science oriented students were more careless about little things and sloppier [messier].
5. Female work was much neater and easier to follow.
6. Calculus: those who took it performed slightly better, but not by a lot – BUT of those who got everything correct ALL had taken calculus.
7. Age had no effect as all were 17,18, or 19 years of age.
8. The results for inflow that was constant [high, then low, then high; as compared with variable inflow [shaped like teeth] was EXACTLY the same. This means they had NO IDEA on the nature of how flows really effected stocks, except in the most general way.

Hypothesis to explain results:

1. They were not serious as there was no pressure to have a good result
2. We did very little preparation for this: the questions about the people walking in & out of the department store, we did demo actual water going in/out of a tank, and did prepare a small ppt show on stock/flow System Dynamics.
3. Perhaps the situation was not real for them – too abstract – so they did not put all they could into it.
4. We saw little attention to detail and lots of carelessness – they got the 'general' idea but failed at the details. Ie. Vague understanding only



Bathtub Dynamics – Ottawa – Raw Tables of Results

Parameter	% Bathtub correct	% Cash Flow correct
Totally correct understanding: STOCK	**4**	
Totally correct understanding: FLOW	**17**	
Legibility	40	40
Points	80	25
Initial	80	60
Slope	90	35

Student Performance on The Bathtub and Cash Flow Problems

Diana M. Fisher
Wilson High School
1151 SW Vermont St.
Portland, Oregon USA
(503) 916-5280 (phone)/(503) 916-2705 (fax)
dfisher@pps.k12.or.us or dmfisher@aracnet.com

Both the "bathtub" and "cash flow" problems were administered to four classes at Wilson High School in Portland, Oregon in the spring of 2003. Each problem contained the same explanation, diagram, flow graphs, and blank graph pad (for the solution graph) that were written by Linda Booth Sweeney and John Sterman in their study "Bathtub dynamics: initial results of a systems thinking inventory" published in the System Dynamics Review Volume 16 Number 4 Winter 2000.

There was no special preparation given to the students except a verbal explanation of the problem scenario, an explanation of the shape of the flow graphs (but no comment about a comparison of relative position of each graph to the other), and a request to sketch the graph of the accumulation for each task. It was mentioned that the initial value of the accumulation was to be 100. Students were given the tasks and asked to work independently on each task. It took them an average of 10 minutes to complete both tasks. Those who took longer were allowed the extra time, so time to complete the task was not a factor in the results. The students were told that they would not be graded on the tasks, nor would they receive extra credit. They were told it was for a study that the instructor was conducting.

Two of the classes to whom these tasks were administered were advanced algebra classes (students ages 16 to 17). The other two classes were advanced placement (first year) calculus classes (students ages 17 to 18). Both of the advanced algebra classes and one of the AP calculus classes were taught by the author, where system dynamics modeling is used a few times throughout the year. The second calculus class was taught by another instructor. An attempt was made to get a third instructor to administer the tasks to his advanced algebra class, but the administration of the tasks was given to a substitute teacher who did not follow the administration guidelines, so those papers had to be discarded.

The students were also asked to indicate which category below best described their previous experience with STELLA modeling (the synonym the students use to describe system dynamics modeling at Wilson High School). The categories were:
1. I have had no previous experience with STELLA modeling.
2. I have had one class where STELLA modeling was used a few (2 or 3) times.
3. I have had a class were STELLA modeling was used more than 5 time during the year.
4. I have taken a system dynamics modeling class using STELLA.

Comments: ____

The comments section was used to have the students who selected category 2 or 3 list the classes in which they had used STELLA. If a student had used the STELLA software a few time, but had more than one class were it was used a few times in each class, they were instructed to circle category 3.

Numerical Summary:
The same rubric was used to grade the student papers as was listed in the Booth Sweeny – Sterman article. An attempt was made to interpret each item in the rubric in a manner similar to that listed in the article. There could be error introduced in the comparison of the high school results and the MIT results due to potential differences in interpretation of specific rubric descriptions. In a few places a score of 0.5 or 0.75 was used if the student graph indicated an understanding of at least half or more of the concept described but was not fully correct. The high school results were compiled by one person so there should be minimal error introduced in comparing the results of one high school class to another high school class, in this study. The results are summarized in the table below. (Note: AA tot represents the total results for both advanced algebra classes. c1, c2, c3, c4 represent the four categories of previous modeling experience as described in the previous section. Cal A tot represents the total results of a first year calculus class taught by the same instructor as the two advanced algebra classes. Cal B tot represents the total results of a first year calculus class taught by an instructor who does not use SD modeling in the instruction process.)

Analysis of Bathtub Problem

Class	AA tot	AA c1	AA c2	AA c3	AA c4	Cal A tot	C A c1	C A c2	C A c3	C A c4	Cal B tot	C B c1	C B c2	C B c3	C B c4
Number of students	45	2	42	1	0	30	0	19	9	2	16	10	4	0	2
1. When the inflow exceeds the outflow, the stock is rising	.71	0	.74	1.00		.93		.89	1.00	1.00	1.00	1.00	1.00		1.00
2. When the outflow exceeds the inflow, the stock is falling.	.73	0	.76	1.00		.93		.89	1.00	1.00	1.00	1.00	1.00		1.00
3. The peaks and troughs of the stock occur when the net flow crosses zero. (ie t=4,8,12,16)	.69	0	.71	1.00		.93		.89	1.00	1.00	1.00	1.00	1.00		1.00
4. The stock should not show any discontinuous jumps (it is continuous)	.91	0	.95	1.00		1.00		1.00	1.00	1.00	.97	.95	1.00		1.00
5. During each segment the net flow is constant so the stock must be rising(falling) linearly.	.61	0	.63	1.00		.97		.95	1.00	1.00	.94	.90	1.00		1.00
6. The slope of the stock during each segment is the net rate (ie ± 25 units/time period)	.57	0	.58	1.00		.83		.89	.67	1.00	.69	.65	.63		1.00
7. The quantity added to (removed from) the stock during each segment is the area enclosed by the net rate (ie 25 units/time period*4 time periods = 100 units, so the stock peaks at 200 units and falls to a minimum of 100 units)	.59	0	.61	1.00		.83		.89	.67	1.00	.69	.65	.63		1.00

The vast majority of advanced algebra students created an M-shaped graph, correctly. Five students did not create any graph. Three students who entered the class half-way through the year did not draw M-shaped graphs. Their graphs were a mimic of the inflow graph. The majority of mistakes made in the rests of the graphs were that the M was drawn with a slightly curved M shape than an M drawn with straight lines and sharp peaks and valleys. Other mistakes involved drawing a more sinusoidal curve than even M-shaped. Finally, there were errors in drawing the correct height between the peaks and valleys.

Almost all the calculus students in group A drew the correct M-shaped graph. Three students drew and M-shaped graph that was not tall enough. One student did not draw a graph. One student drew two M-shaped graphs instead of one, both of which were too short. Finally one student drew a graph that just repeated the inflow graph.

Most of the mistakes made in the calculus group B papers involved M-shaped graphs that were too short.

Analysis of Cash Flow Problem

Class	AA tot	AA c1	AA c2	AA c3	AA c4	Cal A tot	C A c1	C A c2	C A c3	C A c4	Cal B tot	C B c1	C B c2	C B c3	C B c4
Number of students	45	2	42	1	0	30	0	19	9	2	16	10	4	0	2
1. When the inflow exceeds the outflow, the stock is rising	.32	0	.34	0		.70		.63	.78	1.00	.56	.70	0		1.00
2. When the outflow exceeds the inflow, the stock is falling.	.33	0	.35	0		.69		.63	.78	.88	.56	.70	0		1.00
3. The peaks and troughs of the stock occur when the net flow crosses zero. (ie t=2,6,10,14)	.29	0	.31	0		.68		.62	.78	.75	.52	.65	0		.88
4. The stock should not show any discontinuous jumps (it is continuous)	.84	.50	.86	1.00		.90		.84	1.00	1.00	.94	1.00	.75		1.00
5. The slope of the stock at any time is the net rate.	.36	0	.38	0		.59		.51	.67	1.00	.38	.50	0		.50
6. The slope of the stock when the net rate is at its maximum is 50 units/period (t = 0, 8, 16)	.34		.37	0		.63		.51	.78	1.00	.56	.70	0		1.00
7. The slope of the stock when the net rate is at its minimum is -50 units/period (t = 4, 12)	.34	0	.37	0		.63		.53	.78	.88	.56	.70	0		1.00
8. The quantity added to (removed from) the stock during each segment of 2 periods is the area enclosed by the net rate (ie a triangle with area ± (1/2)*50 units/perios*2 periods = ±50 units.) The stock therefore peaks at 150 units and reaches a minimum of 50 units.	.28	0	.30	0		.50		.42	.61	.75	.41	.50	0		.75

The majority of advanced algebra students who did not draw correct graphs drew graphs that either reflected the input graph or were slightly truncated sinusoidal graphs. There were half a dozen graphs that were not strictly periodic, where the peaks on the second half of the graph were higher or lower than the peaks on the first half of the graph. There were six students who chose not to draw any graph at all.

For the calculus students in group A the mistakes centered around graphs that were straight lines with sharp peaks and valleys rather than sinusoidal. There were two papers that had graphs that were shaped as two U's next to each other. Three students chose not to draw graphs. One student drew two M-shaped graphs next to each other in the upper half of the graphing rectangle.

For the calculus students in group B about a third of the papers showed graphs that were composed of M-shaped lines with sharp points at the peaks and valleys. Some of these had the correct height, others did not. Some of the sinusoidal graphs were not of the correct height. One student did not draw a graph. Another student drew a graph that was a series of horizontal segments, connected to each other, that rose and fell in a shape that could roughly mimic the inflow rise and fall without regard to its relative position to the outflow graph.

Comparison of High School Results with MIT Student Results

This table compares the total advanced algebra students, the total first year calculus students in group A and the total first year calculus students in group B with the MIT students. An additional column is introduced in this chart to include those students who had taken a system dynamics modeling class (listed as c4 students above) and including another small group of four independent study system modeling students whom the author taught this year. (Note: One student is in both a c4 category and a calculus A class, but was counted only once.)

Bathtub Problem

Class	AA tot	Cal A tot	Cal B tot	SD tot		MIT
Number of students	45	30	16	7		182
1. When the inflow exceeds the outflow, the stock is rising	.71	.93	1.00	1.00		.80
2. When the outflow exceeds the inflow, the stock is falling.	.73	.93	1.00	1.00		.80
3. The peaks and troughs of the stock occur when the net flow crosses zero. (ie t=4,8,12,16)	.69	.93	1.00	1.00		.86
4. The stock should not show any discontinuous jumps (it is continuous)	.91	1.00	.97	1.00		.89
5. During each segment the net flow is constant so the stock must be rising(falling) linearly.	.61	.97	.94	1.00		.78
6. The slope of the stock during each segment is the net rate (ie ± 25 units/time period)	.57	.83	.69	.93		.66
7. The quantity added to (removed from) the stock during each segment is the area enclosed by the net rate (ie 25 units/time period*4 time periods = 100 units, so the stock peaks at 200 units and falls to a minimum of 100 units)	.59	.83	.69	.93		.63
Mean for all items	.69	.92	.90	.98		.77

Cash Flow Problem

Class	AA tot	Cal A tot	Cal B tot	SD tot		MIT
Number of students	45	30	16	7		150
1. When the inflow exceeds the outflow, the stock is rising	.32	.70	.56	1.00		.47
2. When the outflow exceeds the inflow, the stock is falling.	.33	.69	.56	.96		.44
3. The peaks and troughs of the stock occur when the net flow crosses zero. (ie t=2,6,10,14)	.29	.68	.52	.89		.40
4. The stock should not show any discontinuous jumps (it is continuous)	.84	.90	.94	1.00		.99
5. The slope of the stock at any time is the net rate.	.36	.59	.38	.86		.28
6. The slope of the stock when the net rate is at its maximum is 50 units/period (t = 0, 8, 16)	.34	.63	.56	1.00		.47
7. The slope of the stock when the net rate is at its minimum is -50 units/period (t = 4, 12)	.34	.63	.56	.96		.45
8. The quantity added to (removed from) the stock during each segment of 2 periods is the area enclosed by the net rate (ie a triangle with area ± (1/2)*50 units/perios*2 periods = ±50 units.) The stock therefore peaks at 150 units and reaches a minimum of 50 units.	.28	.50	.41	.57		.37
Mean for all items	.39	.67	.56	.91		.48

General Results:

One would not expect high school students to be more successful than MIT students at analyzing change behavior for each task. Since the results seem to indicate that most of the high school students, the calculus ones especially, performed comparatively well on these tasks it suggests that perhaps more than one factor might be involved in generating these results. Some of the factors to consider are presented in the conclusion section of this paper.

The modeling group performed extremely well, but with only 7 students in that category one must be cautious about drawing too many conclusions. Certainly it can be said that SD modeling appears to aid the analysis of behavior over time graphs, since the course has as one of its instructional strategies the analysis of flow and accumulation graphs. Even the more difficult task (cash flow) was not too challenging for the SD students.

The calculus students from both classes performed well on the bathtub task. It has become part of the calculus curriculum to analyze flow and accumulation graphs in relation to one another, so this task was well within the experience of most of the calculus students, even though they had not seen a separate inflow/outflow scenario before. The cash flow task was more difficult for the calculus students, although many created graphs that were somewhat sinusoidal. There were ten students who specifically drew the accumulation graph as connected parabolas, as they should be drawn. Too many students did not pay close enough attention to the location of the peaks and troughs of the stock. This should have been easy for them. The study of when an accumulation increases and decreases is a fundamental concept in calculus.

The advanced algebra students had the most difficulty, as might be expected. However, they performed well on bathtub task. In class there were motion detector activities throughout the year that required them to study the relationship between distance and velocity curves for linear and quadratic functions. This may have helped their performance on this task. The cash flow problem was another situation altogether. Six students didn't even draw a graph. Still there were a surprising number of graphs (22) that were drawn that were sinusoidal in nature, even though some of them were not the correct height nor had the correct locations for the peaks and valleys.

Implications/Conclusions:

Before discussing the overall results it is important for the reader to understand the change in mathematics instruction that has been undertaken in the United States over the past five years. Due to changes initially adopted in calculus classes, generated by the calculus reform movement of the past 15 years, mathematics instruction now is supposed to provide multiple representations of function behavior for analysis. The success of the calculus reforms prompted the inclusion of many of the multiple representation instructional strategies in all high school algebra and pre-calculus classes. So students today are much more accustomed to analyzing a variety of graphs.

It appears that the new methods of instruction in mathematics in the United States provide a compatible environment for inclusion of tasks such as those represented by the bathtub and cash flow problems. Also, it is evident that instructional strategies used in the lower grades in mathematics, especially in middle school (with 12 and 13 year old students), is supporting the newer (and broader) approach to function analysis, including graphical analysis, at the high school level (14 to 18 year old students). Although the students, especially the advanced algebra students, had more difficulty with the cash flow problem, their work on the bathtub problem seems to indicate that they don't fall too short of MIT students in their performance. While it would be nice to think that high school students are getting smarter, it is more realistic to look at the learning environment and how it might be different today from the high school environment even three or four years ago for the MIT students. It may be that many MIT students studied mathematics in a more traditional instructional environment. This could be due to the fact that some of the MIT participants were older students, hence a significant delay time is present in the curriculum to which they were exposed compared to the curriculum for the current high school student. Additionally, many instructors who teach upper level mathematics in high school are the more senior staff who may or may not have been willing to change a method of instruction that had been successful for many years. So the MIT students may have had a more traditional calculus class, focusing on proofs rather than more conceptual interpretations of representing change behavior. Also, it is necessary to determine whether the interpretation of the high school student papers was done in a manner similar enough to the MIT papers that the numbers do represent a true comparison of the results. This alone could introduce significant error, negating the results.

That said, it is heartening to think that high school students might be gaining skill in the areas that we hope to see develop in future citizens. The results above, as much as can be concluded based on seven students, do indicate that students with significant exposure to SD modeling are adept at analyzing problems similar to the two represented in the tasks given. The bathtub and cash flow tasks are wonderful examples of problems that students should be encouraged to analyze as part of a normal mathematics course of study. It is hoped that more problems of this type can be designed and included in math curriculum. The change needs to occur in the teacher's minds more than the student's minds. The students are capable of much more sophisticated thinking than we have given them opportunity to exercise in the past. But the process must start in elementary school, and be developed throughout each succeeding year of study.

Bathtub Dynamics in Portland at SyMFEST

Ron Zaraza
Portland Public Schools
2544 SE 16th Ave.
Portland, OR 97202
Phone: 503-916-5280
E-mail:s rzaraza@jps.net

Eighty-Two participants at SYMFEST 2002 in Portland participated in an assessment in which they attempted the three systems related tasks first presented by Linda Booth Sweeney and John Sterman in *Bathtub Dynamics: Initial Results of a Systems Thinking Inventory*. The intent was to compare the performance of high school students (ages 14-18) who had taken courses which either explicitly taught system dynamics modeling or used models extensively in presenting/exploring concepts with the performance of the original test sample.

Participants in the assessment were from six Oregon schools (Roseburg HS, De La Salle North HS, La Salle HS, Franklin HS, Wilson HS, West View HS) and one Vermont school (Champlain Valley Union HS). Their systems experience ranged from less than a semester in which they used but did not build models in a course, to five years of instruction in modeling. Their math background ranged from current registration in Algebra 1 to completion of a course in Ordinary Differential Equations. This broad range, coupled with the small sample size makes it almost impossible to draw any conclusions about combinations of systems and mathematics experience. However, some broad trends appear.

Performance on Bath Tub Task 1

Criterion	Average 1.5 years + in systems classes	Average .5 or 1.0 years in systems classes	Average for all students
1. When the inflow exceeds the outflow, the stock is rising.	0.97	0.75	0.84
2. When the outflow exceeds the inflow, the stock is falling.	0.97	0.75	0.84
3. The stock should not show any discontinuous jumps.	0.97	0.875	0.91
4. The peaks and troughs of the stock occur when the net flow crosses zero.	0.97	0.75	0.84
5. During each segment the net flow is constant so the stock must be rising or falling linearly.	0.97	0.73	0.83
6. The slope of the stock during each segment is plus or minus 25 units/time period	0.88	0.5	0.66
7. The stock peaks at 200 units and falls to a minimum of 100 units.	0.88	0.5	0.66
Mean for all items	0.95	0.69	0.80

Those students with more than a year of systems experience performed extremely well on all tests, regardless of mathematics level. It appears that those in higher level mathematics courses performed better, but the sample size is too small to draw any conclusions. On task 1, student performance was consistently higher than the MIT group, averaging 0.18 higher. This strongly suggests that these students had a well developed mastery of rates and accumulations.

Performance on Cash Flow Task 2

Criterion	Average 1.5 years + in systems classes	Average .5 or 1.0 years in systems classes	Average for all students
1. When the inflow exceeds the outflow, the stock is rising.	0.79	0.42	0.57
2. When the outflow exceeds the inflow, the stock is falling.	0.79	0.42	0.57
3. The stock should not show any discontinuous jumps.	0.97	0.98	0.98
4. The peaks and troughs of the stock occur when the net flow crosses zero.	0.79	0.54	0.65
5. The slope of the stock at any time is the net rate.	0.76	0.40	0.55
6. The slope of the stock when the net rate is at its maximum is 50 units per period	0.68	0.23	0.41
7. The slope of the stock when the net rate is at its minimum is 50 units per period.	0.68	0.25	0.43
8. The stock peaks at 150 units and reaches a minimum of 50 units.	0.68	0.23	0.41
Mean for all items	0.77	0.43	0.57

Task 2 results were even more striking. The average score here was 0.29 higher than the reference group. This task is clearly the most difficult mathematically. Since most of the student group were concurrently registered in a math class at the Algebra 2 level or higher, the fact that they were currently using mathematics may have played a role. However, even those students enrolled Algebra 2 averaged 0.17 higher. Those students would not yet have seen periodic/oscillatory functions in their math classes, so their performance suggests that the systems background may be responsible for their high success rate.

There were some difficulties in scoring Criterion 5 on this task. The four separate "bullets" were interpreted conceptually, not rigorously numerically. A score of 1 was given if the graph of the stock showed the correct shape of behavior, and appropriate changes in slope (increasing, decreasing), even if the numerical values were not correct. If they were regraded and required to have the correct numerical values, the performance would closely match Criteria 6 and 7.

Performance on Order/Production/Inventory Task 3

Criterion	Average 1.5 years + in systems classes	Average .5 or 1.0 years in systems classes	Average for all students
1. Production must start in equilibrium with orders.	1.00	0.90	0.94
2. Production must be a constant prior to time 5 and indicate a lag of 4 weeks in the response to the step increase in orders.	0.94	0.60	0.74
3. Production must overshoot orders to replenish the inventory lost during the initial period when orders exceed production. Production should return to (or fluctuate around the equilibrium rate of 11,000 widgets/week .	0.82	0.67	0.73
4. Conservation of material. The area enclosed by production and orders during the overshoot of production must equal the area enclosed by orders less production.	0.65	0.48	0.55
5. Does production oscillate?	0.00	0.00	0.00
6. Inventory must initially decline.	0.97	0.73	0.83
7. Inventory must recover after dropping initially.	0.97	0.69	0.80
8. Inventory must be consistent with the trajectory of production and orders.	0.68	0.31	0.46
Mean for all items	0.75	0.55	0.63

Task 3 results were the most surprising. That problem is one that is outside the students' experience. The business concepts underlying it are not studied in high school.

This means that the students had to translate the problem into rates and accumulations. When the average score is adjusted to reflect the fact that Criterion 5 was not part of the original scoring guide used in the MIT results, the average score is .85, more than double the score in the reference group.

It seems clear that the high school students with more than a year of systems experience were better able to solve the three problems than the group of students at MIT with no prior exposure to systems.

Students with a year or less of systems experienced more mixed results. Though students in this group included some taking calculus, most were taking lower level math classes. Thus, mathematical inexperience may have played a role in their performance. On the first task they performed slightly more poorly than the overall reference group. If only the subgroup at MIT that did the Bath Tub version of Task 1 is considered, their performance is 0.14 lower.

Task 2 presents a similar pattern, although the difference in performance is smaller, averaging only 0.05. As in the more experienced group, Criterion 5 was scored conceptually rather than looking at mathematical accuracy. Nonetheless, this is an unexpectedly good performance. Few students in this group have completed two years of algebra. Yet their performance on this task compares favorably to that of college graduates and graduate students.

Student performance on Task 3 again presents striking results. This group is generally younger than the first group, suggesting that their knowledge of the business concepts involved in this problem (from their stock of general knowledge, not instruction) should be even more poorly developed. This is reflected by an average performance that (adjusted for Criterion 5) is 0.13 lower than the more experienced group. That, however, is still .31 higher than the MIT reference group. This further reinforces the supposition that it is the transference of stock/flow concepts that allowed the students to perform as well as they did on this task.

Looking at both student groups, the question of a link between mathematical expertise and performance is an obvious one. As already noted, the small sample size restricts confidence in any conclusion. It still seems reasonable to look at comparable subgroups from each of the two student groups. Thirteen of the students with more than a year of systems are currently enrolled in Algebra 2. Fourteen students with a year or less of systems are also in Algebra 2. A comparison of their results suggests that the systems experience may be the factor which differentiates performance on the tasks.

Comparative Performance of Students Registered for Algebra 2 on Bath tub Task 1

Criterion	Average in Alg 2 (1.5 years + in systems classes)	Average in Alg 2 (.5 - 1.0 year in systems classes)
1. When the inflow exceeds the outflow, the stock is rising.	1.00	0.80
2. When the outflow exceeds the inflow, the stock is falling.	1.00	0.80
3. The stock should not show any discontinuous jumps.	1.00	0.93
4. The peaks and troughs of the stock occur when the net flow crosses zero.	1.00	0.80
5. During each segment the net flow is constant so the stock must be rising or falling linearly.	1.00	0.80
6. The slope of the stock during each segment is plus or minus 25 units/time period	0.92	0.53
7. The stock peaks at 200 units and falls to a minimum of 100 units.	0.92	0.53
Mean for all items	0.98	0.74

On Task 1, the more experienced group averaged 0.24 higher on all criteria. The less experienced Algebra 2 students score 0.05 higher than the overall average for less experienced students. On Task 2, the more experienced group score 0.32 higher, nearly twice as high. For Task 3, the difference is 0.10. These relatively significant differences suggest that systems experience is connected to performance on the three tasks for students of similar mathematical preparation.

Comparative Performance of Students Registered for Algebra 2, Cash Flow Task 2

Criterion	Average (1.5 years + in systems classes)	Average (.5 or 1.0 years in systems classes)
1. When the inflow exceeds the outflow, the stock is rising.	0.69	0.27
2. When the outflow exceeds the inflow, the stock is falling.	0.69	0.27
3. The stock should not show any discontinuous jumps.	0.92	0.93
4. The peaks and troughs of the stock occur when the net flow crosses zero.	0.69	0.27
5. The slope of the stock at any time is the net rate.	0.62	0.27
6. The slope of the stock when the net rate is at its maximum is 50 units per period	0.54	0.20
7. The slope of the stock when the net rate is at its minimum is 50 units per period.	0.54	0.20
8. The stock peaks at 150 units and reaches a minimum of 50 units.	0.54	0.20
Mean for all items	0.65	0.33

Comparative Performance of Students Registered for Algebra 2 on Order/Production/Inventory Task 3

Criterion	Average for students with 1.5 years or more in systems classes	Average for students with .5 or 1.0 years in systems classes
1. Production must start in equilibrium with orders.	1.00	1.00
2. Production must be a constant prior to time 5 and indicate a lag of 4 weeks in the response to the step increase in orders.	0.92	0.60
3. Production must overshoot orders to replenish the inventory lost during the initial period when orders exceed production. Production should return to (or fluctuate around the equilibrium rate of 11,000 widgets/week .	0.69	0.67
4. Conservation of material. The area enclosed by production and orders during the overshoot of production must equal the area enclosed by orders less production.	0.46	0.53
5. Does production oscillate?	0.00	0.00
6. Inventory must initially decline.	0.92	0.80
7. Inventory must recover after dropping initially.	0.92	0.73
8. Inventory must be consistent with the trajectory of production and orders.	0.54	0.31
Mean for all items	0.68	0.58

These results would be far more convincing with a larger sample. However, the number of secondary students working with systems is still very small. The new tasks will most certainly be administered at SYMFEST 2003, with tighter control of the administration. If it can also be given at the similar event at WPI, the sample size should be large enough to begin drawing conclusions.

Bathtub Dynamics at WPI

James M. Lyneis, Professor of Practice, Worcester Polytechnic Institute
Debra A. Lyneis, Creative Learning Exchange

Jim Lyneis
Worcester Polytechnic Institute
PO Box121, 215 Landgrove RD
Weston, VT 05161
Phone: 802 824-4219
E-mail:jmlyneis@alum.mit.edu

Deb Lyneis
Creative Learning Exchange
PO Box 121, 215 Landgrove RD
Weston, VT 05161
Phone: 802 824-4219
E-mail: lyneisd@clexchange.org

Tasks Administered

We explored students' understanding of stocks and flows using the following "bathtub dynamics" tasks:

- The Department Store Task, Treatment 1
- Bathtub Task 1, square wave
- Bathtub Task 2, sawtooth wave

Jim Lyneis administered the tasks to students in his classes at WPI, primarily out of curiosity. He wanted to see how their performance would compare to that of the Sloan School graduate students in the study by Booth Sweeney and Sterman (2000). Later, when the Creative Learning Exchange suggested gathering and comparing bathtub dynamics results from a broad range of students, he donated his raw data to the cause. While bathtub dynamics tasks were not administered as a complete research study at WPI, they do yield some limited interesting results.

Protocol and Circumstances

Students in two different classes did the tasks. The first was a class of approximately 80 undergraduate students taking Microeconomics in the fall of 2003. (This course is taught using a mix of system dynamics and traditional economics.) Eighty students completed the Department Store Task. Forty nine students also did the Bathtub Task 1 (square wave), while the remaining 31 students did Bathtub Task 2 (sawtooth wave). Students completed the tasks in class at the beginning of the term. They were not graded, and they had as much time as they needed to finish – 10-15 minutes. Later in the term, students did the sawtooth exercise again as part of a graphical integration assignment followed by class discussion and explanation (the second results were not tabulated.)

The second class was Introduction to System Dynamics in the spring of 2003. This time, 14 students completed the Department Store Task and the Bathtub Task 2 (sawtooth wave) in class near the end of the term. Again, students were not graded and had plenty of time to finish. No students were in both classes.

Also, although all 94 students did the department store problem, no students did both the square wave and the sawtooth wave problem. At the time, Jim was not conducting a formal study. In the first class of 80 students, he gave 49 students Task 1 and 31 students Task 2, just to see how they did. Students in the second class, Intro System

Dynamics, did only Task 2 because the square wave was felt to be too easy for the class given that a graphical integration lecture and assignments had been given earlier in the term.

The Students

The students were all undergraduates at Worcester Polytechnic Institute, a competitive four year engineering college in Worcester, Massachusetts. Students completed a biographical survey listing their age, gender, class, major, country of origin and first language. As undergraduates, their ages ranges from 18 to 22, with a fairly even distribution of ages within each of the two courses. Reflecting the larger WPI population, the students were overwhelming male, with only 13 females among the total of 94 students. Students were also primarily engineering majors, with a few math and science majors. We grouped all these together with the assumption that engineering and science majors had similar math interests, backgrounds, and basic course requirements. Of the 94 students, 15 students had majors in Management and 2 had majors in Social Sciences, but even these students would most likely have started out applying to WPI for engineering or science. Except for 5 students, all were from the United States; only one student spoke English as a second language. In conclusion, the WPI students presented a fairly homogeneous group.

Students also reported on whether they had played the Beer Game or taken a previous system dynamics course. None of the 80 microeconomics students had any prior experience with system dynamics, nor had they played the Beer Game. Students in the Introduction to System Dynamics course had played the Beer Game in class.

All WPI students are required to take at least two social sciences courses to fulfill distribution requirements (system dynamics courses are a part of the social science department, along with economics, psychology, and public policy). Students can major in system dynamics, but this is still a very small number of students. For the most part, the students performing the bathtub tasks in the two courses were a wide range of engineering students fulfilling their social sciences requirement.

Results - Department Store Task

Performance on the first two department store questions was much better than on the last two. In fact, of the 80 microeconomics students, only one student answered question 2 incorrectly and one other student answered both 1 and 2 incorrectly. In the system dynamics class, all students answered both correctly. Nearly everyone could interpret the flow graphs.

Microeconomics Course, Beginning of course, Fall 2002,

	Dept. Store Treatment 1			N=80	
	Q1	**Q2**	**Q3**	**Q4**	**score**
Average	0.9875	0.98	0.31	0.28	0.64
Std Dev	0.1118	0.16	0.47	0.45	0.23

The second two questions about accumulations over time were more difficult, however. In the microeconomics class, averages on questions 1 and 2 were .32 and .28, respectively, a dismal showing. Some incorrect responses noted that there was not enough information; others gave the points of the lowest and highest flows. Most students either got all 4 questions correct or missed both of the last two. The mean on all four questions for the entire class was .64 with a standard deviation of .23.

Introductory System Dynamics Course, Spring 2003

	Dept. Store Treatment 1			N=14	
	Q1	**Q2**	**Q3**	**Q4**	**score**
Average	1.00	1.00	0.50	0.36	0.71
Std Dev	0	0	0.52	0.50	0.24

In the introductory system dynamics course, results on questions 3 and 4 were better, with means of .50 and .36, but these should have been higher for students who had studied stocks and flows. (It would be interesting to ask these students what they had not understood about the problem, if the class had not already dispersed.)

Results – Bathtub Task 1

On the square wave bathtub problem, the 49 students in the microeconomics class scored an average of .79 with a standard deviation of .34. Students had the most difficulty with coding item 7, calculating the value of the stock. They had the least problem with discontinuities. Of the 49 students, 31 had perfect responses across the coding items. Those who got it, got it; the rest floundered.

		Bathtub Task 1		N= 49				
	1	**2**	**3**	**4**	**5**	**6**	**7**	**score**
Average	0.84	0.80	0.92	0.78	0.82	0.73	0.67	0.79
Std Dev	0.37	0.41	0.28	0.42	0.39	0.45	0.47	0.34

Results – Bathtub 2

The sawtooth wave was much more difficult. The 31 students in the microeconomics class averaged .48 with a standard deviation of .41. Again, the worst performance was on item 7 calculating the value of the stock, .35. Of the 31 students, 8 had perfect scores, with another 4 missing only the last item (usually by shifting the curve up or down, or making minor arithmetic errors). Whereas for Bathtub 1 most students produced graphs that somewhat approximated the correct answer, for Bathtub 2 the wrong answers were all over the map. Most of the wrong answers were linear graphs, missing entirely the changing slopes. All students had the least problem with discontinuities.

		Bathtub Task 2		N=31					
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**score**

Average	0.45	0.45	0.84	0.48	0.39	0.42	0.42	0.35	0.48
Std Dev	0.51	0.51	0.37	0.51	0.50	0.50	0.50	0.49	0.41

The 14 students in the introductory system dynamics class fared somewhat better, scoring .76 with a standard deviation of .34. Again the arithmetic, item 7, had the most mistakes with an average of .57, but this compares favorably to the .35 of the microeconomics class. Of the 14 students, 6 had perfect scores and another 4 missed only the last arithmetic item. Of the remaining 4 students with errors, one student produced a linear graph, but all the rest were oscillating waves, albeit misplaced.

		Bathtub Task 2		N=14	(Intro SD course)				
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**score**
Average	0.79	0.79	1.00	0.79	0.71	0.71	0.71	0.57	0.76
Std Dev	0.43	0.43	0.00	0.43	0.47	0.47	0.47	0.51	0.34

Controlling for Age

Since age was the only major difference among the students, we wondered if student performance would improve with age as students had taken more college level math and engineering classes. We looked only at the larger microeconomics class because 14 students in the system dynamics class made too small a sample.

Department Store			**Bathtub 1**			**Bathtub 2**		
Age	**N**	**Average**	**Age**	**N**	**Average**	**Age**	**N**	**Average**
18	24	0.59	18	16	0.71	18	8	0.47
19	16	0.73	19	8	0.84	19	9	0.60
20	20	0.63	20	12	0.98	20	8	0.42
21	13	0.64	21	11	0.68	21	6	0.38

Age did not consistently improve performance. For some unknown reason, 19 year olds, sophomores, outperformed all other classes. On the most difficult task, Bathtub 2, seniors did the worst, .38. Although the sample size is small, we can speculate that seniors who have waited until their very last term to satisfy graduation distribution requirements may not be the best research specimens!

Discussion

Overall, the performance of the WPI students was similar to that of the Sloan students on both bathtub tasks with an average on Bathtub 1 of .79 (.34 standard deviation) and on Bathtub 2 of .48 (.41 standard deviation). Sloan students achieved .83 and .46, respectively. Larger samples and a more complete implementation of all the tasks would yield a better basis for comparison.

Within WPI, students had more difficulty with the last two questions of the department store problem and with the sawtooth wave bathtub problem. The students who had just completed the introductory system dynamics course performed better than their peers in microeconomics, but again the samples are small and the administration of all the tasks to both groups of students was incomplete. Although they did better than their peers, we would have expected all of the system dynamics students to score perfectly on the department store problem because they had studied stocks and flows. Another time, we would debrief the students to understand their confusion. These students had a much better grasp of the graphical integration in Bathtub 2 than their peers did, however. They appear to have learned something (at least most of them did, anyway)!

In class, many of the students were engaged in the system dynamics work, doing their best, seeking help when they needed it, and making good progress. A few of the students were less committed, just fulfilling their distribution requirements. These students performed poorly in class, and presumably on the tasks. Yet, all students at WPI come with an interest and background in math/engineering, so they should have been able to do better. In the MIT sample, there probably were not many laggards!

Implications

It appears that even among strong math and engineering students, there is confusion on basic concepts and skills such as graph reading, distinguishing and interpreting stocks and flows, and graphical integration. Many do not have a strong intuitive sense of how a flow accumulates a stock over time. It will be interesting to compare results with those of liberal arts students who would presumably be less math-inclined, and with younger K-12 students.

The broader question is how to teach these skills to students. We agree with Sterman that a good system dynamics education must include the underlying math and theory. Yet, how do we also build the broader problem-solving, systemic thinking/acting skills and attitudes that students will need to deal with dynamic complexity. Most approaches to systems education, especially K-12, include little math, much less any graphical integration. Is this OK in the long run? Students miss out on a deeper understanding based on the elegant math underlying the practice of system dynamics. Yet, if, as we see with bathtub dynamics, people are thrown off by the simplest stock/flow problems, how do we move forward without scaring them away. Also, how do we establish that graphical integration does actually transfer to thinking and acting systemically? Big issues. "Bathtub dynamics" gives us all a chance to think about them.

Booth Sweeney, L., Sterman J., Bathtub Dynamics: Initial results of a systems thinking inventory, *System Dynamics Review*,**9**(2): 249-286.

Back to the Top