Can people learn behaviours of stock and flow using their ability to calculate running total? An experimental study

Tony L. K. Phuah

System Dynamics Group, University of Bergen, Norway

Tony.Phuah@student.uib.no

Abstract

Stocks and flows are basis of dynamics. Understanding of stock and flow is crucial in comprehending and managing problems such as global warming and national debt. Yet previous experimental studies have found that people perform poorly in simple stock-flow tasks. However, many do have a notion of accumulation in terms of calculating running total, adding or subtracting items to keep track of a running tally. Here a pre-test-treatment-post-test experiment was designed to test the hypothesis that people's understanding of stock and flow behaviours will improve after being asked to reflect on a cognitive conflict, generated by utilizing their running total calculation. Comparisons with a conventional approach to teach stock and flow dynamics and without teaching were also done. Results show that improvements were not significant; the hypothesis lacks support. On the other hand, the conventional approach produced significant improvement. Possible explanations of the results and their implications for education on dynamics, communication of complex dynamic problems and policy insights are discussed.

Key words: *stocks and flows, accumulation, graphical integration, Stock-Flow-Thinking, System Dynamics education, misperceptions of dynamics, dynamic complexity, dynamic decision making*

1 Introduction

Groundbreaking experiment by Sweeney and Sterman (2000) had astonished System Dynamics (SD) community by finding general poor (or no) understanding of stock and flow, even among highly educated people. This finding testified the static mental model hypothesized earlier (Moxnes, 1998). Stock and flow is the most basic foundation of dynamics and is supposed to be well understood for people to deal with dynamic complexity of the world. Yet this finding has been replicated (Ossimitz, 2002) and it is robust (Cronin, Gonzalez, & Sterman, 2009), irrespective of the ways of information display, number of data points, cover stories and incentive given or not.

Cronin et al. (2009) also confirmed the conjecture that many people use *correlation heuristic* or *pattern matching* on stock-flow task, assuming the most salient flow directly and instantaneously influences the stock. This may be related to human tendency to think of cause must have immediate and direct effect, but this mentality overlooks the fact that flow-affects-stock causal relationship is accumulative by nature. The laws of accumulation dictate the distinctive behaviours of flows and their resulted stock, as depicted by Figure 1.

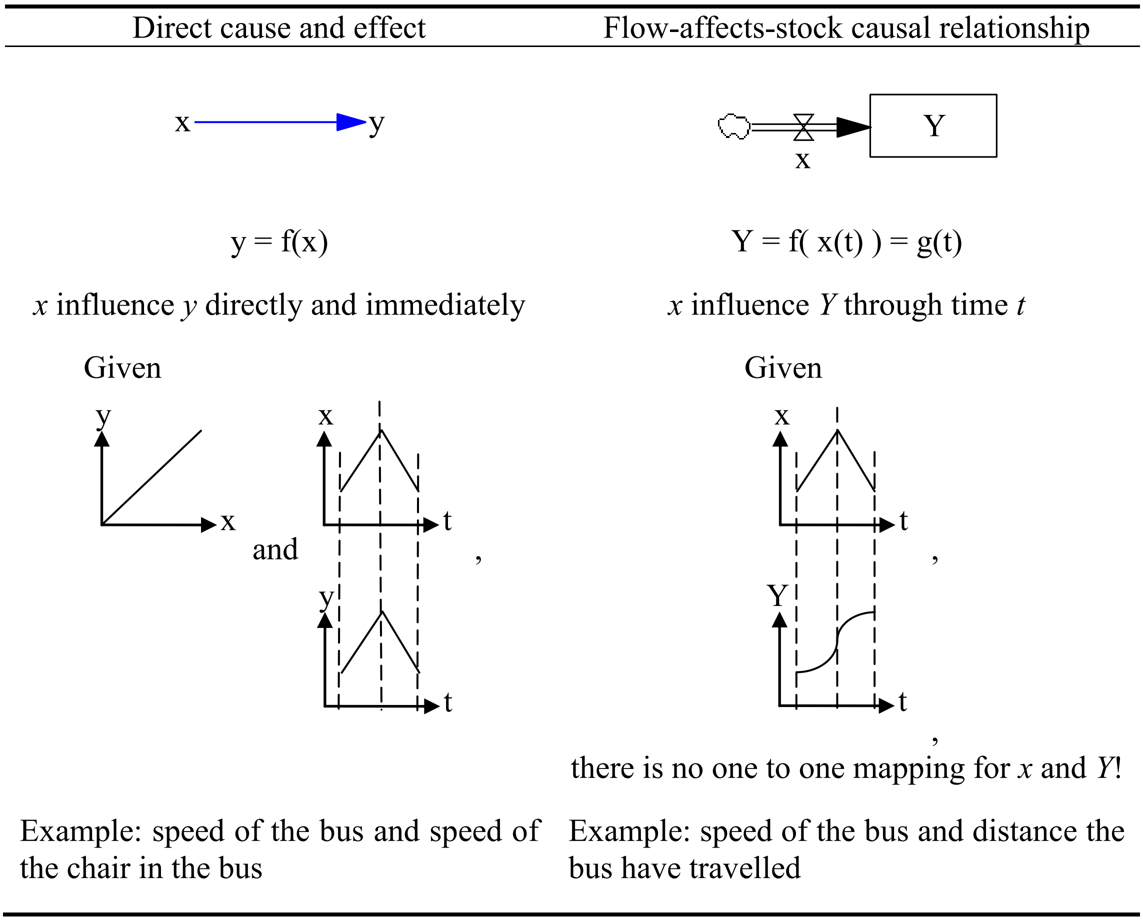


Figure 1 Difference between direct and flow-affects-stock causal relationship

Unfortunately, this gap of understanding creates a challenge for SD educators and consultants when explaining the behaviours of stock and flow. Without a proper model of stock and flow, people will have wrong expectations of behaviours, which can have serious policy implications. Notable examples include misperceptions of climate change mechanism (Sterman & Sweeney, 2007), renewable resources management (Moxnes, 1998), debt-deficit relationship (Ossimitz, 2002) and fallacious congestion indicators in the criminal justice reform (Olaya, Diaz, Ramos, & Pabon, 2008).

These consequences motivate the following questions: ***How to improve people's understanding of stock and flow?*** How to empower people to identify stocks and flows in everyday life and reason about their behaviours?

Few studies have attempted to address this question. Kainz and Ossimitz (2002) used a 90-minute "crash course" introducing stock and flow concept and reported improvement. Among 64 subjects who chose to continue post-test tasks 4 weeks later, 67% were correct in post-test Graphic Hospital (GHP) task, compared to 19% in pre-test Graphic Parking Lot (GPL) task; mean score for pre- and post-test pair Surge Tank / Bath Tub

(ST/BT) Task were 0.36 and 0.54. However, selection bias might have occurred, since not all subjects went through both tests. Pala and Vennix (2005) employed 13 weeks introductory SD course and the results turned out to be ambiguous – significant improvement in department store task (success rates on Question 3 and 4 were improved from 38% to 60% and from 27% to 45% respectively, number of subjects, N = 163), surprisingly worse for manufacturing task (average performance fell from 81% to 73%, N = 107) and mixed in CO_2 zero emissions task (CO_2 trajectory post-test result 76% is a bit higher than pre-test 71%, but global mean temperature trajectory decreased from 31% to 24%, N = 70). On the other hand, Sterman (2009) obtained favourable result (only 25% responded incorrectly, compared to 46% in Cronin et al. (2009) Experiment 5) from his half-semester introductory SD course, showing wide variations of the effects of SD education on improving stock and flow understanding.

These inconclusive effects indicate the need to move towards systematically designed instruction in SD education, based on well tested instruction design principles, so that intended learning outcome is reproducible. This study takes a first step in this direction, exploring effectiveness of two different teaching approaches to imparting graphical integration skill. Graphical integration requires the ability to deduce behaviour of stock based on information about behaviours of its inflows and outflows. Therefore, this should be a good starting point for people to learn about stock and flow.

The first approach is a variant of the current way of teaching graphical integration, which involves guiding learners through the steps of doing graphical integration (like the one in Table 7-2, Sterman, 2000, p. 236) using some exercises. These exercises also enable practices and feedbacks, two elements that are frequently emphasized by educators.

The second approach is developed from a novel idea of trying to connect people's existing understanding of accumulation with learning stock and flow behaviours. Accumulation is a universal process and many people do have basic ideas about it. As Forrester (2009) commented, "Any child who can fill a water glass or take toys from a playmate knows what accumulation means." People do know how to keep a running tally and add or subtract things to calculate running total. By building upon people's prior knowledge of running total, there should be a potential to design effective instruction to promote deeper understanding of accumulation. Cognitive conflict and reflection, a common strategy to foster conceptual change, is used for this purpose.

Minute details of my implementation of these two teaching approaches will be described by Section 2. Experiment results in Section 3 show that the first approach makes significant gain in the graphical integration task, but improvement yielded by the second approach is unapparent. I will discuss the possible reasons that may explain the results and illuminate their implications for SD education in Section 4. Finally, Section 5 concludes this study and suggests further research that can help to advance the profession of SD education.

2 Experimental design

This experiment uses a pre-test-treatment-post-test design; the treatments are the two teaching approaches. The pre-test and post-test are derived from the same task, so that improvement of performance can be identified. To address the possibility that improvement in post-test could be the result of increasing experience after the pre-test, a treatment where no teaching is given, is added to serve as control group. Figure 2 shows the organization of the experiment.

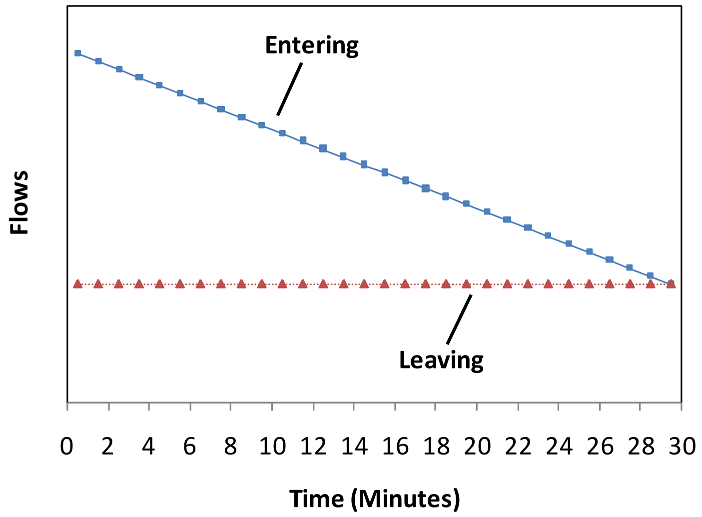
Pre-test	T1 Base No teaching.	Post-test
	T2 Graphical integration guidance Step-by-step queries guiding subject to do graphical integration.	
	T3 Running total and reflection Compare expected answer with calculated running total and reflect on the reason of discrepancy (if any) to discover laws of accumulation.	

Figure 2 Organization of the experiment

2.1 The task

To measure how well the subjects do graphical integration, I use the graphical department store task, first introduced by Cronin et al. (2009) Experiment 5, and also used by Sterman (2009) as post-test. This allows comparison to their results. Its discrete characteristic also eases the calculation of the running total. The only adaptations are a change of wording from “department store” to “supermarket” and an additional sentence to further explain the task. These changes were motivated by subjects’ questions in pilot experiment. These adaptations helped to reduce frequency of questions raised. Figure 3 presents the adapted task for one particular set of flows.

The graph below shows the number of people **entering** and **leaving** a supermarket over a 30 minute period.



In the space below, graph the number of people in the supermarket over the 30 minute interval. You do not need to specify numerical values. The dot at time zero shows the initial number of people in the supermarket.

In other words, draw a line or curve to show how the number of people in the supermarket changes over the 30 minute interval, starting from the black dot (●) in the space below.

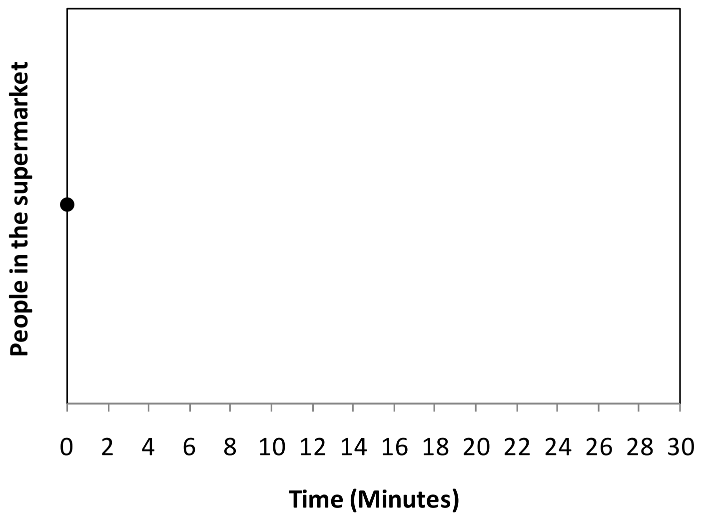


Figure 3 Adapted graphical department store task used in this experiment

From the stock and flow perspective, the underlying model of the task is extremely simple, as portrayed by Figure 4. There is only one stock, one inflow and one outflow. No feedback, delay or nonlinearity is involved. The key to solve this task is first to recognize that the number of people in the supermarket is a stock while entering and leaving are its inflow and outflow. Then one must infer behaviour of the stock based on how flows develop. The number of people in the supermarket must increase when entering is larger than leaving, stay the same when number of entering equals leaving, decrease only when leaving is greater than entering. The rate of increase or decrease in the stock depends on the gap between entering and leaving.

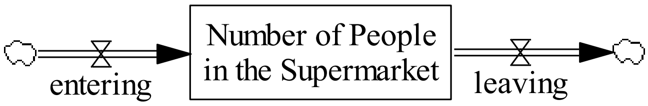


Figure 4 Stock and flow structure of department store task

2.2 Pre-test and post-test

All treatments are identical with respect to pre-test and post-test. Pre- and post-test differ only in their entering and leaving flow graphs. Figure 5 lists pre- and post-test questions together with their correct responses. The pre- and post-test are devised to be of equal complexity – constant inflow and outflow (and thus net flow) in Question 1; constant net flow (despite changes in inflow and outflow) in Question 2; and constant outflow and linear inflow in Question 3. Flow graphs of post-test are inverted versions of their pre-test counterparts. They are arranged in a sequence of increasing difficulty level to encourage learning.

2.3 Treatments

As exhibited by Figure 2, the experiment has three treatments. All treatments were administrated using paper and pen.

2.3.1 T1 The base treatment

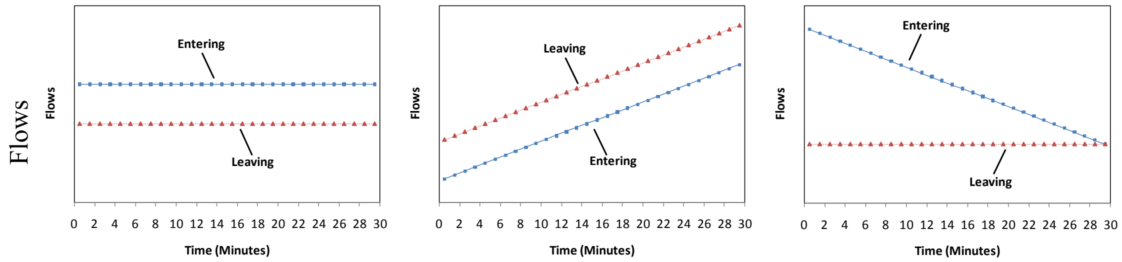
Since the subjects do not receive any teaching, they will only go through pre-test and post-test, total 6 questions. Pre-test must be finished before subject can start the post-test. Post-test will be given after subject completed pre-test and indicate so by raising his/her hand. The difference in performance of pre-test and post-test can then be identified, if any.

Pre-test

1. Constant Flows; I > O

2. Linear increase in both I and O, Constant Net Flow; O > I

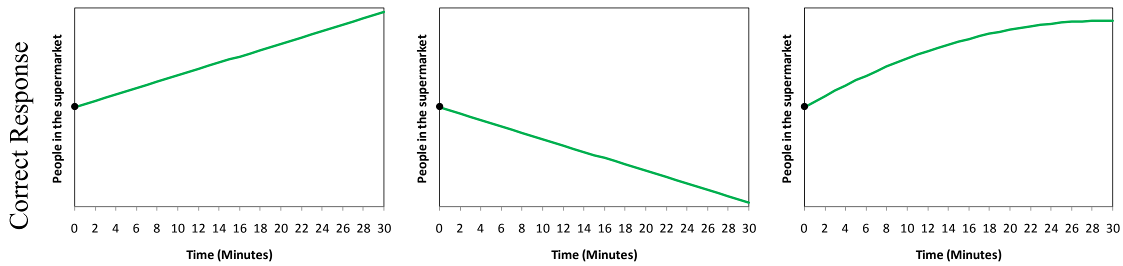
3. Constant Outflow, Linear decline in Inflow; I ≥ O



Net flow is constant and > 0. Stock rises linearly.

Net flow is constant and < 0. Stock falls linearly.

Net flow > 0, falls linearly to 0 by t = 30. Stock rises at decreasing rate, reaches equilibrium at t = 30.

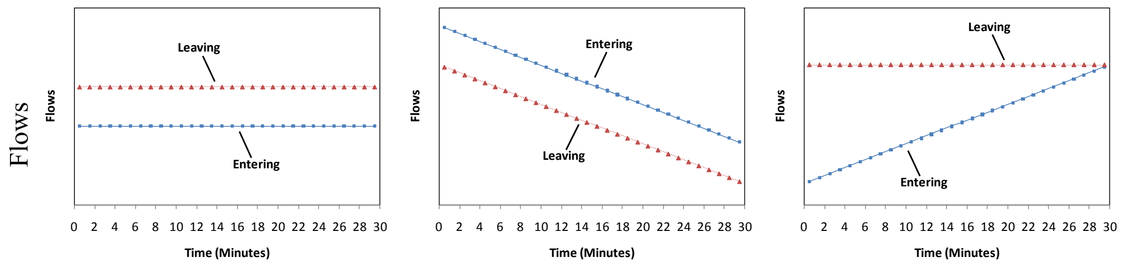


Post-test

1. Constant Flows; I < O

2. Linear decline in both I and O, Constant Net Flow; I > O

3. Constant Outflow, Linear increase in Inflow; I ≤ O



Net flow is constant and < 0. Stock falls linearly.

Net flow is constant and > 0. Stock rises linearly.

Net flow < 0, rises linearly to 0 by time 30. Stock falls at decreasing rate, is constant at t = 30.

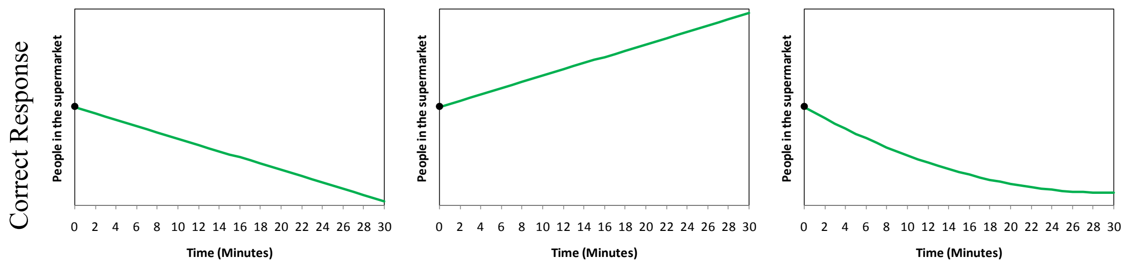


Figure 5 Flow graphs for the pre- and post-test questions and their correct responses

2.3.2 T2 Graphical integration guidance

To figure out how the number of people in the supermarket changes, one needs to ask three basic questions, about its net flow, direction of change, and shape of the change: Is the net flow positive, negative, or zero? Is the stock increasing, or decreasing? Is the stock increasing or decreasing faster and faster, slower and slower, or at a constant rate? These questions are essentially a simplified version of standard steps of graphical integration (Table 7-2, Sterman, 2000, p. 236).

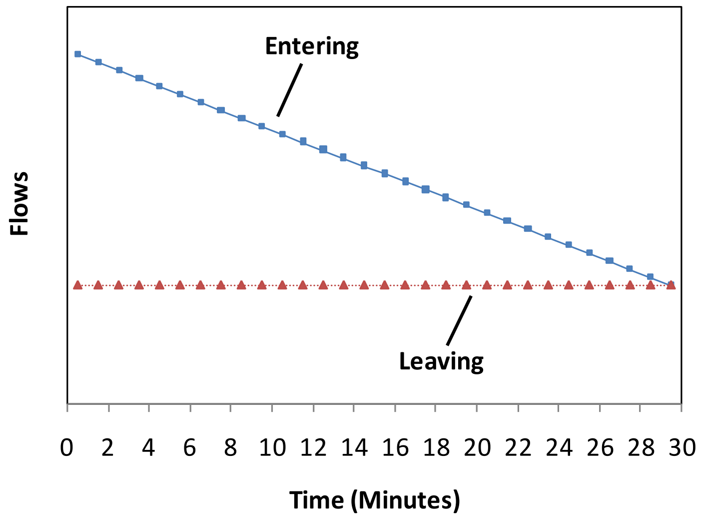
The graphical integration guidance uses these questions to design a step-by-step guided work-out exercise, as presented in Figure 6. There are three such exercises in the treatment, each based on one pre-test question. After subjects complete each exercise, feedback is given to inform whether their answers are correct or not. Drawing is judged qualitatively correct if both direction and shape are correct. If the drawing is wrong, correct trajectory will be drawn. Checking on the subjects' corresponding pre-test answers are also done on the spot. Subjects can then learn from mistakes (if any) before they move on to the next question. This outcome feedback is given after every exercise to maximize potential for learning.

2.3.3 T3 Running total and reflection

Prompted by Experiment 4 of Cronin et al. (2009), the idea of calculating running total of a stock should be familiar to most people, even though they may not know what is stock and flow and infer their behaviours intuitively. The obvious question is then, *how to make use of people's ability to calculate running total to improve their understanding of stock and flow?* Or more specifically in this case, to teach graphical integration? Inspired by Moxnes and Saysel (2009), this treatment utilizes people's ability to calculate running total to evoke cognitive conflict, then assists people to resolve this conflict by carefully crafted instructions for reflection. The rationale behind this is that when people's expected answers are different from the correct answer calculated by themselves, the cognitive conflict induced should be the greatest, demolishing people's overconfidence, which is widely found to be an impediment to learning stock and flow (Cronin, et al., 2009; Moxnes & Saysel, 2009). Hopefully the deep reflection triggered will spark off the discovery of laws of accumulation.

In this treatment, subjects are asked to both calculate and plot the graph of number of people in the supermarket, which necessitates the use of the running total method. In other words, subjects have to determine the number of people in the supermarket at next point of time by adding net inflow in the ensuing period to the existing number of people in the supermarket ($S_t = I_t - O_t + S_{t-1}$). The entering and leaving graphs are the same as in the pre-test, but numbers are labelled to enable numerical accumulation. Number of data points is kept to minimum to reduce cognitive burden of the required computations. Figure 7 demonstrates the so called calculation instruction.

The graph below shows the number of people **entering** and **leaving** a supermarket over a 30 minute period.

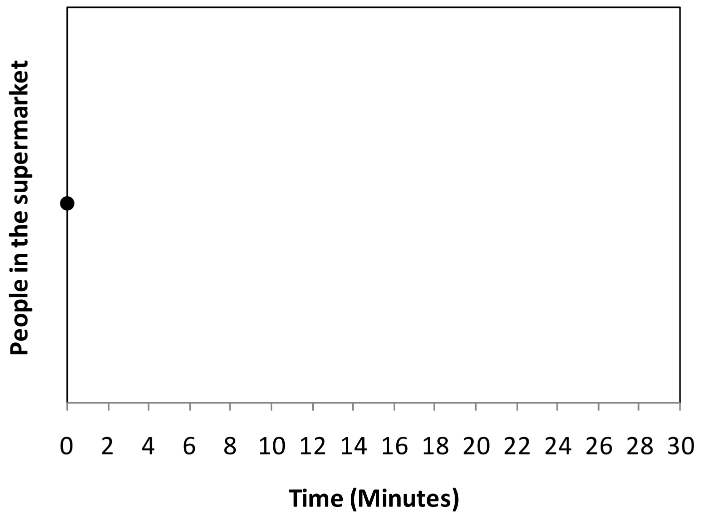


a. Which flow is larger? ☐ Entering larger than leaving ☐ Leaving larger than entering ☐ Both same

b. In what direction will the number of people in the supermarket change?
☐ Increase ☐ Decrease ☐ No change

c. (Skip this if no change) How will the change in the number of people be?
☐ increase or decrease at a faster and faster rate ☐ increase or decrease at a slower and slower rate
☐ increase or decrease at a constant rate

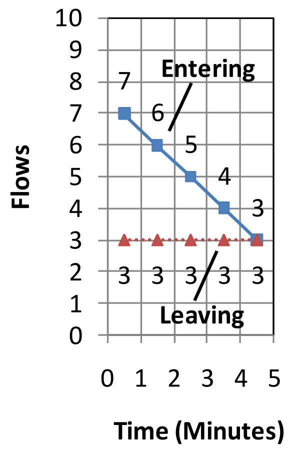
d. In the space below, graph the number of people in the supermarket over the 30 minute interval. You do not need to specify numerical values. The dot at time zero shows the initial number of people in the supermarket.



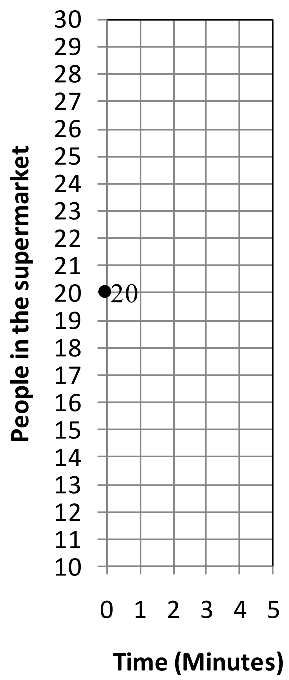
Please raise your hand after having completed this question. Do not continue to the next question before administrator has commented on your answer.

Figure 6 One of the graphical integration guidance exercise

The graph below shows the number of people **entering** and **leaving** a supermarket each minute over a 5 minute period.



In the space below, graph the number of people in the supermarket over the 5 minute interval. The dot shows that at time zero there are 20 people in the supermarket. *Calculate how the number of people in the supermarket develops from minute to minute and plot the numbers accurately in the graph.*



Please raise your hand after having completed this question. Do not continue before administrator has commented on your answer.

Figure 7 Calculation instruction in Running total and reflection treatment

Subjects have received feedbacks on their plotted graphs before they proceed to reflective instruction. The graph must have all points correct, otherwise it is considered wrong, and the correct graph will be shown. This rigorous checking is necessary to generate the needed cognitive conflict for learning. Subjects may know how to do it accurately, but fail to do so because they want to finish quickly. This feedback urges them to answer more carefully, and increases their chances of learning.

After calculation checking, subjects' previous answers during pre-test are returned and checked by referring to the correctly calculated answers. Subsequent instruction (Figure 8) calls subjects' attention to the equivalence of pre-test and treatment questions. If subjects' expected answers in the pre-test were different from the correctly calculated answer, the declaration of error should be immediately thought-provoking. If not, the "Why?" queries followed should indirectly request subjects to reflect, in order to give explanations. The queries are split into two parts to help subjects notice characteristics of the stock's behaviour by concentrating on one characteristic at a time. If subjects' pre-test answers were correct, they can skip the queries ands continue to the next question. This is to prevent that subjects who have already mastered the principle of accumulation, feel bored about unnecessary reflection. This could degenerate their performance on the post-test later.

Now check your answer to Part 1 Question 3.

Note that the entering and leaving graphs in Part 1 Question 3 and Part 2 Question 3 are the same. If your answer to Part 1 Question 3 is different, it must be **wrong**.

i. (Skip this if your answer to Part 1 Question 3 was right)
Why should the number of people in the supermarket **increase**?

____
____
____

ii. (Skip this if your answer to Part 1 Question 3 was right)
Why should the number of people increase **at a slower and slower rate**?

____
____
____

Figure 8 Reflective instruction in Running total and reflection treatment
Part 1 and Part 2 in this figure indicate pre-test and calculation instruction in the Running total and reflection treatment.

2.4 Hypotheses

My primary hypothesis is that the performance in Graphical integration guidance and Running total and reflection treatments will yield better results that the base treatment.

The performance is quantified by the sum of subjects' correct answers to pre-test and post-test questions (separate count for pre-test and post-test). So a subject's pre- or post-test scores could vary from 0 to 3. Two different types of improvement of the scores can then be identified: the Breadth of Improvement (BI) and the Depth of Improvement (DI).

The BI measures how wide of the improvement reached by the subjects, by calculating the proportion of subjects in the treatment who scored higher in post-test than in pre-test:

$$BI = \frac{Number\ of\ subject\ who\ improved}{Total\ number\ of\ subjects}$$

Equation 1 Breadth of improvement

The DI gauges how deep the effect of treatments was on the subjects, constrained by the maximum potential effect. The *normalized change* proposed by Marx and Cummings (2007), is adopted as DI indicator. If subject's post-test performance improved from pre-test, Equation 2a will be used; if the performance worsened, an analogous expression, which is the ratio of the actual decrease to the maximum possible decrease (Equation 2d), will be used. If the subject's pre-test score equalled post-test score, no improvement or decrease in performance, DI = 0 (Equation 2c) except when the subject earned perfect score on both pre-test and post-test. In the latter case this subject's score will be excluded from the analysis for the reason that the subject's performance is beyond the scope of the measurement instrument. Likewise, a subject who scores 0 on both the pre-test and post-test should also be excluded from the analysis (Equation 2b).

$$DI = c = \begin{cases} \dfrac{\%post - \%pre}{100 - \%pre} & \%post > \%pre \quad (a) \\ Drop & \%post = \%pre = 100\% \text{ or } 0 \quad (b) \\ 0 & \%post = \%pre \quad (c) \\ \dfrac{\%post - \%pre}{\%pre} & \%post < pre \quad (d) \end{cases}$$

Equation 2 Depth of improvement

Hence more precisely, the null hypotheses of the experiment are:

H1: There is no significant difference in the BI between treatment T1 and T2.

H2: There is no significant difference in the BI between treatment T1 and T3.

H3: There is no significant difference in the DI between treatment T1 and T2.

H4: There is no significant difference in the DI between treatment T1 and T3.

2.5 Method

I conducted two experimental sessions with 53 undergraduate students from the University of Bergen, Faculty of Social Sciences. Their average age was 21.4 (range 19-38), and 76.0% were female. They were all randomly assigned into the three treatments.

The experiment took place in classrooms. General instructions (the first page of the test papers) were read aloud to subjects before the experiment began. To motivate subjects to try their best in the experiment, they were given 50 Norwegian kroner and were told in the general instruction that their participation were very important for my master thesis project. Anonymity was assured as their names were decoupled from the test papers (Smith, 1982). No time limit was imposed but all subjects completed the pre-test, treatment and post-test within 50 minutes.

3 Results

The experiment results are summarized in Table 1. Improvement was obtained by 31.3% of the base treatment (T1) subjects, 71.4% of those receiving Graphical integration guidance (T2) and 43.8% of those with Running total and reflection (T3). Since the data do not approximate the normal distribution, Mann-Whitney U tests are used. The difference in the Breadth of Improvement (BI) between T1 and T2 is statistically significant (p value = .031); null hypothesis H1 is rejected. However, difference of BI between T1 and T3 is not significant (p = .472), null hypothesis H2 cannot be rejected.

For the base treatment, average score increased from 16.6% to 25.0%; from 19.0% to 69.0% for T2 and from 20.8% to 45.8% for T3. The line graph in Figure 9 helps visualize the improvements. The Depth of Improvement (DI) between T1 and T2 is significantly different (p = .014); null hypothesis H3 is rejected. But null hypothesis H4 cannot be rejected because DI between T1 and T3 does not differ significantly enough (p = .219).

Table 1 BI, DI, average pre-test and post-test scores across teaching approaches

Treatment			T1	T2	T3
N			16*	14*	16**
BI (%)			31.3	71.4	43.8
Score (%)	Pre-test	Average	16.6	19.0	20.8
		Std. dev.	29.8	25.2	36.3
	Post-test	Average	25.0	69.0	45.8
		Std. dev.	25.8	30.6	40.1
DI		Average	0.167	0.590	0.433
		Std. dev.	0.264	0.479	0.473

* One subject was removed from each treatment because they had done or learned about this kind of task before and earned perfect score on both pre-test and post-test.

** Five subjects were removed from the dataset because helpers did not carry out experiment properly with them.

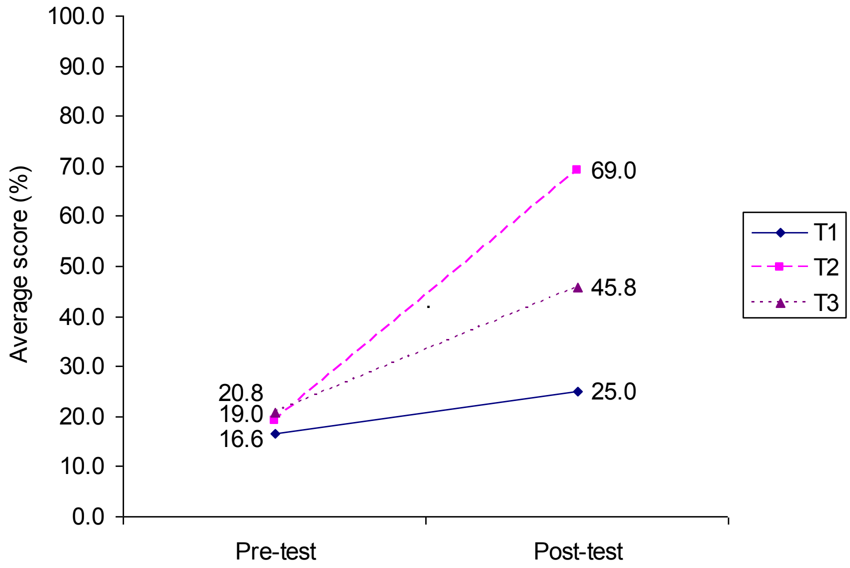


Figure 9 Average pre-test and post-test scores across teaching approaches

Further break down of the overall performance into each pre-test and post-test question is tabulated in Table 2. Most of the increase in T1's performance comes from Question 1 (43.8% correct on post-test Question 1 compared to 18.8% on pre-test Question 1), the easiest question in the tests. Performance is actually poorer on Question 2 (from 18.8% to 12.5%) and just slightly better on Question 3 (from 13.3% to 18.8%). In contrast, T2 and T3 improved for all questions, although the absolute gain is smaller on Question 3 (42.9% for T2; 12.5% for T3).

Table 2 Success rate of each pre-test and post-test question across teaching approaches

Treatment			T1	T2	T3
N			16	14	16
% Correct	Pre-test	Question 1	18.8	28.6	31.3
		Question 2	18.8	28.6	18.8
		Question 3	13.3	0.0	12.5
	Post-test	Question 1	43.8	85.7	56.3
		Question 2	12.5	78.6	56.3
		Question 3	18.8	42.9	25.0

Fisher's exact test shows that the success rate for each pre-test question is not significantly different (at 5%-level) across treatments, thus we can directly compare success rates for each post-test question to examine treatment effects. T2 has significantly stronger effect than T1 on post-test Question 1 and 2 (p = .026 and .001, respectively) while T3 outperforms T1 on post-test Question 2 (p = .023).

In summary, the Breadth and Depth of Improvement in Graphical integration guidance are significantly greater than the base treatment, manifested also by its post-test Question 1 and 2 success rates. Improvement in Running total and reflection treatment is only significant different than the base treatment on post-test Question 2.

4 Discussions

The results show that the conceived Running total and reflection approach, which evokes cognitive conflict using people's ability to calculate running total and guides reflection towards discovery of the principle of stock and flow dynamics, falls short of the expectation that it will definitely improve people's understanding of stock and flow dynamics. There is improvement, to be fair, in all of the performance indicators compared to no teaching, showing the idea is workable; but the improvement is not large enough to be significant, in which the underlying causes are to be found out.

One possible reason for the insignificant effect may be because its calculation and reflection requires quite a bit of effort from subjects. Some subjects in the Running total and reflection treatment did not answer reflective queries. If people refuse to reflect, they will have no learning. Strangely, some subjects showed signs of learning when answering reflective queries but no or little improvement in post-test. They seemed to not relate the answers with the subsequent graphical department store task, possibly because they had not been able to reorganize the discoveries into a coherent conceptual framework.

In retrospect, the teaching intervention could be too limited to realise its potential. The time spent is short, guidance provided for reflection is restricted, and there is not much feedback for subjects to get clues from. It might be a bit demanding to expect that people will swiftly discover the pattern of stock and flow relationships and form an abstract principle based on just three examples, especially if they are not familiar with this kind of learning strategy. Therefore the result is not totally surprising.

On the other hand, under the same constraints, the success of Graphical integration guidance, a series of step-by-step guided work-out exercises, is particularly remarkable. More people are able to do graphical department store task afterwards and their abilities to do graphical integration improve. However, constrained by the experimental design, I did not test its retention and transfer of learning to farther context. In other words, we do not know how deep the understanding of stock and flow dynamics was. Therefore it is only safe to conclude that Graphical integration guidance is effective in improving people's graphical integration skill. Nevertheless, this shows that Graphical integration guidance must contain some effective instructional design principles that work for this purpose, which can be explored in future.

Inspecting performance of individual questions, a majority of subjects still failed on post-test Question 3. Twenty five percent (25%) of Running total and reflection treatment's subjects did it correctly, even with the Graphical integration guidance, only 42.9% were able to figure it out. Fifty eight percent (58.3%) of Running total and reflection treatment subjects who erred, drew paths that matched the shape of inflow or net flow positively or inverted (Table 3). This denotes the persistence of difficulty for people to deduce dynamics for non-constant net flow, thus they resorted to intuitively appealing correlation heuristic to get out of the predicament.

Since Question 3 of the pre-test, together with Question 1, 2 and 3 of the post-test, are identical to Condition 5, 1, 2 and 3 of the Cronin et al. (2009) Experiment 5 and Sterman (2009) post-test, it is possible to compare[1] the results obtained with their results, see Table 3. It appears that Sterman's half-semester introductory SD course had larger impact, yet considering the short intervention time and different characteristics of subject population (initial performance of this study is poorer), the progress made by Graphical integration guidance is still impressive. This is actually a quite common scenario, for example, to attempt to explain behaviour implication of principle of accumulation to someone who has little time and incentive to learn further. It is not an either-or comparison, as these teaching techniques can be integrated into introductory SD course to enhance its performance; while ample time of introductory SD course allow more activities to be arranged to reinforce retention and transfer of learning.

Table 3 Results of this study compared[1] to Cronin et al. (2009) and Sterman (2009)

<table>
<tr><th colspan="3">This study</th><th>Pre-test Question 3</th><th>Post-test Question 1</th><th>Post-test Question 2</th><th>Post-test Question 3</th></tr>
<tr><th colspan="3">Cronin et al. (2009) and Sterman (2009)</th><th>Condition 5</th><th>Condition 1</th><th>Condition 2</th><th>Condition 3</th></tr>
<tr><td rowspan="4">% Incorrect</td><td colspan="2">Cronin et al.</td><td>44.4</td><td>-</td><td>-</td><td>-</td></tr>
<tr><td colspan="2">Sterman</td><td>-</td><td>4.8</td><td>25.0</td><td>22.7</td></tr>
<tr><td rowspan="2">This study</td><td>T2</td><td>100.0</td><td>14.3</td><td>21.4</td><td>57.1</td></tr>
<tr><td>T3</td><td>87.5</td><td>43.8</td><td>43.8</td><td>75.0</td></tr>
<tr><td rowspan="4">% Incorrect exhibiting correlation</td><td colspan="2">Cronin et al.</td><td>80.0</td><td>-</td><td>-</td><td>-</td></tr>
<tr><td colspan="2">Sterman</td><td>-</td><td>0.0</td><td>40.0</td><td>40.0</td></tr>
<tr><td rowspan="2">This study</td><td>T2</td><td>71.4</td><td>0.0</td><td>100.0</td><td>37.5</td></tr>
<tr><td>T3</td><td>71.4</td><td>42.9</td><td>71.4</td><td>58.3</td></tr>
<tr><td rowspan="4">N</td><td colspan="2">Cronin et al.</td><td>35</td><td>-</td><td>-</td><td>-</td></tr>
<tr><td colspan="2">Sterman</td><td>-</td><td>21</td><td>20</td><td>22</td></tr>
<tr><td rowspan="2">This study</td><td>T2</td><td colspan="4">14</td></tr>
<tr><td>T3</td><td colspan="4">16</td></tr>
</table>

So what are the implications of these results for SD instruction?

First, look at the possible keys to success for Graphical integration guidance. Seeing the development of the answers of the group on the treatment exercises as well as pilot experiment experiences, I would attribute the improvement largely to the continuous feedbacks and work-out practices, implanting the steps of doing graphical integration

[1] The comparison has limitation. Adaptation of the task, demography of participants or other unmeasured sources of variation could confound the comparison.

until they are internalized. This is an effective way to teach graphical integration, and is possibly also effective for other well-defined skills, which SD educators could make use of (e.g. stocks and flows identification and mapping exercises).

Second, notice that the strategy employed by Running total and reflection treatment – first give one's *expected* answer, then get the *actual* answer (by hand calculation or computer simulation), and if they disagree, reflect or find out the root cause of disagreement – is basically the manner we learn insights when modelling. Therefore, though the Running total and reflection seems to be no advantage over the Graphical integration guidance, if our instructional goal is not only acquiring the skill of graphical integration, but also acquire the way of SD reasoning, this teaching approach may be preferable. Cognizant of the time and efforts demanded by it, we learn that motivation is a crucial factor. It should be helpful to strengthen the extrinsic motivation for learning by course credit, or intrinsic motivation by articulating the utility of understanding of stock and flow dynamics or SD way of reasoning. Besides, necessary guidance, or scaffolding, is very supportive especially to the novice learner, when they are still not used to the SD learning strategy. Start with the simple task, with prompting along the way. As the learner's capability improved, these assistances can, and in fact should, fade away. Literature on conceptual change research, e.g. Limon & Mason (2002), should be useful to facilitate deep learning.

Back to the learning of stock and flow, it is now clear that overcoming misperception of stock and flow dynamics is harder than we presumed. Even genuine cognitive conflict (generated by people themselves) does not guarantee improvement. This may shed light on many communication headaches on complex dynamic problems and policy insights, where people do not respond to cognitive conflict revealed by System Dynamists. Contemporary education does not help either – 82% of all subjects say "no" to the question "Do you think your educational background has prepared you for this kind of task?" This signifies more work needs to be done to disseminate principles of dynamics. The importance of basic education to aid the communication of complex dynamic problems and policy insights cannot be overstated.

5 Conclusion

Being aware of the general poor (or no) understanding of stock and flow, this study explored the potential of exploiting people's existing ability to calculate running total to improve their understanding of stock and flow behaviours, using a pre-test-treatment-post-test experiment. Results suggest that it is possible for people to learn stock and flow dynamics using their existing running total knowledge; on average subjects made modest improvement. The idea of the new Running total and reflection approach does make sense, but the ideal is harder to achieve than previously expected. On the other hand, conventional teaching approach in the form of step-by-step guided work-out exercise, is effective to teach graphical integration skill. But as mentioned previously, retention and transfer of learning from the Graphical integration guidance remains to be tested.

The results could be explained in terms of motivation, guidance, feedbacks and practices, as discussed. In light of the above, I propose enhancements applicable for the new Running total and reflection approach. The present teaching interventions should be seen as initial efforts in improving people's understanding of stock and flow. Further works include graphical integration exercises in different framing (e.g. continuous case), non-examples (e.g. speed of the bus and speed of the chair in the bus), other stock-flow tasks (e.g. CO2 zero emissions task), identifying and mapping stocks and flows exercises. To cultivate the habit of thinking in stock and flow, people need to regularly apply them throughout their lives.

Recognizing the tedious and error-prone procedure to carry out these teaching interventions, it is beneficial to computerize them. In the meantime, adaptation and trial of the tasks are very welcomed. Data collected under controlled conditions permits us to evaluate the extent to which the teaching technique is appropriate in bringing about a particular type of learning to a particular type of student. After all, rise to the challenge of improving people ability to deal with increasingly dynamic complex world, clever blend of teaching approaches allows us to achieve the ultimate desired learning outcomes most effectively.

Acknowledgement

The author would like to thank Professor Erling Moxnes for his guidance and financial support on conducting the experiment and reviewing this manuscript. Also thanks to helpers for carrying out the experiment treatments and all experiment participants including pilot experiment. Thanks to four anonymous reviewers for their useful feedbacks to improve this paper. Of course any faults found in this paper rest wholly with the author. Thanks isee systems, this conference committee and Faculty of Social Sciences, University of Bergen for financially supporting the author to attend the conference. Lastly, the author is grateful for being part of the System Dynamic Group at University of Bergen, together with Professor Pål I. Davidsen, Professor Erling Moxnes and Associate Professor I. David Wheat.

References

Cronin, M. A., Gonzalez, C., & Sterman, J. D. (2009). Why don't well-educated adults understand accumulation? A challenge to researchers, educators, and citizens. *Organizational Behavior and Human Decision Processes, 108*(1), 116-130.

Forrester, J. W. (2009). Some Basic Concepts in System Dynamics. Retrieved from http://www.clexchange.org/ftp/documents/system-dynamics/SD2009-02SomeBasicConcepts.pdf

Kainz, D., & Ossimitz, G. (2002). *Can Students Learn Stock-Flow-Thinking? An Empirical Investigation.* Paper presented at the 2002 International System Dynamics Conference, Palermo, Italy.

Limon, M., & Mason, L. (2002). *Reconsidering Conceptual Change: Issues in Theory and Practice*. Hingham, MA, USA: Kluwer Academic Publishers.

Marx, J. D., & Cummings, K. (2007). Normalized change. *American Journal of Physics, 75*(1), 87-91.

Moxnes, E. (1998). Overexploitation of renewable resources: The role of misperceptions. *Journal of Economic Behavior & Organization, 37*(1), 107-127.

Moxnes, E., & Saysel, A. K. (2009). Misperceptions of global climate change: information policies. *Climatic Change, 93*(1), 15-37.

Olaya, C., Diaz, G., Ramos, A. M. S., & Pabon, F. D. (2008). *The Power of the Stock: Accumulations in the Colombian Accusatory System Reform.* Paper presented at the 2008 International System Dynamics Conference, Athens, Greece.

Ossimitz, G. (2002). *Stock-Flow-Thinking and Reading stock-flow-related Graphs: An Empirical Investigation in Dynamic Thinking Abilities.* Paper presented at the 2002 International System Dynamics Conference, Palermo, Italy.

Pala, Ö., & Vennix, J. A. M. (2005). Effect of system dynamics education on systems thinking inventory task performance. *System Dynamics Review, 21*(2), 147-172.

Smith, V. L. (1982). Microeconomic Systems as an Experimental Science. *The American Economic Review, 72*(5), 923-955.

Sterman, J. D. (2000). *Business dynamics: systems thinking and modeling for a complex world*. Boston: Irwin McGraw-Hill.

Sterman, J. D. (2009). *Does formal system dynamics training improve people's understanding of accumulation?* Paper presented at the 2009 International System Dynamics Conference, Albuquerque, New Mexico, USA.

Sterman, J. D., & Sweeney, L. B. (2007). Understanding public complacency about climate change: adults' mental models of climate change violate conservation of matter. *Climatic Change, 80*(3), 213-238.

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