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

Tony Leong Keat Phuah

Thesis Submitted in Partial Fulfillment
Of the Requirement for the Degree of
Master of Philosophy in System Dynamics

System Dynamics Group
Department of Geography
University of Bergen

Supervised by Professor Erling Moxnes

May, 2010

Abstract

Stock and flow is 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 discovered that people performed poorly in simple stock-flow tasks. However, many do have notion of accumulation in terms of calculating running total. 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 asking them to verify their expected behaviour using running total calculation and reflect if their expected behaviour was wrong. Comparisons with conventional approach to teach stock and flow dynamics and without teaching were also done, to my knowledge, the first time in controlled experiment. Results show that improvement is not significant; the hypothesis lacks support. On the other hand, conventional approach obtains 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*

Acknowledgement

This work benefited immensely from the advice, criticism and encouragement of many teachers and friends. I would like to convey my deepest gratitude to my supervisor, Prof. Erling Moxnes, for his invaluable guidance and financial support on conducting the experiment, as well as comments throughout this thesis writing. Thanks for assisting me to transform an initial vague idea into a reality.

Rarely are works of any size the result of a singular effort. Thanks Shi Lili, Shao Wei, Hiba Abdelrahman M. Ahmed, Dennis Osei Kontor, Iman Sebtiu Mohammed, Sebastian Alcaraz Martinez, Inger Bjørgo Hustvedt, Richard Allistan McDowall, Warakan Supinajaroen, Maria Federica Dinola, Sebastian, Carmen Lee, Lee Yik Nam and Wu Yun for their helps to carry out the experiment treatments. Thanks also Miheala Tabacaru for helping me to move the experiment equipments. Thanks all experiment participants including pilot experiment. Thank you for making the experiment a success. Thanks also to those who had directly or indirectly contributed in one way or another towards the accomplishment of this study.

I am particularly grateful to be the student of Assoc. Prof. I. David Wheat and Prof. Pål I. Davidsen. Together with Prof. Erling Moxnes, you have helped me to build up the basic foundation of System Dynamics, a useful tool that allowed me to understand our Sustainability challenge better. I will look for ways to contribute what I have learned here for the benefit of the world.

Finally, endless love and support of my family is always my motivation to overcome all the challenges faced. I love you, Mom, Dad, Gilbert and Bernard.

Table of Contents

List of Figures

List of Tables

1 Introduction

Groundbreaking experiment by Sweeney and Sterman (2000) had astonished System Dynamics 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 already 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, as if time does not play any role in changes of variable; but this static mentality overlooks the fact that flow-affects-stock causal relationship is accumulative by nature, the function of flow is more to controlling the rate of changes, as illustrated by Figure 1-1. The laws of accumulation dictate the peculiar behaviours of flows and their resulted stock.

Unfortunately, this gap of understanding creates a challenge for System Dynamics educators and consultants to explain the behaviours of stock and flow. Without proper model of stock and flow, people will have wrong expectation of behaviours, which can have serious policy implications. Notable examples include misperceptions of climate change mechanism (Sterman & Sweeney, 2007), renewable resources management (Moxnes, 2004), 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 so that they can identify stocks and flows in everyday life and reason about their behaviours?

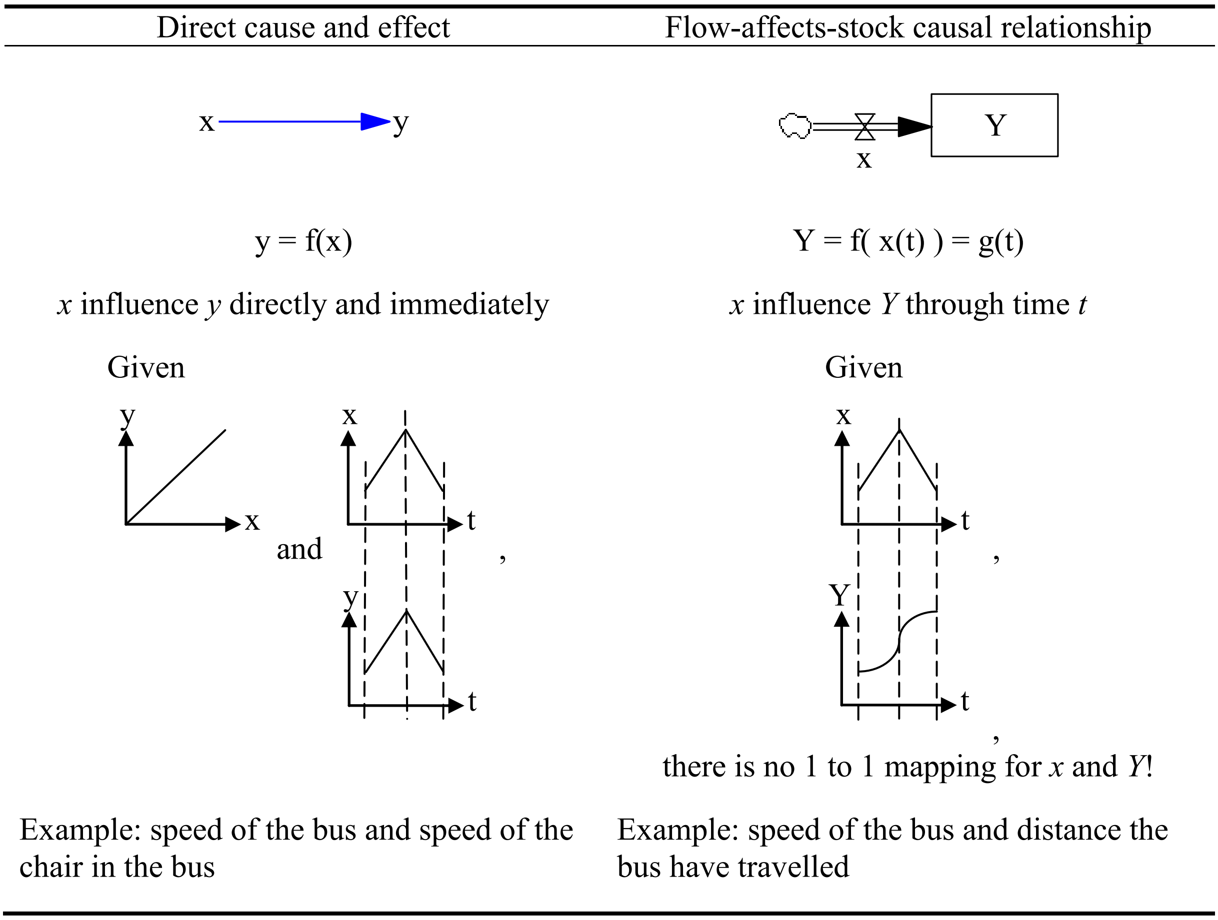


Figure 1-1 Difference between direct and flow-affects-stock causal relationship

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 System Dynamics 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 System Dynamics course, showing wide variations of the effects of System Dynamics education on improving stock and flow understanding.

These inconclusive effects indicate the need to move towards systematic designed instruction in System Dynamics education, based on well tested instructional 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 serve for practices and feedbacks, two elements that 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 universal process and many people do have basic idea of 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 Chapter 2. Experiment results in Chapter 3 show that the first approach makes significant gain in 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 System Dynamics education in Chapter 4. Finally, Chapter 5 concludes this study and suggests further research that can help to advance the profession of System Dynamics 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 2-1 shows the organization of the experiment.

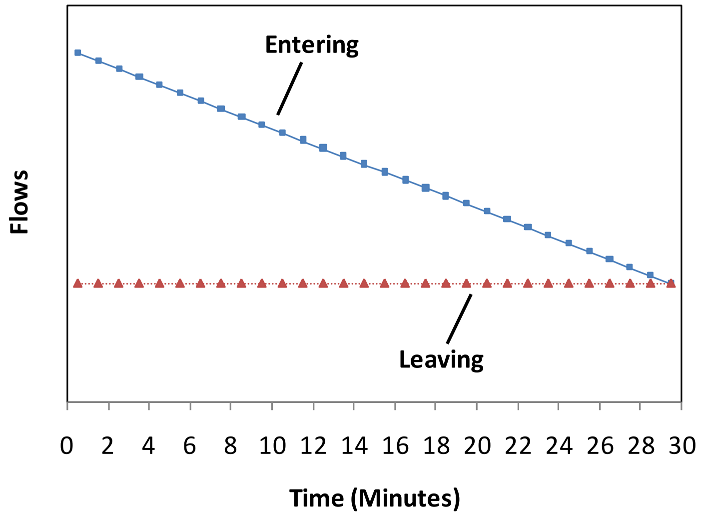
Pre-test	T1 Base No teaching.	Post-test
	T2 Graphical integration guidance Step-by-step queries guiding subject to do graphical integration.	
	T3 Running total and reflection Compare expected answer with calculated running total and reflect on the reason of discrepancy (if any) to discover laws of accumulation.	

Figure 2-1 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 the questions raised. Figure 2-2 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.

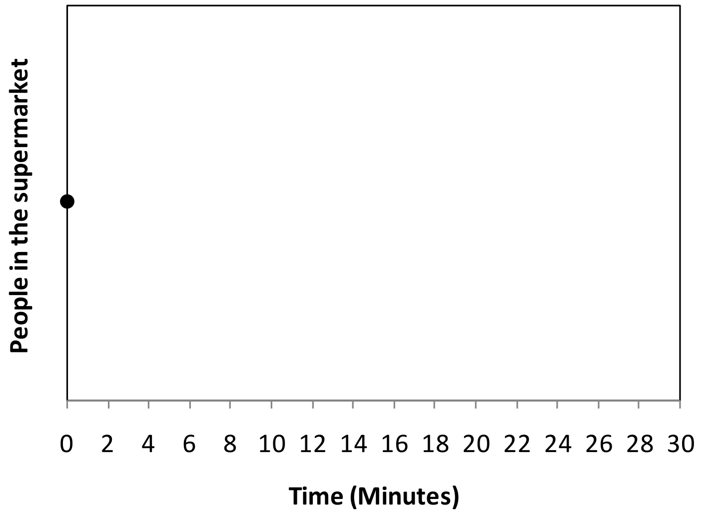


Figure 2-2 Adapted graphical department store task used in this experiment

From the stock and flow perspective, the underlying model of the task is extremely simple, as portrayed by Figure 2-3. 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.

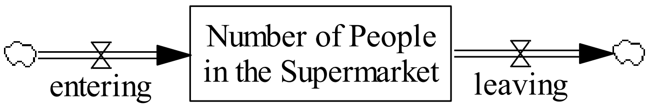


Figure 2-3 Stock and flow structure of department store task

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 2-4 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.

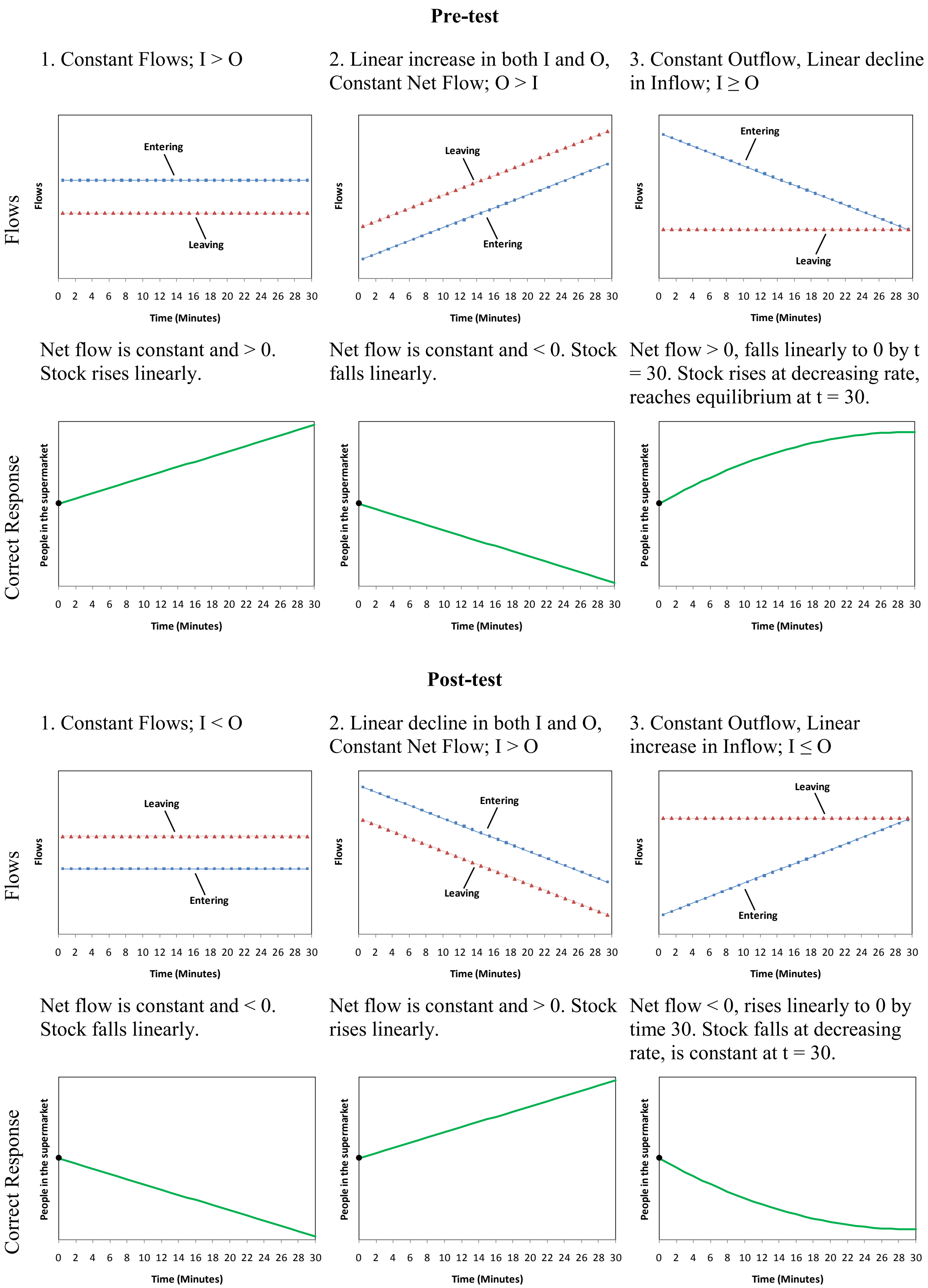


Figure 2-4 Flow graphs for the pre- and post-test questions and their correct responses

2.3 Treatments

As exhibited by Figure 2-1, the experiment has three treatments. All treatments were administrated using paper and pen.

2.3.1 T1 The base treatment

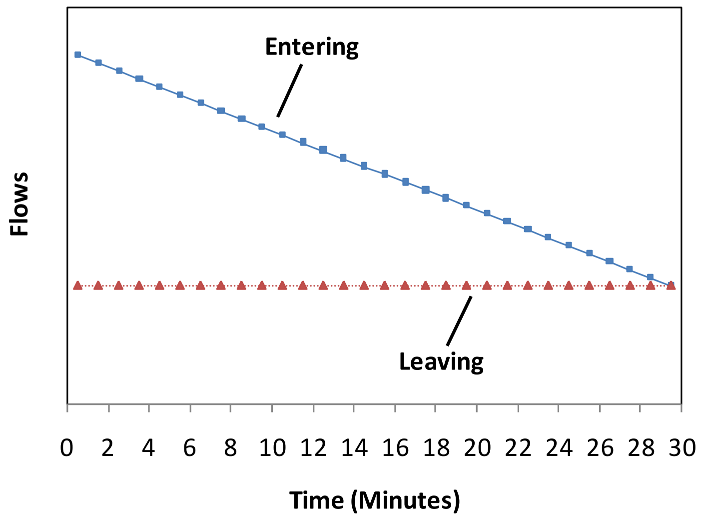
Since the subjects do not receive any teaching, they will only go through pre-test and post-test, total 6 questions. Pre-test must be finished before subject can start the post-test. Post-test will be given after subject completed pre-test and indicate so by raising his/her hand. The difference in performance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

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).

Therefore graphical integration guidance uses these questions to provide step-by-step guided work-out exercise, as presented in Figure 2-5. There are three such exercises in the treatment, each based on one pre-test question. After subject completed each exercise, feedback is given to inform subject whether his or her answer is correct or not. Drawing is judged qualitatively correct if both direction and shape are correct. If it is wrong, the correct trajectory will be drawn. Checking on the subject's corresponding pre-test answer is also done on the spot. Subject can then learn from mistake (if any) before move on to the next question. This outcome feedback is given after every exercise to maximize potential for learning.

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

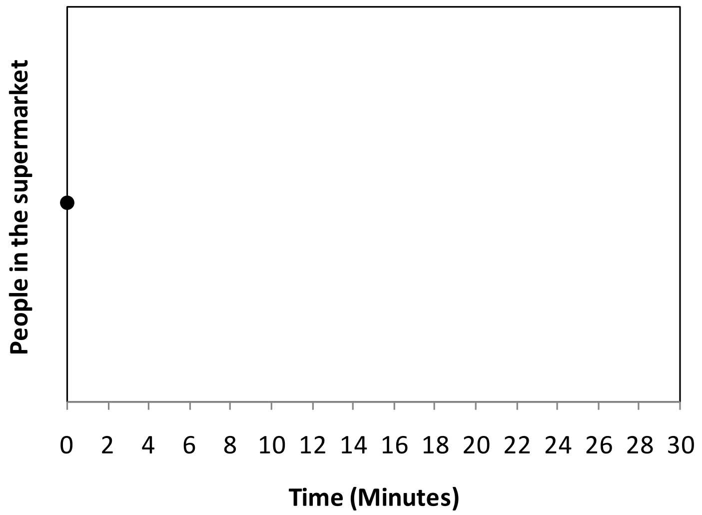


a. Which flow is larger? ☐ Entering larger than leaving ☐ Leaving larger than entering ☐ Both same

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.

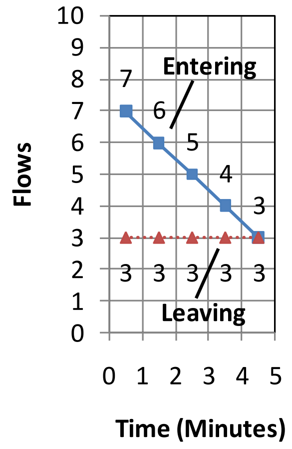
Figure 2-5 One of the graphical integration guidance exercise

2.3.3 T3 Running total and reflection

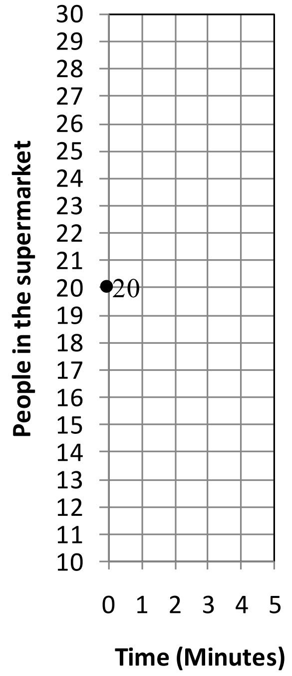
Prompted by Experiment 4 of Cronin et al. (2009), the idea of calculate running total of a stock should be familiar to most people, even though they may not know what is stock and flow and infer their behaviours intuitively. The obvious question is then, *how to make use of people's ability to calculate 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) strategy, this treatment utilizes people's ability to calculate running total to evoke cognitive conflict, then assists people to resolve this conflict by carefully crafted reflective instructions. 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 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 this moment 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 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 2-6 demonstrates the so called calculation instruction.

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 2-6 Calculation instruction in Running total and reflection treatment

Subject's plotted graph will receive feedback before he or she proceeds to reflective instruction. The graph must pass or has all points correctly, 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 make it because they want to finish it quickly. This feedback urges them to answer more carefully, and increase their chances of learning.

After the calculation checking, subject's previous answer during pre-test is returned and checked by referring to the correctly calculated answer. Subsequent instruction (Figure 2-7) calls subject's attention to the equivalence of pre-test and treatment questions. If subject's expected answer in pre-test was different from the correctly calculated answer, the declaration of error should be thought-provoking. If not, the "Why?" queries followed will indirectly request subject to think about it in deep, in order to give explanation. The queries are split into two parts to help subject notice characteristics of the stock's behaviour easier by concentrating on one characteristic at a time. If subject's pre-test answer was correct, he or she can skip the queries to continue the next question. This is to prevent subject who has already mastered principle of accumulation feeling bored about unnecessary reflections, which may 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 2-7 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 post-test performance in Graphical integration guidance and Running total and reflection treatments will improve, compared to the base treatment.

The performance is quantified as sum of subject's correct answer 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 the effect of treatment was on 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 was the improvement reached by the subjects, constrained by the maximum potential improvement. 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 have run the experiment with two groups of undergraduate students from University of Bergen, Faculty of Social Sciences. First in February 2010, 31 students (SV1) took the test. Later in March 2010, another 22 students (SV2) were given the test to affirm the initial finding, as well as to increase the sample size. They are demographically similar, with average age 21.4 (range 19-38) years. The only statistical significant different characteristics is gender, where 87.1% of SV1 are female, compared to 57.9% of SV2 (Fisher's exact test yields $p = 0.038$). Anyway, both groups received three treatments and the assignment of treatments is random. This should balance any possible effect of subject characteristics. Table 2-1 shows the distribution over subject groups and treatments.

Table 2-1 Distribution of subjects over groups and treatments

Treatment	T1	T2	T3	Total
SV1	12	10	9	31
SV2	5	5	12	22
Total	17	15	21	53

The experiments took place in classroom. General instruction (the first page of test papers) was 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. Privacy 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

Experiment results for the two subject groups are analyzed separately at first. Their results are similar, hence I pool SV1 and SV2 for further analysis. The pooled result is summarized in Table 3-1. 31.3% subjects of the base treatment (T1), 71.4% of Graphical integration guidance (T2) and 43.8% of Running total and reflection (T3) experienced improvement. Since the data does not approximate normal distribution, Mann-Whitney U test is 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. The bar charts in Figure 3-1 help to visualize their BI.

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, as depicted by Figure 3-2. 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 3-1 BI, DI and 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.

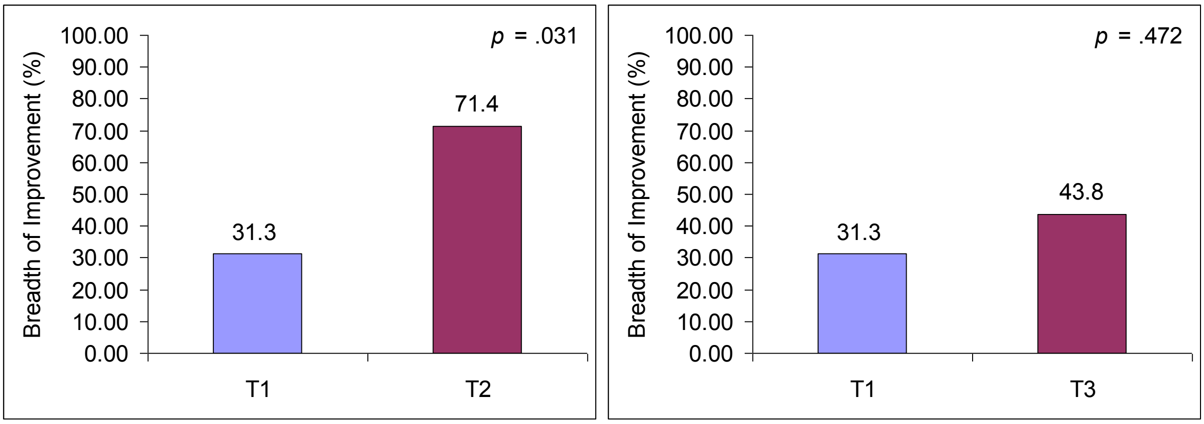


Figure 3-1 Breadth of Improvement of T2 and T3 compared to T1

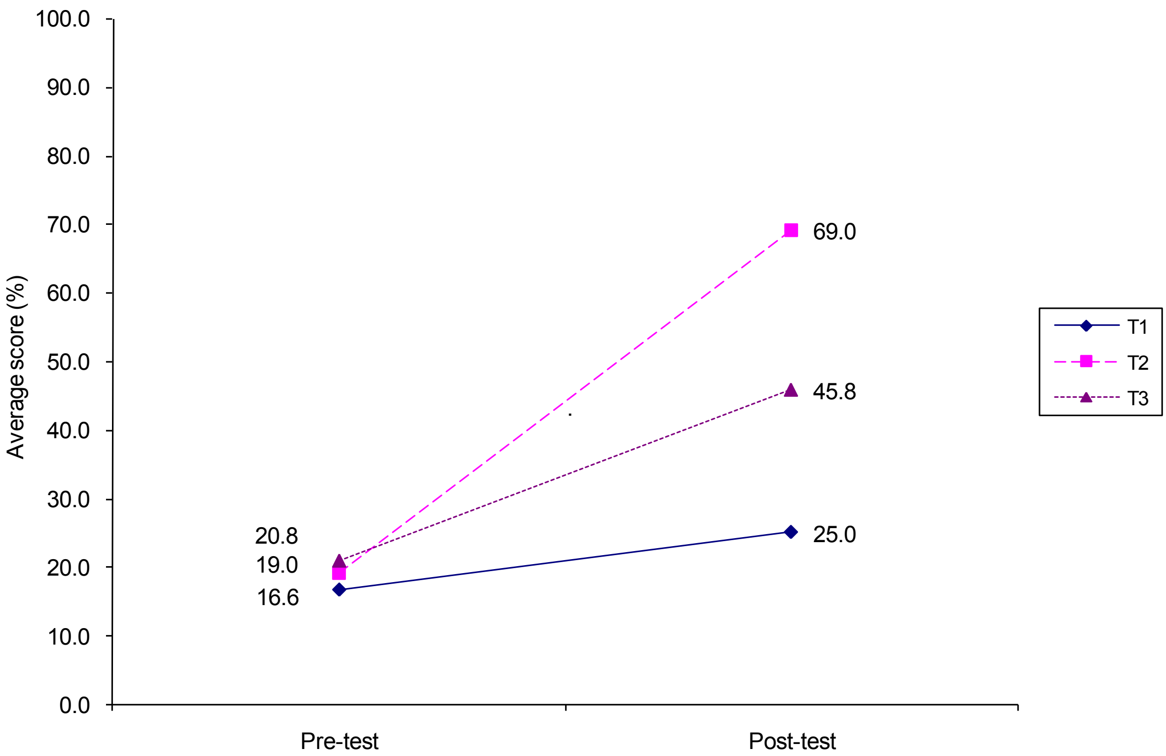


Figure 3-2 Average pre-test and post-test scores across teaching approaches

Further break down the overall performance into each pre-test and post-test question is tabulated in Table 3-2. 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 post-test. 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 the questions, although the absolute gain is smaller on Question 3 (42.9% for T2; 12.5% for T3).

Table 3-2 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 different significantly (at 5%-level) across treatment, thus we can directly compare success rate for each post-test question to examine effect of treatments. T2 has significant positive effect than T1 on post-test Question 1 and 2 (p = .026 and .001, respectively) while T3 outperforms T1 on post-test Question 2 significantly (p = .023).

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

In comparison, Graphical integration guidance, a series of step-by-step guided work-out exercises, clearly helps people to learn graphical integration. More people are able to do graphical department store task afterwards and their level of graphical integration is rising. However, constrained by the experimental design, post-test did not test its retention and transfer of learning to farther context. In other word, we do not know how deep the understanding of stock and flow dynamics had been improved. 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 will be explored later. But first let me compare the results with previous works.

4.1 Comparison to earlier literature

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, as Table 4-1 displayed. It appears that Sterman's half-semester introductory System Dynamics 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 quite common scenario, for example, to attempt to explain behaviour implication of the principle of accumulation to average Joe 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 System Dynamics course to enhance its performance; while ample time available for introductory System Dynamics course allow more activities to be arranged to reinforce retention and transfer of learning.

Table 4-1 Results of this study compared[1] to Cronin et al. (2009) and Sterman (2009)

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

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

Regarding the gender effect discovered by Sterman (2009), I did not find the males performed better than female, which probably owing to small sample size and dominance of female in my subject population.

Inspecting performance of individual questions in my experiment, majority of subjects still failed on post-test Question 3. 75% of Running total and reflection treatment subjects did it incorrectly, even with the Graphical integration guidance, 57.1% still unable to figure it out. 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 4-1). 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.

4.2 Why Running total and reflection resulted smaller improvement than expected?

The same rationale may also give a hint about the disappointing result of Running total and reflection treatment. Only 37.0% of subjects (N = 46) consistently use correlation heuristic in all their three pre-test answers. Some also commented that the graphical department store task is difficult, indicating that they actually have no idea how to do it. They may sometime use correlation heuristic as wild guess, but do not place much confidence in it. If this is the case, comparison between wrong expected answer and correctly calculated answer may not produce cognitive conflict as strong as expected. As a result, people may not eager to reflect to find out reason of the discrepancy. This is especially detrimental since the graphical department store task is already lack of personal relevance to the subjects ("Why should I need to know how number of people in a supermarket change?").

This motivational factor is important because this approach requires calculation and reflection, which demands more efforts from subjects to think. It is found that some subjects did not answer reflective queries intentionally. 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 signs of improvement in post-test. They seemed to not relate the answers with the subsequent graphical department store task. This reveals that in order for Running total and reflection approach to work, several conditions have to be met: (1) People can understand the task, read the graph, calculate running total and plot graph; (2) People know how to compare expected answer with actual answer and aware of the error (if any); (3) People can discover the relationship of stock and flow behaviours from running total calculation; (4) People able to reorganize the discoveries into a coherent conceptual framework and apply it. Failure in any of these chains of actions will not complete the intended learning. Without purposeful guidance to assist people in overcoming the learning curve, people may be overwhelmed by the difficulties of this learning approach, ended up with fragmented learning.

In retrospect, the teaching intervention is likely too limited to realise its potential. The time spent is short, guidance provided for reflection is restricted, and not much feedback for subjects to get clues from. It might be a bit demanding to suppose people will swiftly discover the pattern of stock and flow relationships and form abstract principle based on just three examples by themselves, especially if they are not familiar with this kind of learning strategy. Therefore it is not totally surprised that subjects' post-test performance is not markedly better than their pre-test.

4.3 Why Graphical integration guidance works so well?

On the other hand, under the same constraints, the success of Graphical integration guidance is particularly remarkable. This has offered an opportunity to study the effective instructional design principles embedded. Seeing the development of answers of the group on treatment exercises as well as pilot experiment experiences, I would attribute the improvement largely to the work-out practices and continuous feedbacks, which successfully implanted the steps of doing graphical integration. Feedback is not only making learning possible (learning is a feedback process, see Sterman (2000)), but also can boost the subject's morale to try hard on the next question if his/her answer on previous question was reported correct. In addition, presenting the steps in a form of short questions should ease memorization and recall.

4.4 What are the implications of these results?

First, these possible keys to success for Graphical integration guidance are effective to teach graphical integration, and likely for other well-defined skills, which System Dynamics educators could make use of (e.g. stocks and flows identification and mapping exercises).

Second, in view of the subtle guidance needed, Running total and reflection seems to have no advantage over the Graphical integration guidance. However, notice that the strategy employed by the Running total and reflection – first give our *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 from modelling process. Therefore, if our instructional goal is not only acquiring the skill of graphical integration, but also acquire the way of System Dynamics reasoning, this teaching approach may be preferable. This shows that instructional goal affects teaching option.

Cognizant of the time and efforts required, we learn that motivation is a crucial factor to both learners and teachers. 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. Besides, necessary guidance, or scaffolding, is very supportive especially to the novice learner, when they are still not used to the System Dynamics 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, be faded away. Literature on conceptual change research, e.g. Limón & 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 did not response to cognitive conflict revealed by the System Dynamists as expected. Contemporary education does not help either – 82.0% 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 works need to be done to disseminate principle of dynamics. The importance of basic education on dynamics to aid the communication of complex dynamic problems and policy insights cannot be overstated.

5 Conclusion

In 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 people can learn dynamics of stock and flow using their existing running total knowledge; the improvement is modest. 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, its retention and transfer of learning remain to be tested.

The results could be explained in terms of motivation, guidance, feedbacks and practices, as discussed. In light of the above, enhancements applicable for the new Running total and reflection approach are proposed. These teaching interventions should be seen as initial effort 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. CO_2 zero emissions task), identifying and mapping stocks and flows. 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 teachings, it is beneficial to computerize them. In the mean time, 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 more and more dynamic complex world, clever blend of teaching approaches allows us to achieve the ultimate desired learning outcomes most effectively.

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.

Limón, M., & Mason, L. (2002). *Reconsidering Conceptual Change: Issues in Theory and Practice*. Dordrecht ; Boston: 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. [doi: DOI: 10.1016/S0167-2681(98)00079-1]. *Journal of Economic Behavior & Organization, 37*(1), 107-127.

Moxnes, E. (2004). Misperceptions of basic dynamics: the case of renewable resource management. *System Dynamics Review, 20*(2), 139-162.

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.

Appendix A The tests

The experiment is divided into pre-test, treatment and post-test. Treatment T1 will only contain pre-test and post-test, so they are named Part 1 and Part 2. Treatment T2 (Graphical integration guidance) and T3 (Running total and reflection) have three parts, so Part 1 is pre-test, Part 2 is their corresponding treatment and Part 3 is post-test. Carry out the experiment based on the helper guides in the following appendix.

The subject no. is usually tagged in alphanumeric (e.g. A1, B2) to identify them quickly. These are equivalent:
A = T1 The base treatment
B = T2 Graphical integration guidance
C = T3 Running total and reflection

Part 1 [pre-test]

Subject No: __________

Dear participant,

Thank you for participating in this test. Your participation is very important for my master thesis project.

Instructions:

1. Please answer ALL questions, unless specifically instructed to skip a question. If you are uncertain about the meaning of a question, please ask the administrator for clarification.

2. The test consists of TWO (or THREE) parts. **Please raise your hand when you are asked to do so**, to continue the test.

(3. When the administrator marks 'R' on your answer, it means your answer is right. If it is wrong, the administrator will show you the right answer.)

Your answers will not be coupled with your names, you remain anonymous. However, we do check that you have done your best before you receive 50 kroner for participating. For accounting purposes we need your signature on a separate sheet of paper.

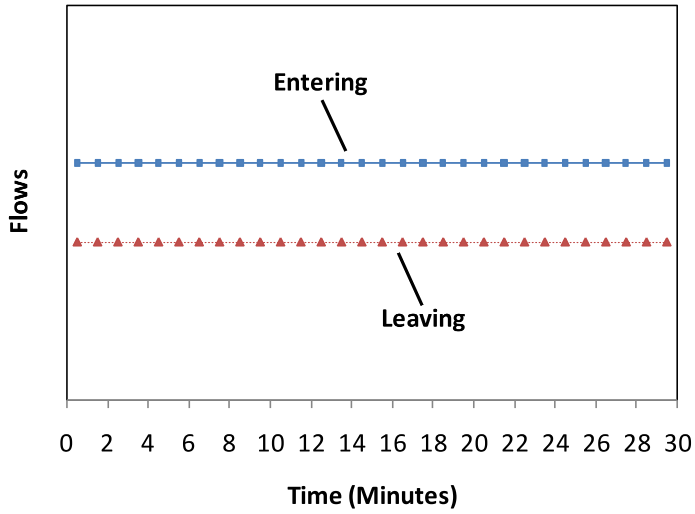
Thank you very much for participating and for doing your best!

Part 1 [pre-test]

Subject No: __________

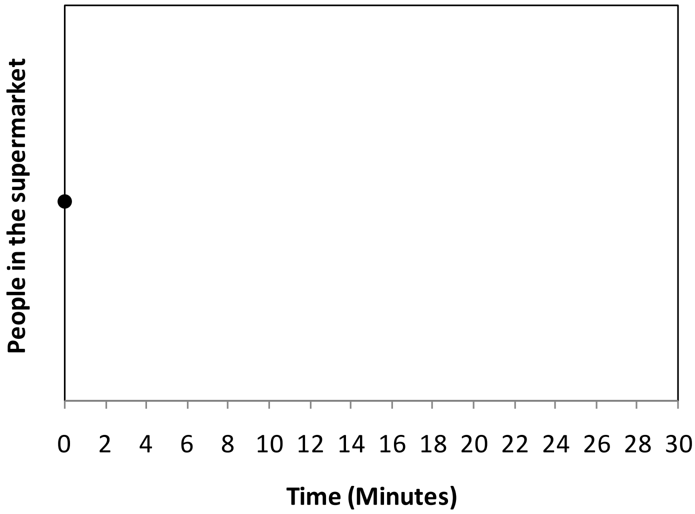
1.

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.

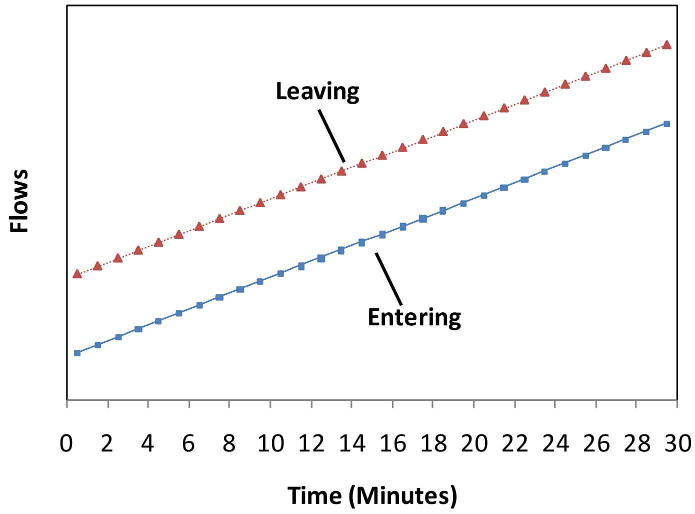


Part 1 [pre-test]

Subject No: __________

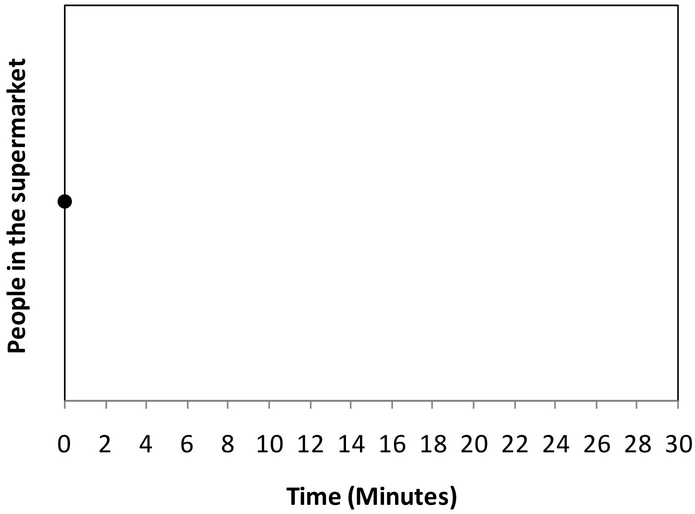
2.

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.

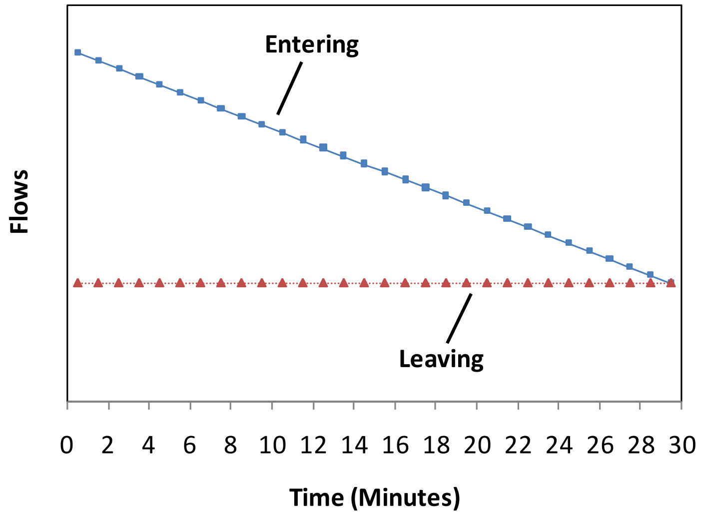


Part 1 [pre-test]

Subject No: __________

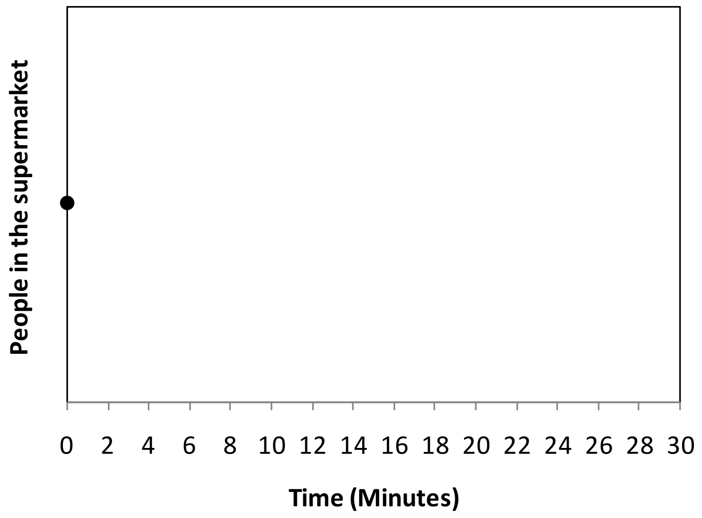
3.

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.



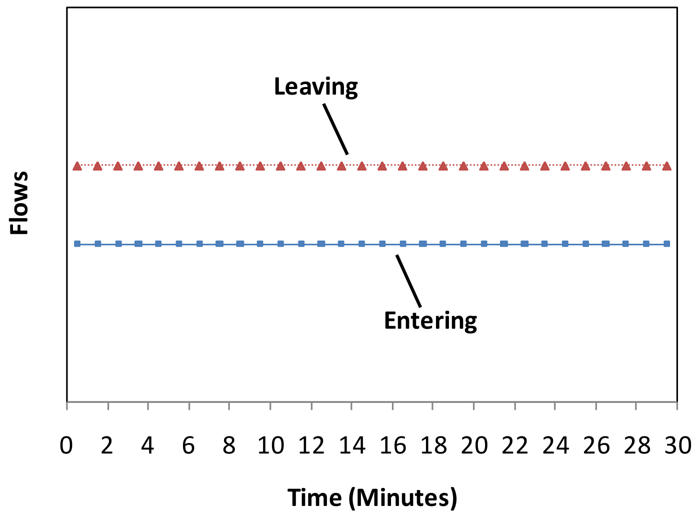
Please raise your hand after having completed these questions.

Part 2 or 3 [post-test]

Subject No: __________

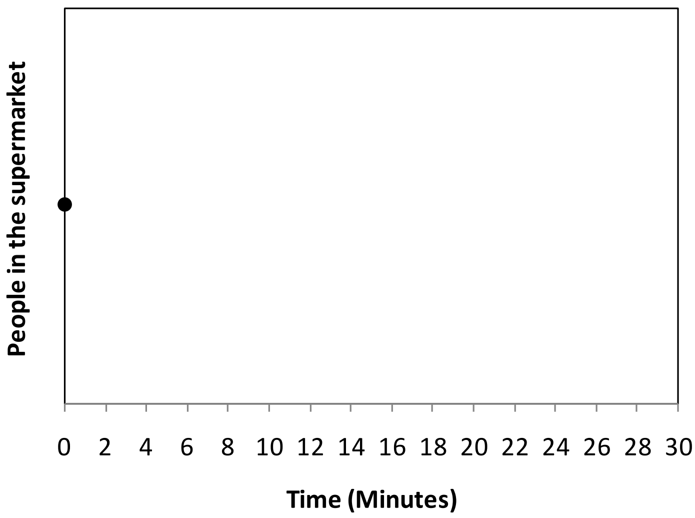
1.

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.

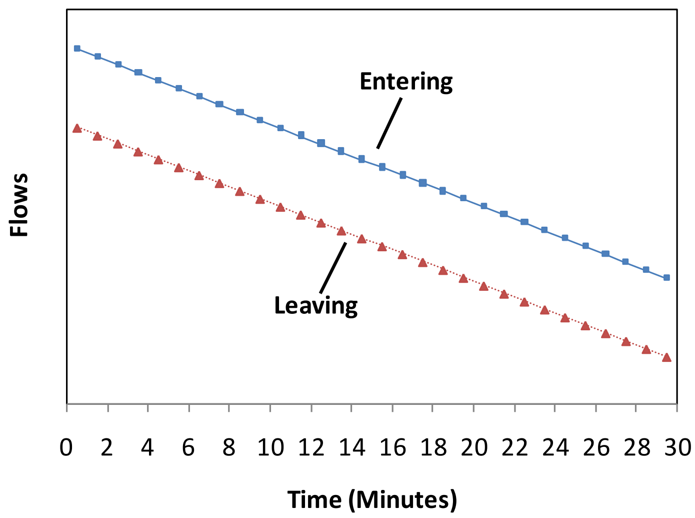


Part 2 or 3 [post-test]

Subject No: __________

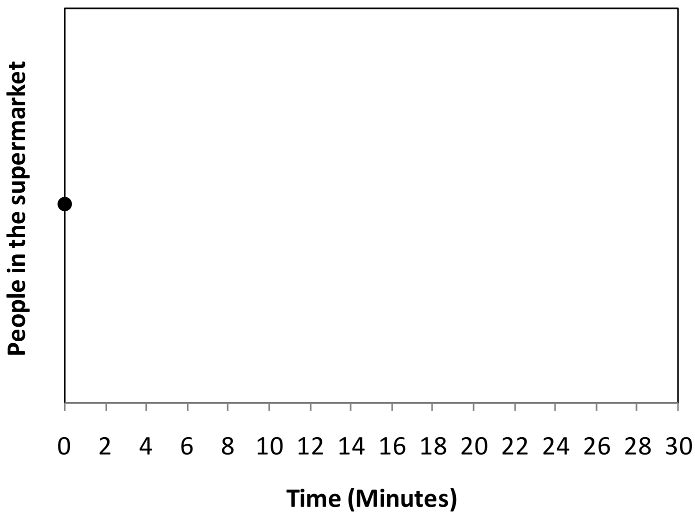
2.

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.

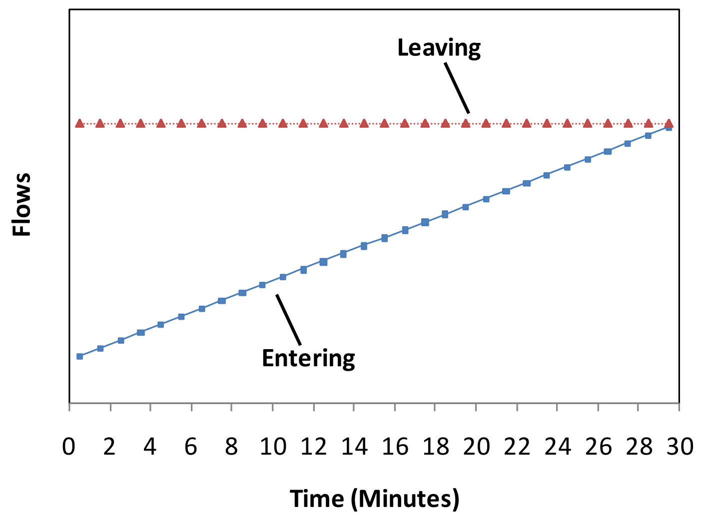


Part 2 or 3 [post-test]

Subject No: __________

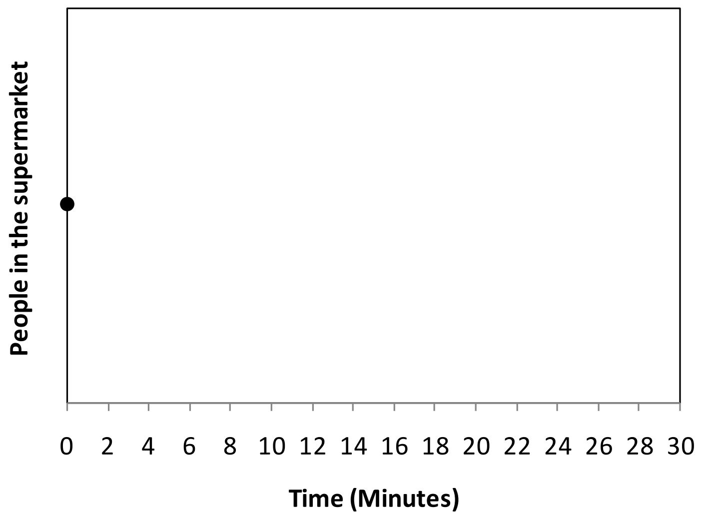
3.

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.



Part 2 or 3 [post-test]

Subject No: __________

Gender: □ Male □ Female

Age: ________

Educational background: □ Natural Sciences, Technology, Engineering or Mathematics
□ Social Sciences □ Humanities
□ Other: ____________________________________

Is English your native language? □ Yes □ No

Have you done or learned about this kind of task before? □ Yes □ No

Do you think your educational background has prepared you for this kind of task? □ Yes □ No

Please write down any comment you may have about this experiment, and your participation in it.

____
____
____
____
____

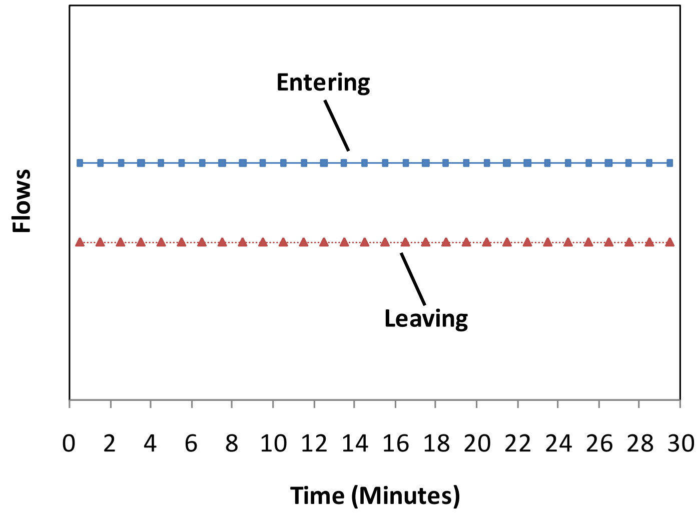
Please bring the completed test to the experiment leader to get your payment and to sign your name for having received the money. Again thanks so much for participating.

Part 2 [treatment for T2]

Subject No: __________

1.

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

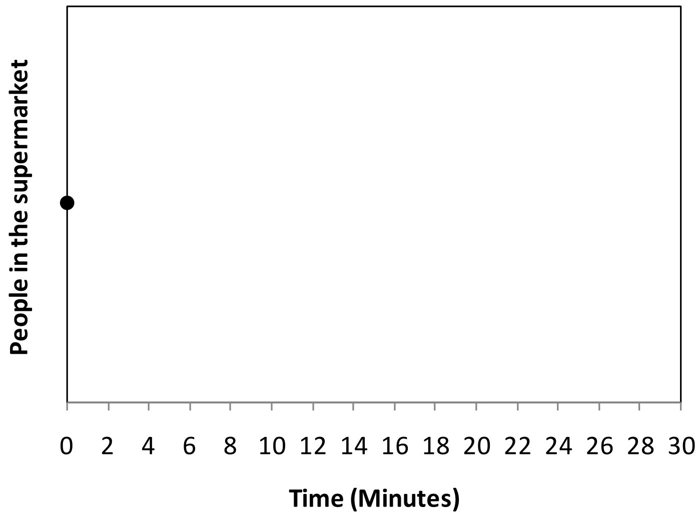


a. Which flow is larger? □ Entering larger than leaving □ Leaving larger than entering □ Both same

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.



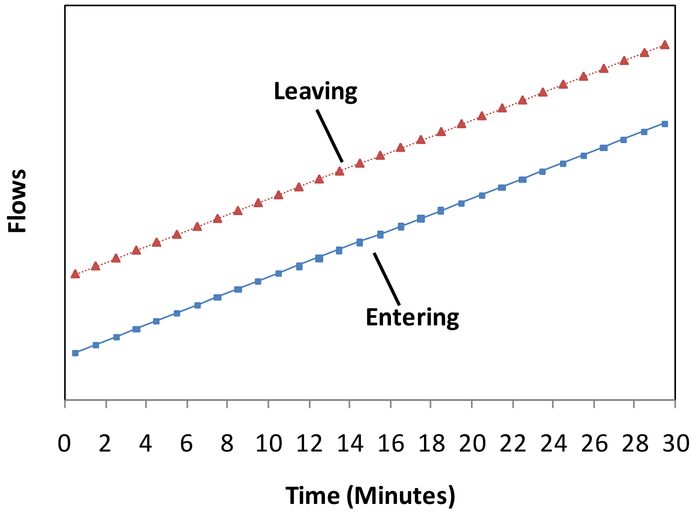
Please raise your hand after having completed this question. Do not continue to the next question before administrator has commented on your answer.

Part 2 [treatment for T2]

Subject No: __________

2.

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

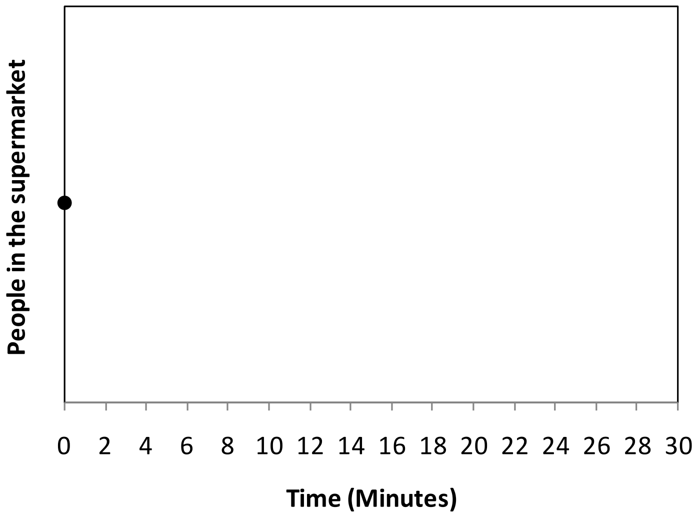


a. Which flow is larger? □ Entering larger than leaving □ Leaving larger than entering □ Both same

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.



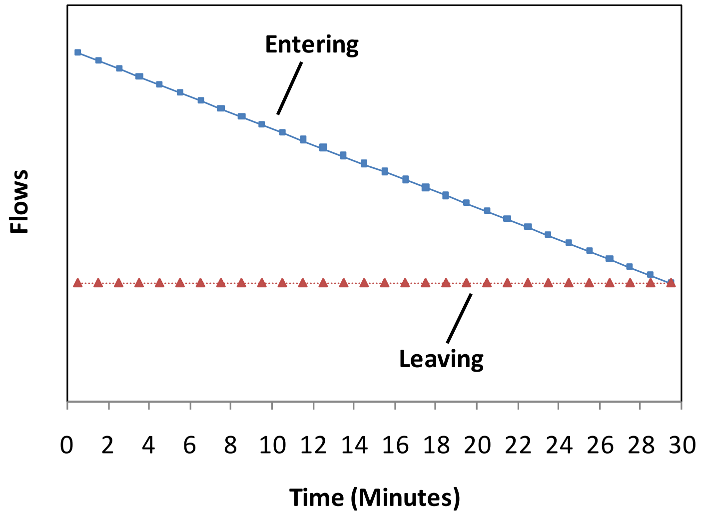
Please raise your hand after having completed this question. Do not continue to the next question before administrator has commented on your answer.

Part 2 [treatment for T2]

Subject No: __________

3.

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

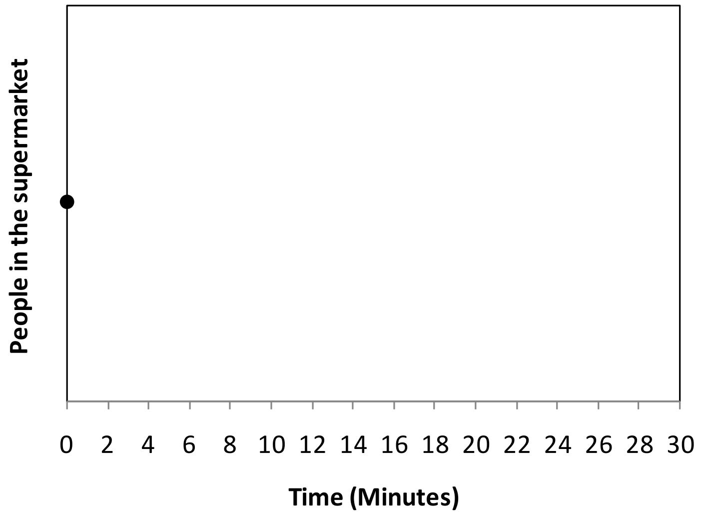


a. Which flow is larger? □ Entering larger than leaving □ Leaving larger than entering □ Both same

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.



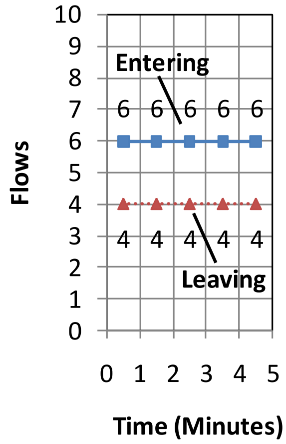
Please raise your hand after having completed this question. Do not continue to the next question before administrator has commented on your answer.

Part 2 [treatment for T3]

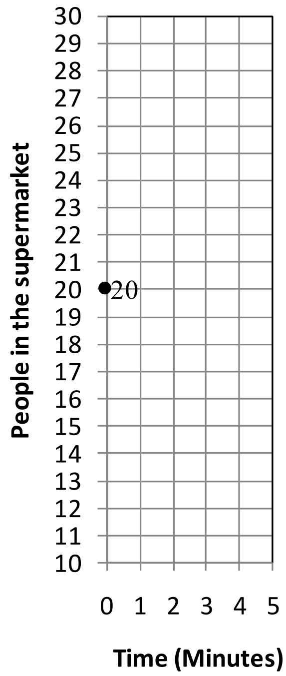
Subject No: __________

1.

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.

Part 2 [treatment for T3]

Subject No: __________

Now check your answer to Part 1 Question 1.

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

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

____
____
____

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

____
____
____

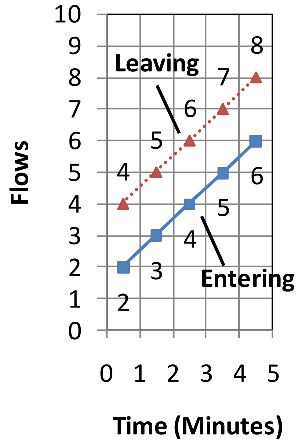
Continue to the next question after having completed the question.

Part 2 [treatment for T3]

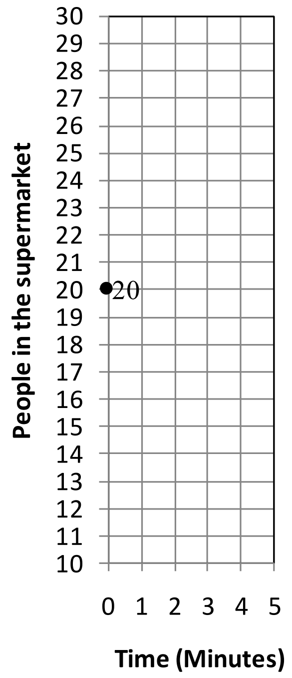
Subject No: __________

2.

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.

Part 2 [treatment for T3]

Subject No: __________

Now check your answer to Part 1 Question 2.

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

i. (Skip this if your answer to Part 1 Question 2 was right)
   Why should the number of people in the supermarket **decrease**?

   ____
   ____
   ____

ii. (Skip this if your answer to Part 1 Question 2 was right)
   Why should the number of people decrease **at a constant rate**?

   ____
   ____
   ____

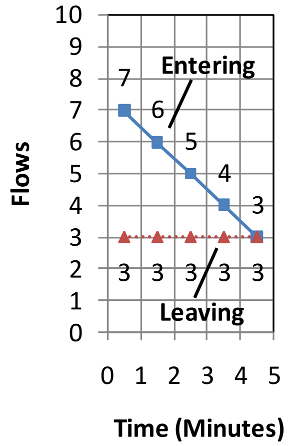
Continue to the next question after having completed the question.

Part 2 [treatment for T3]

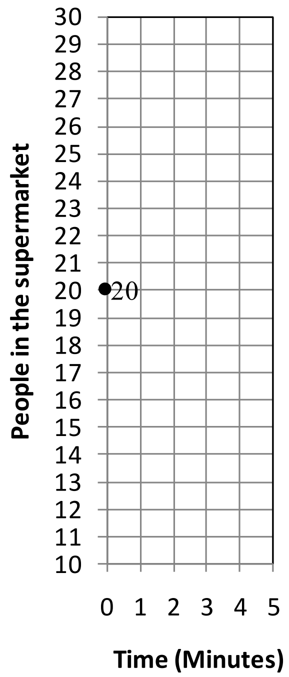
Subject No: __________

3.

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.

Part 2 [treatment for T3]

Subject No: __________

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**?

   ____
   ____
   ____

Please raise your hand after having completed the question.

Appendix B Helper guides

The experiment requires a number of well-trained helpers to help on carrying out the procedure. Each helper will be trained in one particular treatment and took care of assigned subjects from that treatment (e.g. A1-A5). Treatment T1 (the base treatment) is just giving pre-test and post-test one at a time, therefore no helper guide is needed. Try to source helpers who are familiar with graphical integration to help on treatment T2 (Graphical integration guidance) and T3 (Running total and reflection). It is advisable to get helpers who are more used to qualitative thinking to be trained on treatment T2 while those who are good at calculation (quantitative) in charge of treatment T3.

To train helpers, first let the helper read the helper guide of his or her treatment. Then carry out the treatment with the helper (the helper is the participant) so that he or she can comprehend the helper guide better. Since practice makes perfect, create opportunity for helper to practise (e.g. rehearsal with other helpers being the participants) and monitor the steps helper carried out. Correct it if it is wrong.

Well-trained helpers are the important element of this experiment. Please recruit and plan out in advance.

Helper Guide (for treatment B Graphical integration guidance)

Objective:
To help subjects learn graphical integration.

The test papers are divided into 3 parts:
1. Part 1 – 3 graphical integration questions
2. Part 2 – 3 graphical integration guidance questions
3. Part 3 – 3 graphical integration questions

Things needed to prepare:
Blue pens for subjects, red pen for yourself. Test papers (Part 1, 2 and 3) with Subject No. filled.

Subjects:
Each **well-trained** helper should be able to take care of 5 assigned subjects (e.g. B1, B2, B3, B4 and B5) at the same time.

Duty:
When your assigned subject **raise hand**, approach the subject, based on the condition, take the corresponding action:

Condition	Action
Subject complete Part 1	Take away Part 1, hand in Part 2.
Subject complete Part 2 question (2 times per subject)	1. Check the subject's answer. 2. Then take out subject's Part 1 question and check. 3. Direct subject move to next question by turn to next page.
Subject complete Part 2 Question 3	1. Check the subject's answer. 2. Then take out subject's Part 1 question and check. 3. Take away all papers, hand in Part 3.

Experiment flow from single subject perspective
(**Helper actions are marked as bold red color**):

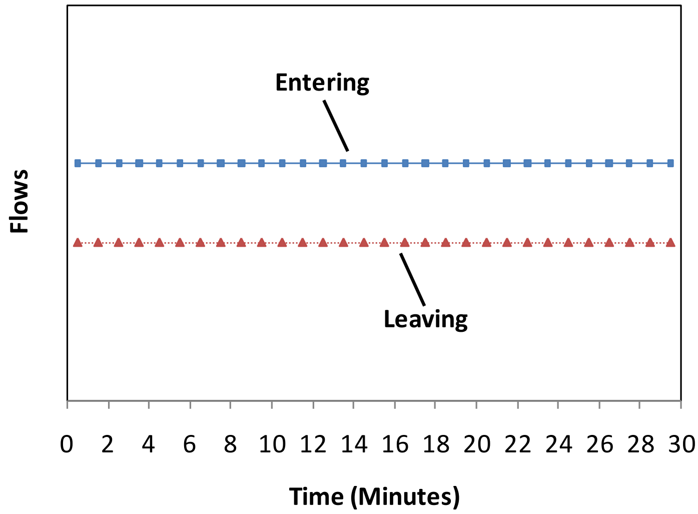
1. Experiment leader place Part 1 on table before start, then assign the subject to helper (e.g. B1-B5). **Helper takes the corresponding subjects' Part 2 and 3 tests.**
    - Experiment leader read out general instruction, clarify if any question.
    - Experiment start, subjects doing Part 1.

2. Subject completed Part 1 and raise hand.

    i. **Take away Part 1, hand in Part 2.**

3. Subject completed Part 2 Question 1 and raise hand.
   i. **Check Part 2 Question 1 (Please refer page 51-52 on how to check them. Below shows the right answer).**

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

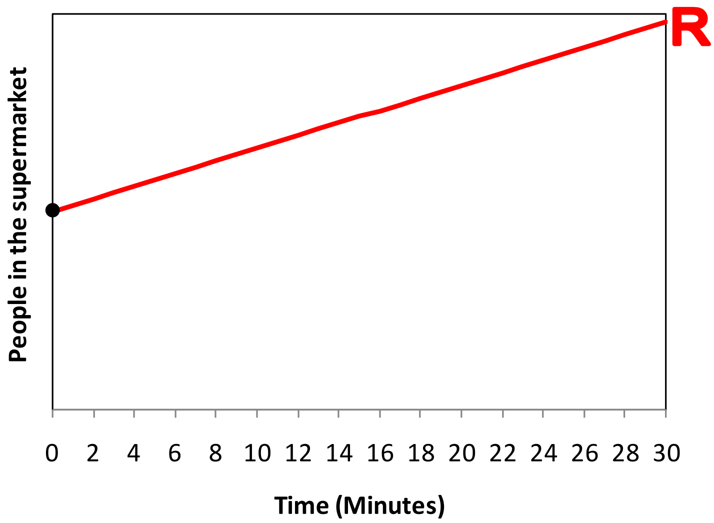


a. Which flow is larger? ☐ Entering larger than leaving **R** ☐ Leaving larger than entering ☐ Both same

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

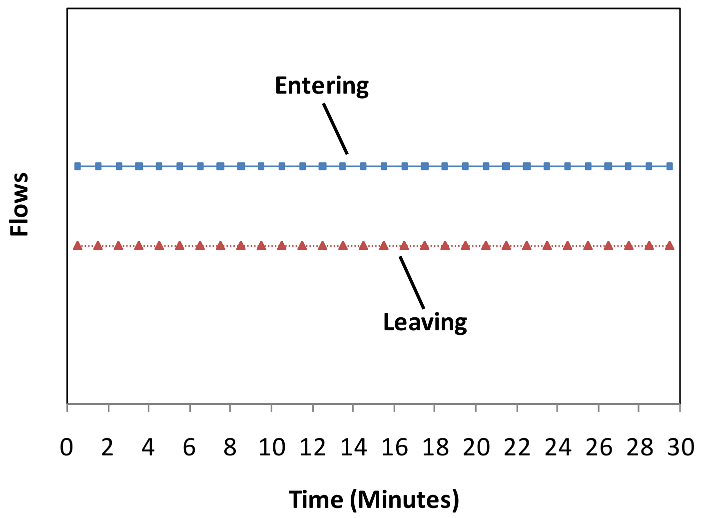
c. (Skip this if no change) How will the change in the number of people be?
☐ increase or decrease at a faster and faster rate ☐ increase or decrease at a slower and slower rate
☐ increase or decrease at a constant rate **R**

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.



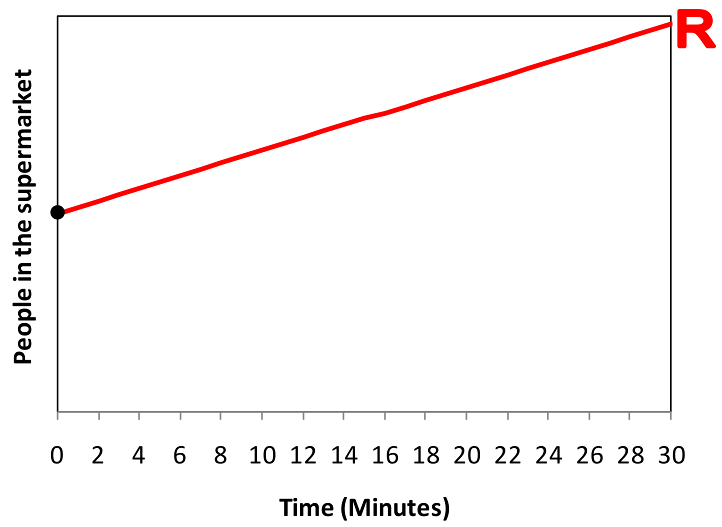
ii. **Take out subject's Part 1 Question 1 and check his/her answer by referring to Part 2 Question 1 right answer.**

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.

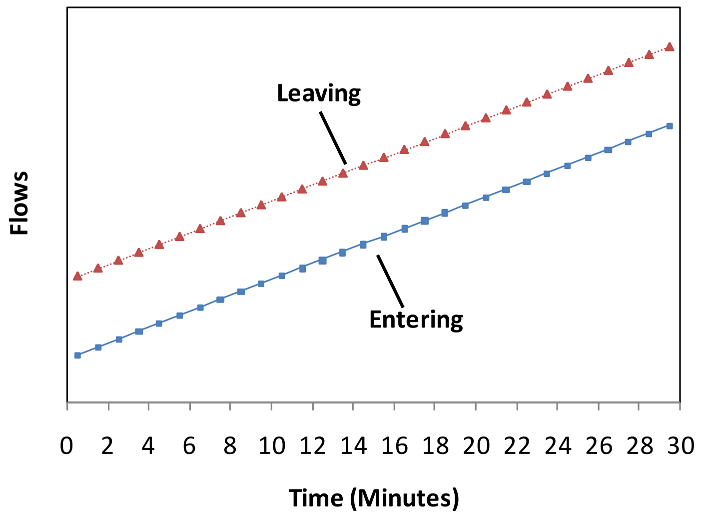
In other words, draw a line or curve to show how the number of people in the supermarket changes over the 30 minute interval, starting from the black dot (●) in the space below.



iii. **Direct subject move to Part 2 Question 2.**

4. Subject completed Part 2 Question 2 and raise hand.
   i. **Check Part 2 Question 2.**

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

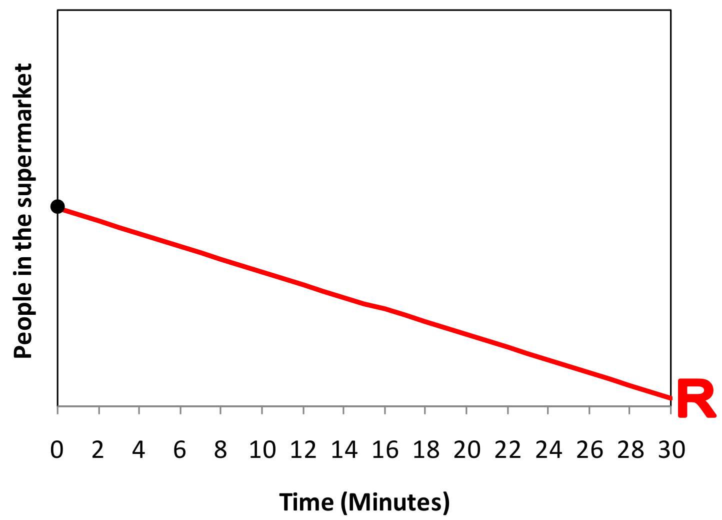


a. Which flow is larger? □ Entering larger than leaving □ Leaving larger than entering R □ Both same

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

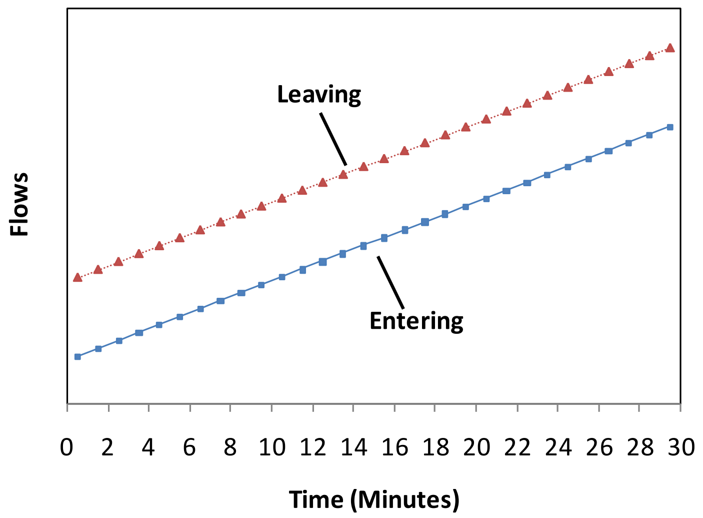
c. (Skip this if no change) How will the change in the number of people be?
□ increase or decrease at a faster and faster rate □ increase or decrease at a slower and slower rate
□ increase or decrease at a constant rate R

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.



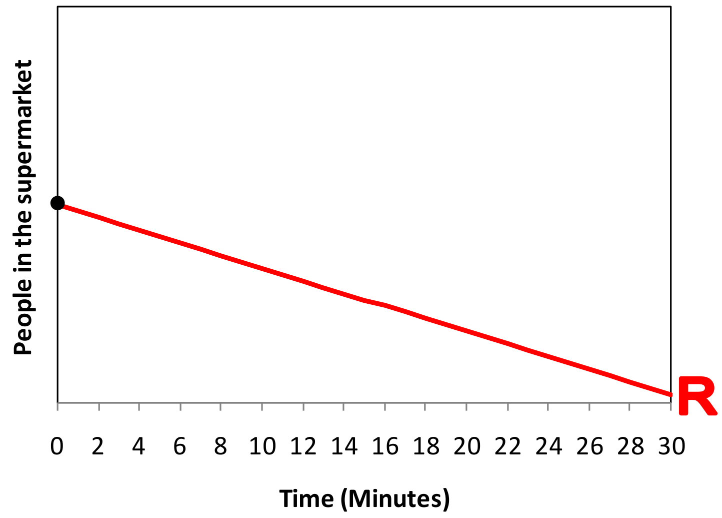
ii. Take out subject's Part 1 Question 2 and check his/her answer by referring to Part 2 Question 2 right answer.

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.

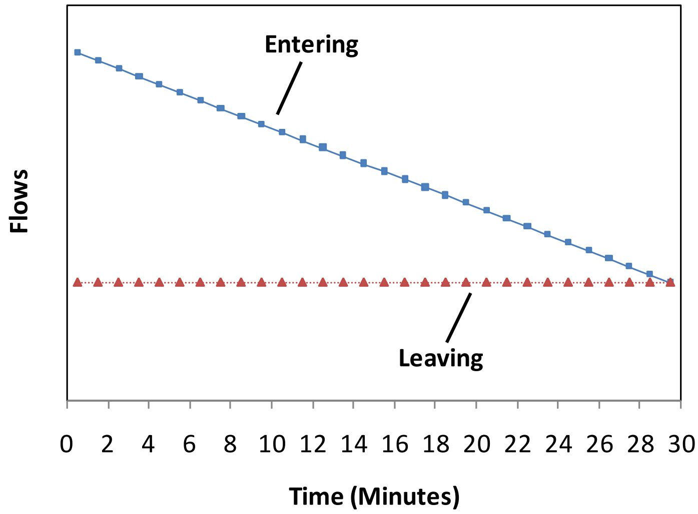
In other words, draw a line or curve to show how the number of people in the supermarket changes over the 30 minute interval, starting from the black dot (●) in the space below.



iii. Direct subject move to Part 2 Question 3.

5. Subject completed Part 2 Question 3 and raise hand.
   i. **Check Part 2 Question 3:**
      Subject can circle, check, cross or other marking to indicate their answer.
      If it is right, mark 'R', else circle the right answer.

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

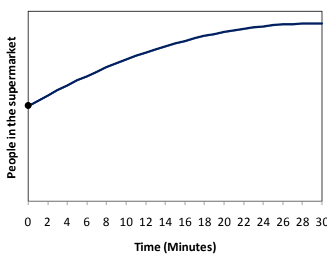
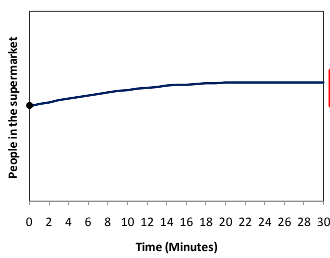
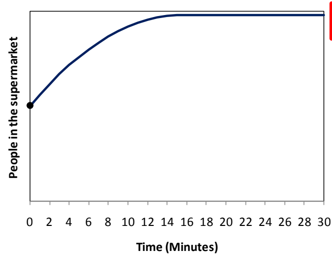


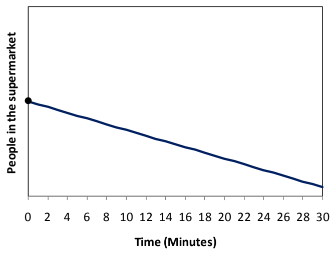
a. Which flow is larger? ☐ Entering larger than leaving R ☐ Leaving larger than entering ☐ Both same

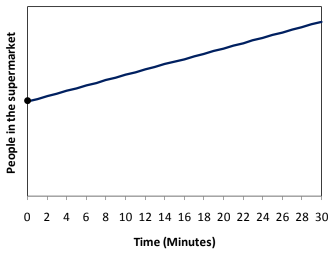
b. In what direction will the number of people in the supermarket change?
   ☑ Increase R ☐ Decrease ☐ No change

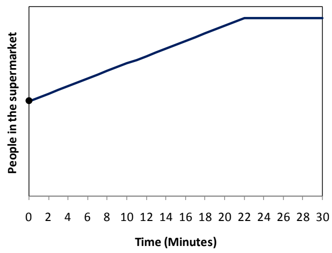
c. (Skip this if no change) How will the change in the number of people be?
   ☑ increase or decrease at a faster and faster rate ☐ increase or decrease at a slower and slower rate
   ☐ increase or decrease at a constant rate

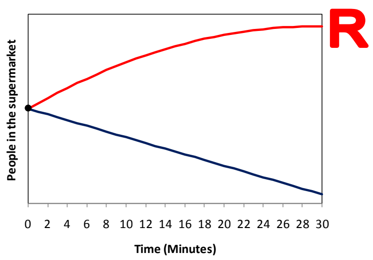
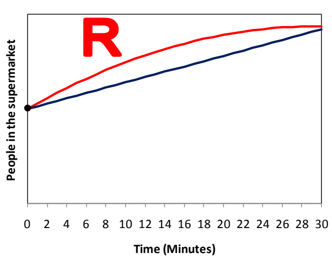
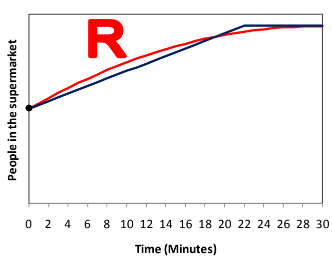
d. In the space below, graph the number of people in the supermarket over the 30 minute interval. You do not need to specify numerical values. The dot at time zero shows the initial number of people in the supermarket.

All these drawing are considered right:		
 R	 R	 R
Direction right, shape right	Less steep and becomes flat sooner, but its direction and shape are still right	Steeper also OK, as long as its direction and shape are still right

<table>
<tr><th colspan="3">All these are wrong:</th></tr>
<tr>
<td>

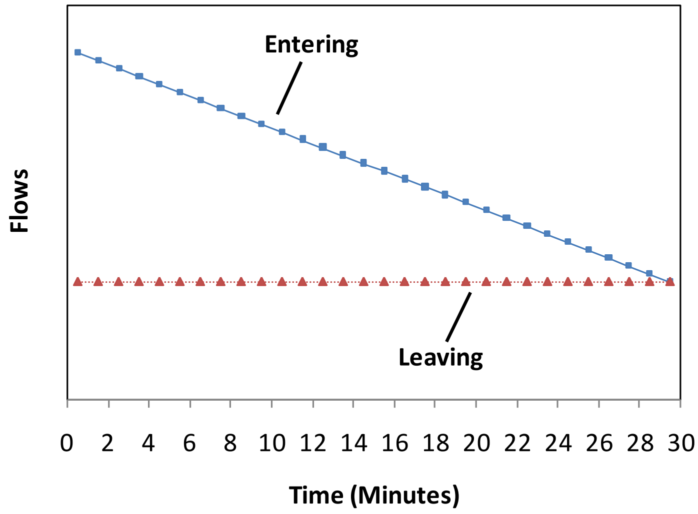
Direction already wrong</td>
<td>

Direction right, but shape wrong</td>
<td>

Shape is still not right (curve, not lines)</td>
</tr>
<tr><th colspan="3">To check them:</th></tr>
<tr>
<td>
</td>
<td>
</td>
<td>
</td>
</tr>
</table>

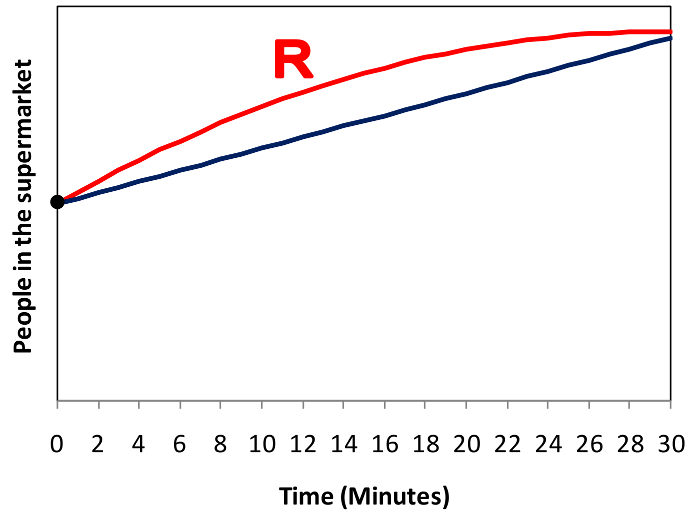
ii. **Take out subject's Part 1 Question 3 and check his/her answer by referring to Part 2 Question 3 right answer.**

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.

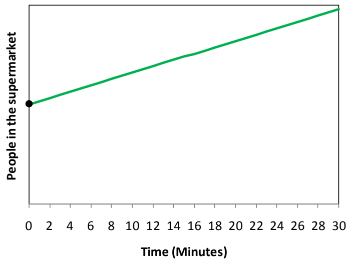


6. **Take away all papers, hand in Part 3. Viola! It's done. Thanks for helping me carry out the treatment!**
   i. Subject completed Part 3, leave seat and bring the completed test to the experiment leader.

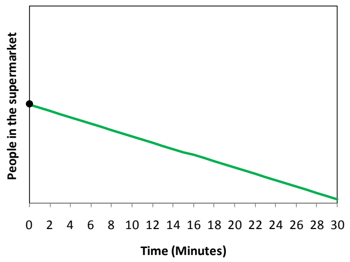
Summary

In summary, right answers for each question (print out this page if you need it to help you check the questions, but **don't let subjects to see it**):

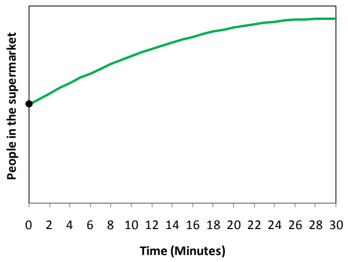
1.



2.



3.



Any question? Please ask me if you need any clarification so that you can help me to carry out this experiment smoothly.

Helper Guide (for treatment C Running total and reflection)

Objective:
To help subjects learn graphical integration by exploiting their ability to calculate running total (keep a running tally of stock and add or subtract the flows at each point of time).

The test papers are divided into 3 parts:
1. Part 1 – 3 graphical integration questions
2. Part 2 – 3 * (running total + reflective) questions
3. Part 3 – 3 graphical integration questions

Basic idea:
1. First, let subjects try graphical integration questions in Part 1 – what is their **expected answer**?
2. Then, in Part 2, ask subjects to calculate running total – what is the **right answer**?
3. Subjects should be surprised if their expected answer is different from the right answer, and hopefully they can figure out **why** during reflection.
4. Lastly, give subjects another set of graphical integration questions (Part 3) to see if they have learned.

Things needed to prepare:
Blue pens for subjects, red pen for yourself. Test papers (Part 1, 2 and 3) with Subject No. filled, Part 1 and 2 should be 1 page per sheet. Staple Part 2.

Subjects:
Each **well-trained** helper should be able to take care of 3 assigned subjects (e.g. C1, C2 and C3) at the same time.

Duty:
When your assigned subject **raise hand**, approach the subject, based on the condition, take the corresponding action:

Condition	Action
Subject complete Part 1	Take away Part 1, hand in Part 2.
Subject complete Part 2 plotting (3 times per subject)	1. Take away previous Part 1 question, if any (see below). 2. Check the plotting. 3. Then take out subject's Part 1 question and check. 4. Direct subject continue by turn to next page. 5. Leave Part 1 question with the subject for reflection (will collect it next checking, see above).
Subject complete Part 2 all questions	Take away all papers, hand in Part 3. (Now you should have subject's all Part 1 and Part 2 questions)

Experiment flow from single subject perspective
(**Helper actions are marked as bold red color**):

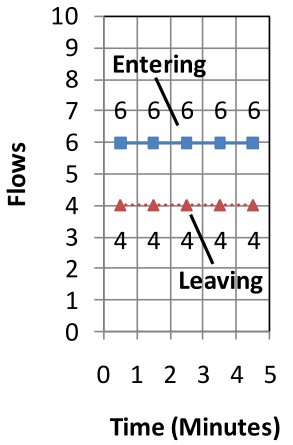
1. Experiment leader place Part 1 on table before start, then assign the subject to helper (e.g. C1-C5). **Helper takes the corresponding subjects' Part 2 and 3 tests.**
    - Experiment leader read out general instruction, clarify if any question.
    - Experiment start, subjects doing Part 1.

2. Subject completed Part 1 and raise hand.

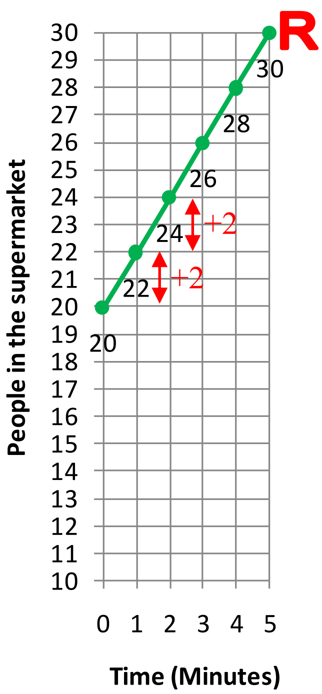
    i. **Take away Part 1, hand in Part 2.**

3. Subject completed Part 2 Question 1 and raise hand.
   i. **Check Part 2 Question 1 plotting (Please refer page 63-64 on how to check it. Below shows the right answer).**

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

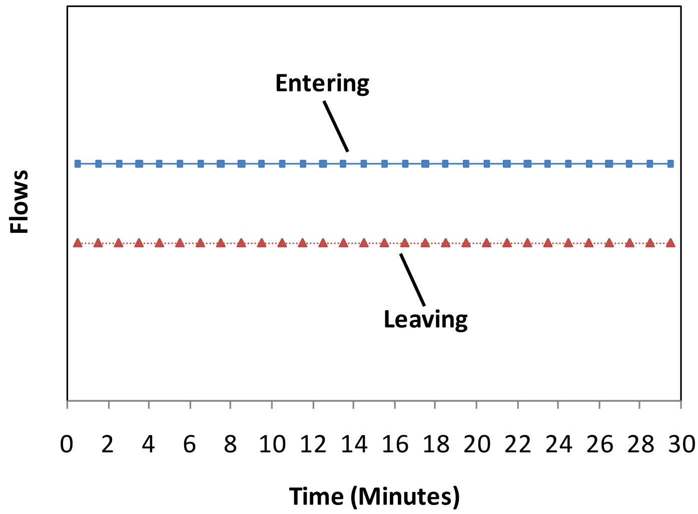


In the space below, graph the number of people in the supermarket over the 5 minute interval. The dot shows that at time zero there are 20 people in the supermarket. *Calculate how the number of people in the supermarket develops from minute to minute and plot the numbers accurately in the graph.*



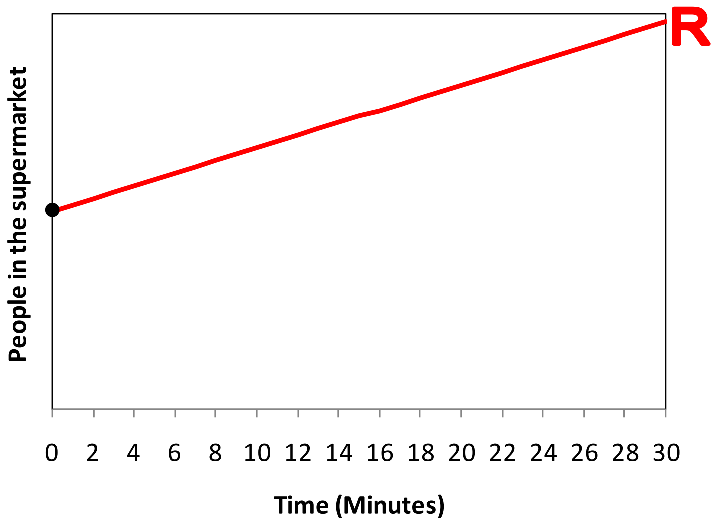
ii. **Take out subject's Part 1 Question 1 and check his/her answer by referring to Part 2 Question 1 right answer (Please refer page 65-66 on how to check it. Below shows the right answer).**

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.



iii. **Direct subject continue next page. Leave Part 1 question with subject for reflection (They can skip these reflective queries if their answer to Part 1 Question 1 was right).**

Now check your answer to Part 1 Question 1.

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

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

____

____

____

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

____

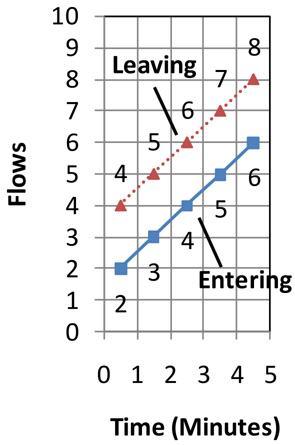
____

____

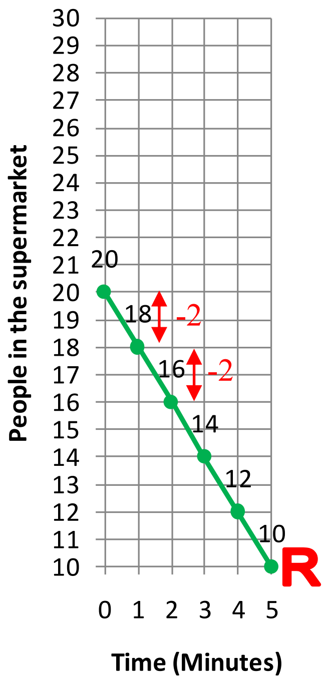
Continue to the next question after having completed the question.

4. Subject completed reflective queries (they can skip if their answer to Part 1 Question 1 was right) and Part 2 Question 2 plotting and raise hand.
    i. **Check Part 2 Question 2 plotting.**

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

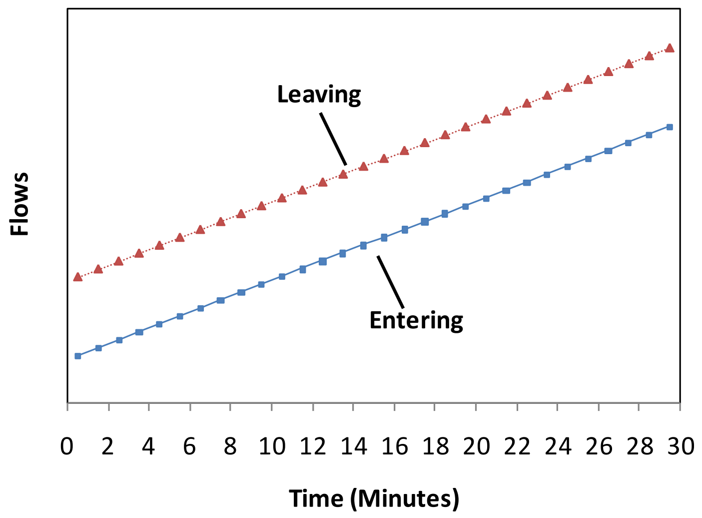


In the space below, graph the number of people in the supermarket over the 5 minute interval. The dot shows that at time zero there are 20 people in the supermarket. *Calculate how the number of people in the supermarket develops from minute to minute and plot the numbers accurately in the graph.*



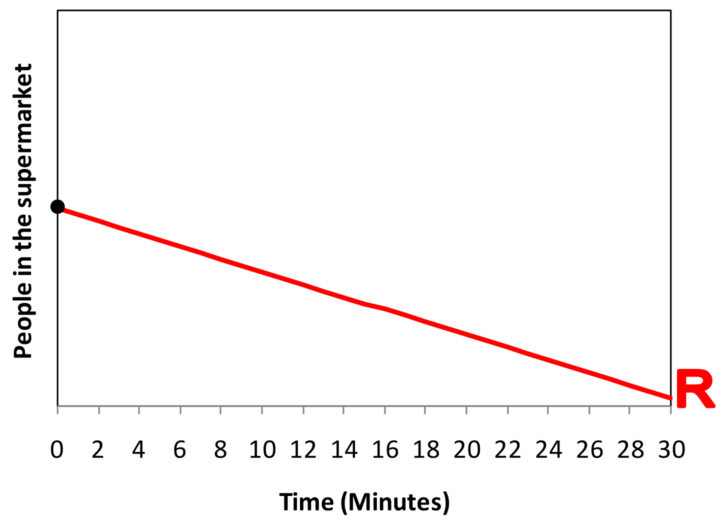
ii. Take out subject's Part 1 Question 2 and check his/her answer by referring to Part 2 Question 2 right answer.

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.



iii. **Direct subject continue next page. Leave Part 1 question with subject for reflection (They can skip these reflective queries if their answer to Part 1 Question 2 was right).**

Now check your answer to Part 1 Question 2.

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

ii. (Skip this if your answer to Part 1 Question 2 was right)
Why should the number of people in the supermarket **decrease**?

____

____

____

iv. (Skip this if your answer to Part 1 Question 2 was right)
Why should the number of people decrease **at a constant rate**?

____

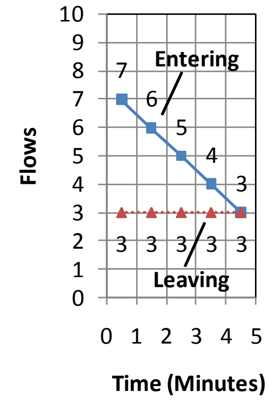
____

____

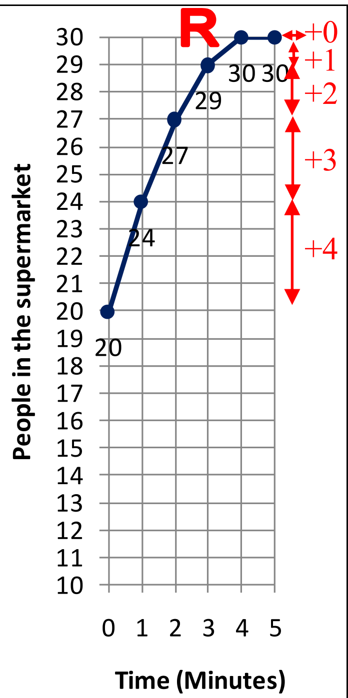
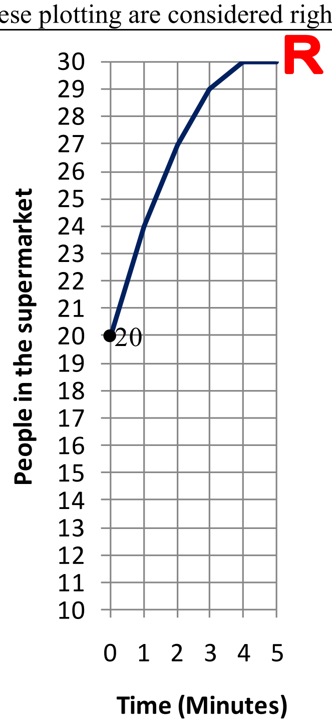
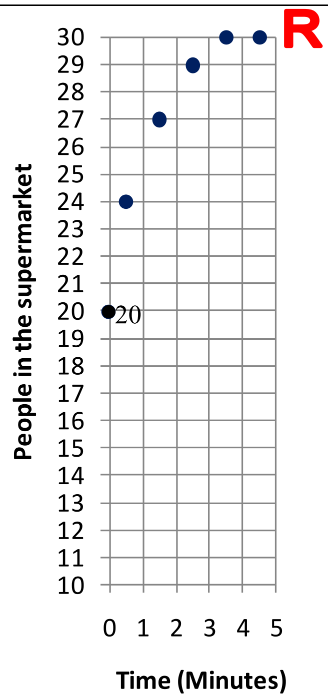
Continue to the next question after having completed the question.

5. Subject completed reflective queries (they can skip if their answer to Part 1 Question 2 was right) and Part 2 Question 3 plotting and raise hand.
   i. **Check Part 2 Question 3 plotting.**
      - **Many times subjects know how to calculate running total and plot accurately, but not doing so because they want to finish it quickly. So we will check their plotting and feedback to them, to urge them to answer more carefully. Most will get the next plotting right.**

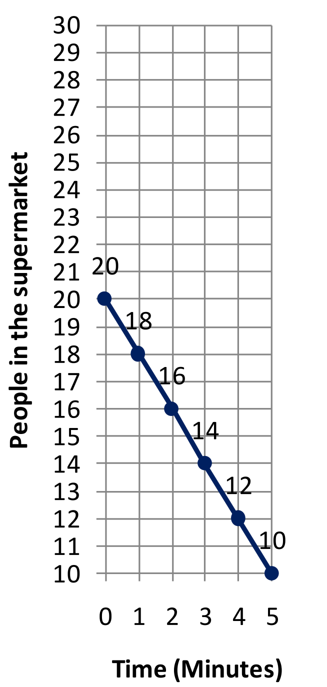
The graph below shows the number of people **entering** and **leaving** a supermarket each minute over a 5 minute period.



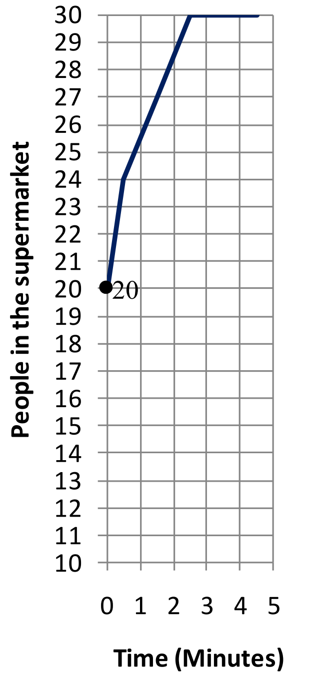
In the space below, graph the number of people in the supermarket over the 5 minute interval. The dot shows that at time zero there are 20 people in the supermarket. *Calculate how the number of people in the supermarket develops from minute to minute and plot the numbers accurately in the graph.*

All these plotting are considered right:		
		
Perfect answer	No number no point also OK, as long as it pass all points	No line or plot in the middle also OK, just make sure all points are right

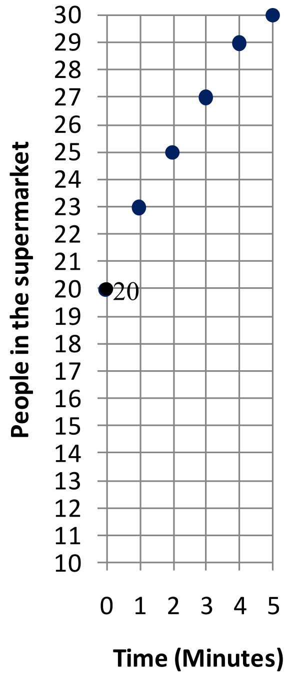
All these are wrong:



Totally wrong

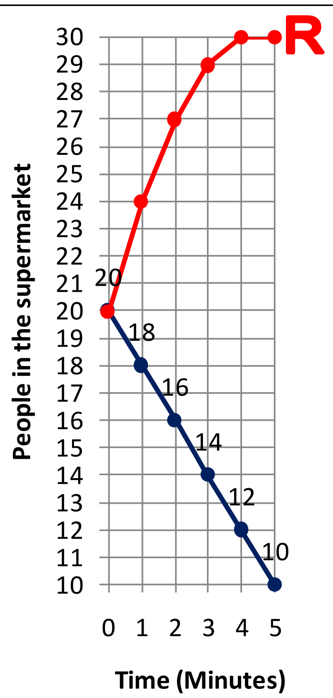


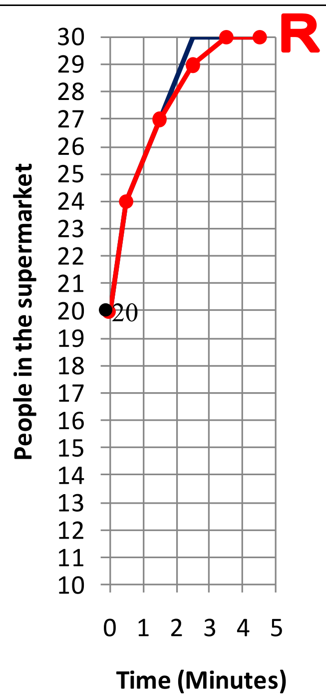
Points at Min 2.5 and 3.5 are wrong



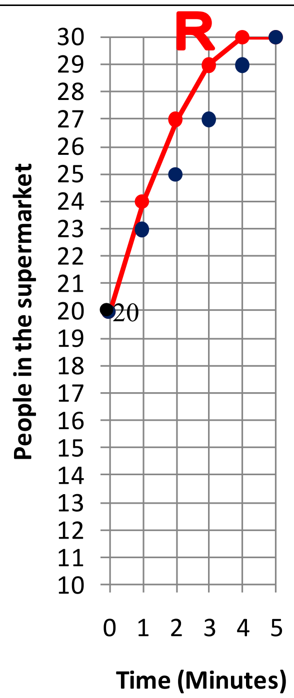
All points are inaccurate, distorted the shape

To check them:



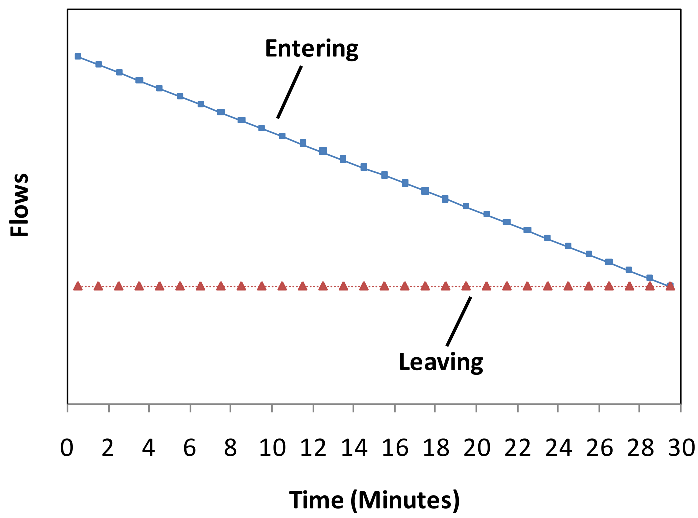


If subject plot in the middle, we also plot the right answer in the middle, so as not to confuse the subject



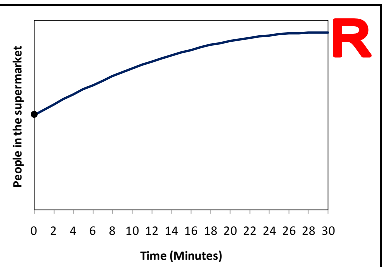
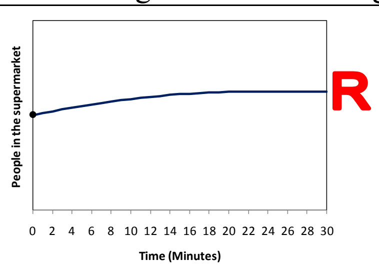
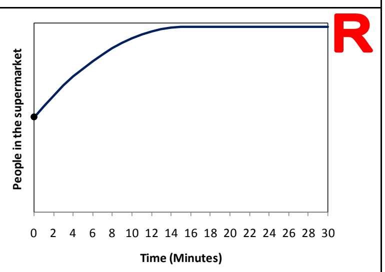
ii. **Take out subject's Part 1 Question 3 and check his/her answer by referring to Part 2 Question 3 right answer.**
   - **Sometimes subjects don't know how to compare Part 1 and Part 2. So we take out subject's Part 1 drawing and check it by referring to Part 2 right answer.**

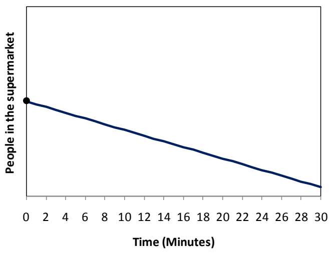
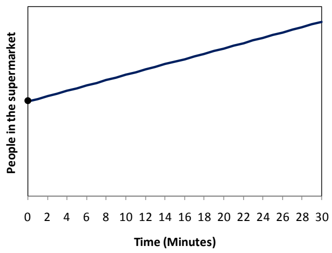
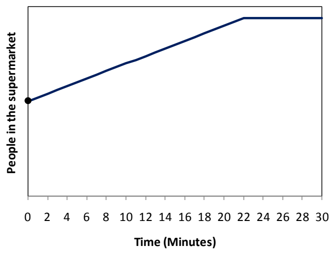
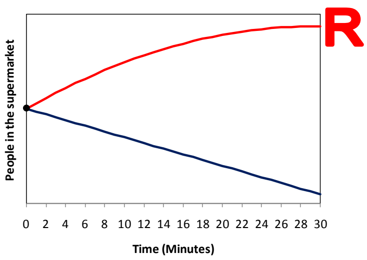
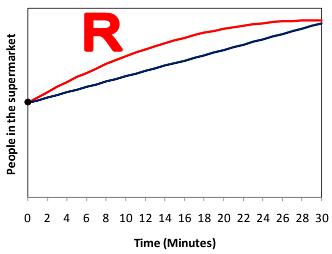
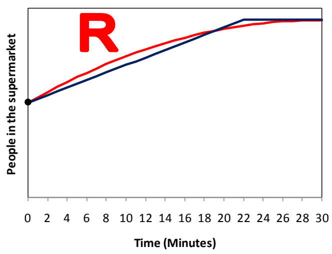
The graph below shows the number of people **entering** and **leaving** a supermarket over a 30 minute period.



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.

All these drawing are considered right:		
		
Direction right, shape right	Less steep and becomes flat sooner, but its direction and shape are still right	Steeper also OK, as long as its direction and shape are still right

All these are wrong:		
 Direction already wrong	 Direction right, but shape wrong	 Shape is still not right (curve, not lines)
To check them:		
		

iii. **Direct subject continue next page. Leave Part 1 question with subject for reflection (They can skip these reflective queries if their answer to Part 1 Question 3 was right).**

Now check your answer to Part 1 Question 2.

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

iii. (Skip this if your answer to Part 1 Question 2 was right)
Why should the number of people in the supermarket **decrease**?

____
____
____

v. (Skip this if your answer to Part 1 Question 2 was right)
Why should the number of people decrease **at a constant rate**?

____
____
____

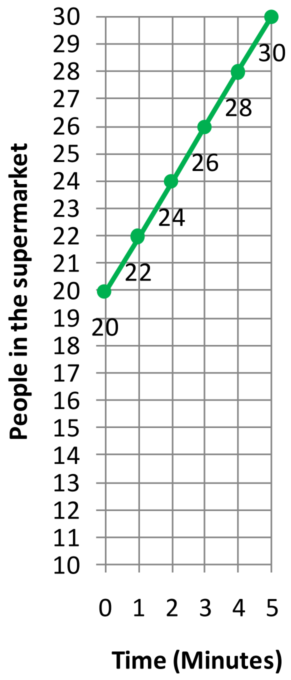
6. Subject completed reflective queries (they can skip if their answer to Part 1 Question 3 was right) and raise hand.
   i. **Take away all papers, hand in Part 3. Viola! It's done. Thanks for helping me carry out the treatment!**
      - Subject completed Part 3, leave seat and bring the completed test to the experiment leader.

Summary

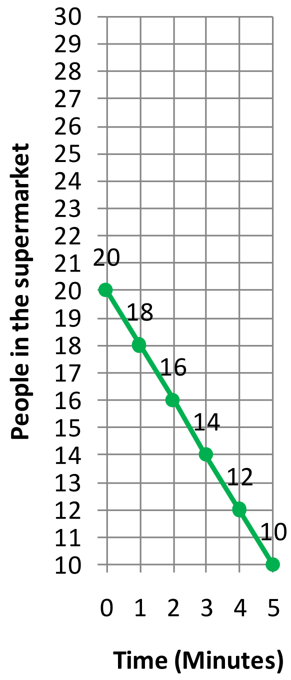
In summary, right answers for each question (print out this page if you need it to help you check the questions, but **don't let subjects to see it**):

Part 2:

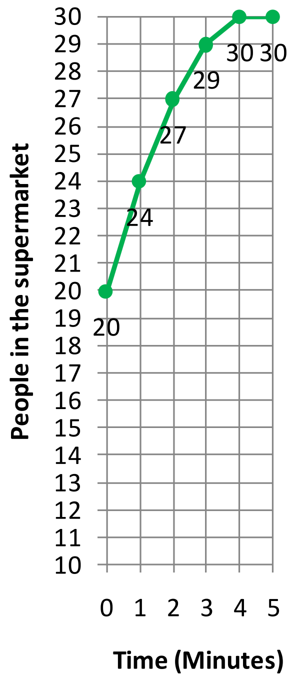
1.



2.

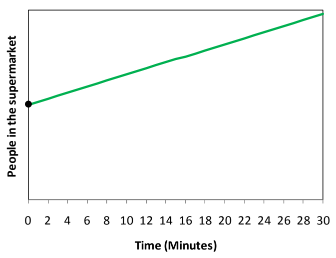


3.

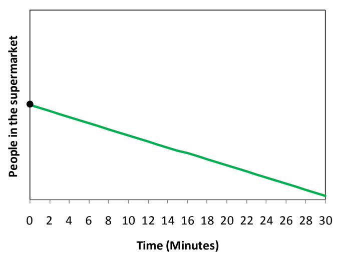


Part 1:

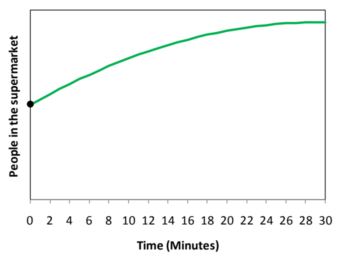
1.



2.



3.



Any question? Please ask me if you need any clarification so that you can help me to carry out this experiment smoothly.

Appendix C Data

Demography data

Empty cell indicates missing value.

Subject	Treatment	Gender	Age	Background	EnglishNative	TaskBefore	EducationPrepared
A1	1	2	21	SocialSciences	0	0	1
A2	1	2	19	SocialSciences	0	0	0
A3	1	1	21	SocialSciences	0	0	0
A4	1	2	20	SocialSciences	0	0	0
A5	1	2	19		0	0	0
A6	1	2	20	SocialSciences	0	0	0
A7	1	2	22	SocialSciences	0	1	0
A8	1	2	21	SocialSciences	0	0	0
A9	1	2	20	Psychology	0	0	0
A10	1	2	20	SocialSciences	0	0	0
A11	1	2	21	SocialSciences	0	0	0
A12	1	2	20	Humanities	0	0	1
B1	2	2	21	HighSchool	0	0	0
B2	2	2	20	SocialSciences	0	0	0
B4	2	2	37	Economy	0	0	0
B5	2	2	23	SocialSciences	0	0	0
B6	2	2	21	SocialSciences	0	0	1
B7	2	1	21	SocialSciences	0	0	0
B8	2	2	21	SocialSciences	0	0	0
B9	2	2	20	SocialSciences	0	0	0
B10	2	1	20	SocialSciences	0	0	0
C2	3	1	21	SocialSciences	0	0	0
C4	3	2	20	SocialSciences	0	0	0
C5	3	2	20	SocialSciences	0	0	0
C6	3	2	38	SocialSciences	0	0	0
C8	3	2	20	SocialSciences	0	0	0
C9	3	2	20	SocialSciences	0	0	0
A13	1	1	20	SocialSciences	0	0	0
A14	1	2	20	SocialSciences	0	0	0
A15	1	2	22	SocialSciences	0	0	0
A17	1	1	19	SocialSciences	0	0	0
B11	2	2	21	SocialSciences	0	1	1
B12	2	2	19	SocialSciences	0	0	0
B13	2	1	26	SocialSciences	0	0	0

B15	2	2	22	Humanities	0	0	0
B16	2	1	21	SocialSciences	0	0	0
C10	3	1	21	SocialSciences	0	0	0
C11	3	2	20	SocialSciences	0	0	0
C12	3						
C14	3	1	20	SocialSciences	0	0	0
C17	3	1	24	SocialSciences	0	0	0
C18	3	2	21	SocialSciences	0	0	0
C19	3	2	21	SocialSciences	0	0	0
C20	3	2	21	SocialSciences	0	0	0
C21	3						
C22	3						

Pre-test data

Prefix ‘Pr’ stands for pre-test; suffix ‘D’ stands for direction; ‘S’ stands for shape; so ‘PrQ1D’ means “Pre-test Question 1 direction right?”. ‘PrQ1Cor’ means “Pre-test Question 1 correlate?”; ‘PrQ2CorIn’ means “Pre-test Question 2 correlate with inflow?”; ‘PrQ2CorNet’ means “Pre-test Question 2 correlate with net flow?”; ‘PrQ3CorP’ means “Pre-test Question 3 correlate positively?” and ‘PrQ3CorN’ means “Pre-test Question 3 correlate negatively?”.

Subject	PrQ1D	PrQ1S	PrQ1	PrQ1 Cor	PrQ2D	PrQ2S	PrQ2	PrQ2 CorIn	PrQ2 CorNet	PrQ3D	PrQ3S	PrQ3	PrQ3 CorP	PrQ3 CorN
A1	0	1	FALSE	1	1	1	TRUE	0	0	0	0	FALSE	1	0
A2	0	1	FALSE	1	0	1	FALSE	0	1	0	0	FALSE	1	0
A3	0	1	FALSE	1	0	1	FALSE	1	0	0	0	FALSE	0	0
A4	0	1	FALSE	1	0	1	FALSE	0	1	1	0	FALSE	0	1
A5	1	1	TRUE	0	0	0	FALSE	0	0	1	1	TRUE	0	0
A6	1	0	FALSE	0	1	0	FALSE	0	0	0	0	FALSE	1	0
A7	0	1	FALSE	1	0	1	FALSE	0	1	0	0	FALSE	1	0
A8	0	1	FALSE	1	0	1	FALSE	1	0	1	0	FALSE	0	1
A9	1	1	TRUE	0	1	1	TRUE	0	0	1	1	TRUE	0	0
A10	0	1	FALSE	1	0	0	FALSE	0	0	0	0	FALSE	0	0
A11	0	1	FALSE	1	0	1	FALSE	0	1	0	0	FALSE	1	0
A12	1	1	TRUE	0	0	1	FALSE	1	0	0	0	FALSE	1	0
B1	0	1	FALSE	1	1	1	TRUE	0	0	1	0	FALSE	0	1
B2	0	1	FALSE	1	1	1	TRUE	0	0	0	0	FALSE	1	0
B4	1	1	TRUE	0	0	1	FALSE	0	1	0	0	FALSE	1	0
B5	0	1	FALSE	1	0	1	FALSE	0	1	0	0	FALSE	1	0
B6	1	1	TRUE	0	1	1	TRUE	0	0	0	1	FALSE	0	0
B7	0	1	FALSE	1	0	1	FALSE	0	1	0	0	FALSE	0	0

B8	0	0	FALSE	0	0	0	FALSE	0	0	0	0	FALSE	0	1
B9	0	1	FALSE	1	0	1	FALSE	0	1	0	0	FALSE	0	0
B10	0	1	FALSE	1	0	1	FALSE	0	1	0	0	FALSE	1	0
C2	0	1	FALSE	1	0	1	FALSE	0	1	0	0	FALSE	1	0
C4	0	1	FALSE	0	0	1	FALSE	0	1	0	0	FALSE	1	0
C5	1	1	TRUE	0	1	1	TRUE	0	0	1	1	TRUE	0	0
C6	1	1	TRUE	0	1	1	TRUE	0	0	0	0	FALSE	0	0
C8	0	1	FALSE	1	0	0	FALSE	1	0	0	0	FALSE	1	0
C9	0	1	FALSE	1	0	0	FALSE	1	0	1	0	FALSE	0	1
A13	0	1	FALSE	1	1	1	TRUE	0	0	0	0	FALSE	0	0
A14	0	1	FALSE	1	0	1	FALSE	1	0	0	0	FALSE	1	0
A15	0	1	FALSE	1	0	1	FALSE	0	1					
A17	0	0	FALSE	0	0	0	FALSE	0	0	0	0	FALSE	0	0
B11	1	1	TRUE	0	0	1	FALSE	0	1	0	0	FALSE	1	0
B12	0	1	FALSE	1	0	1	FALSE	1	0	0	0	FALSE	1	0
B13	0	1	FALSE	1	1	0	FALSE	0	0	1	0	FALSE	0	0
B15	0	1	FALSE	1	0	1	FALSE	0	1	0	0	FALSE	1	0
B16	1	1	TRUE	0	1	1	TRUE	0	0	0	0	FALSE	1	0
C10	1	1	TRUE	0	0	1	FALSE	0	1	0	0	FALSE	1	0
C11	0	1	FALSE	1	0	1	FALSE	1	0	0	0	FALSE	1	0
C12	0	1	FALSE	1	0	1	FALSE	0	1	0	0	FALSE	1	0
C14	0	0	FALSE	0	0	1	FALSE	1	0	0	0	FALSE	0	0
C17	1	0	FALSE	0	1	0	FALSE	0	0	0	0	FALSE	0	0
C18	0	1	FALSE	1	0	1	FALSE	1	0	0	0	FALSE	1	0
C19	0	1	FALSE	1	0	1	FALSE	0	1	0	0	FALSE	1	0
C20	0	1	FALSE	1	0	1	FALSE	1	0	0	1	FALSE	0	0
C21	1	1	TRUE	0	0	1	FALSE	0	1	0	0	FALSE	1	0
C22	1	1	TRUE	0	1	1	TRUE	0	0	1	1	TRUE	0	0

Post-test data

Prefix ‘Po’ stands for post-test; the rest follows pre-test conventions.

Subject	PoQ1D	PoQ1S	PoQ1	PoQ1 Cor	PoQ2D	PoQ2S	PoQ2	PoQ2 CorIn	PoQ2 CorNet	PoQ3D	PoQ3S	PoQ3	PoQ3 CorP	PoQ3 CorN
A1	0	0	FALSE	1	1	1	TRUE	0	0	0	1	FALSE	0	0
A2	0	0	FALSE	1	0	1	FALSE	0	1	0	0	FALSE	1	0
A3	0	1	FALSE	1	0	0	FALSE	0	0	0	0	FALSE	0	0
A4	0	1	FALSE	0	1	1	TRUE	0	0	0	0	FALSE	1	0
A5	1	1	TRUE	0	1	0	FALSE	0	0	1	1	TRUE	0	0
A6	1	1	TRUE	0	1	0	FALSE	0	0	0	0	FALSE	0	0
A7	0	1	FALSE	1	0	1	FALSE	1	0	0	0	FALSE	1	0
A8	0	1	FALSE	1	0	1	FALSE	1	0	0	0	FALSE	1	0

A9	1	1	TRUE	0	1	0	FALSE	0	0	1	1	TRUE	0	0
A10	1	1	TRUE	0	0	0	FALSE	0	0	0	0	FALSE	0	0
A11	1	1	TRUE	0	0	1	FALSE	0	1	0	0	FALSE	1	0
A12	1	1	TRUE	0	0	1	FALSE	1	0	0	1	FALSE	0	0
B1	1	1	TRUE	0	1	1	TRUE	0	0	0	0	FALSE	0	0
B2	1	1	TRUE	0	1	1	TRUE	0	0	1	1	TRUE	0	0
B4	1	1	TRUE	0	0	1	FALSE	1	0	0	0	FALSE	1	0
B5	1	1	TRUE	0	1	1	TRUE	0	0	0	0	FALSE	1	0
B6	0	1	FALSE	0	1	1	TRUE	0	0	1	1	TRUE	0	0
B7	1	1	TRUE	0	1	1	TRUE	0	0	0	0	FALSE	0	0
B8	0	1	FALSE	0	0	1	FALSE	0	1	0	0	FALSE	1	0
B9	1	1	TRUE	0	1	1	TRUE	0	0	1	1	TRUE	0	0
B10	1	1	TRUE	0	1	1	TRUE	0	0	1	1	TRUE	0	0
C2	1	1	TRUE	0	0	0	FALSE	1	0	0	0	FALSE	1	0
C4	0	1	FALSE	0	1	1	TRUE	0	0	1	0	FALSE	0	1
C5	1	1	TRUE	0	1	1	TRUE	0	0	1	0	FALSE	0	1
C6	1	1	TRUE	0	1	1	TRUE	0	0	0	1	FALSE	0	0
C8	0	0	FALSE	0	0	0	FALSE	0	0	0	0	FALSE	1	0
C9	1	1	TRUE	0	1	1	TRUE	0	0	0	0	FALSE	1	0
A13	1	1	TRUE	0	0	1	FALSE	0	1	1	1	TRUE	0	0
A14	0	1	FALSE	1	0	1	FALSE	1	0	0	0	FALSE	1	0
A15	0	1	FALSE	1	0	1	FALSE	0	1	0	0	FALSE	0	0
A17	0	0	FALSE	0	0	0	FALSE	0	0	0	0	FALSE	0	0
B11	1	1	TRUE	0	1	1	TRUE	0	0	1	1	TRUE	0	0
B12	1	1	TRUE	0	1	1	TRUE	0	0	0	0	FALSE	0	0
B13	1	1	TRUE	0	1	1	TRUE	0	0	0	0	FALSE	0	0
B15	1	1	TRUE	0	1	1	TRUE	0	0	1	1	TRUE	0	0
B16	1	1	TRUE	0	0	1	FALSE	0	1	0	0	FALSE	0	0
C10	1	1	TRUE	0	0	1	FALSE	0	1	0	0	FALSE	1	0
C11	1	1	TRUE	0	1	1	TRUE	0	0	1	1	TRUE	0	0
C12	0	1	FALSE	1	0	1	FALSE	0	1	1	0	FALSE	0	1
C14	0	1	FALSE	1	0	1	FALSE	0	1	0	0	FALSE	0	0
C17	0	0	FALSE	0	0	0	FALSE	0	0	0	0	FALSE	0	0
C18	0	1	FALSE	0	1	1	TRUE	0	0	0	0	FALSE	0	0
C19	0	1	FALSE	1	0	1	FALSE	0	1	0	0	FALSE	0	0
C20	1	1	TRUE	0	1	1	TRUE	0	0	1	1	TRUE	0	0
C21	1	1	TRUE	0	1	1	TRUE	0	0	1	1	TRUE	0	0
C22	1	1	TRUE	0	1	1	TRUE	0	0	1	1	TRUE	0	0

Treatment 2 Graphical integration guidance data

Following pre-test convention, ‘T2Q3dS’ means “Treatment 2 Question 3d shape right?”. Besides right (0) or wrong (1), subject may intentionally give no response (2).

Subject	T2 Q1a	T2 Q1b	T2 Q1c	T2 Q1d D	T2 Q1d S	T2 Q1d	T2 Q2a	T2 Q2b	T2 Q2c	T2 Q2d D	T2 Q2d S	T2 Q2d	T2 Q3a	T2 Q3b	T2 Q3c	T2 Q3d D	T2 Q3d S	T2 Q3d
B1	1	1	1	1	1	TRUE	1	1	1	1	1	TRUE	1	1	0	0	0	FALSE
B2	1	1	1	1	1	TRUE	1	1	0	1	0	FALSE	1	1	0	1	1	TRUE
B4	0	0	2	0	1	FALSE	0	0	2	0	1	FALSE	0	0	1	0	0	FALSE
B5	1	1	2	1	1	TRUE	1	1	2	1	1	TRUE	0	0	2	0	0	FALSE
B6	1	1	0	1	1	TRUE	1	1	1	1	1	TRUE	1	1	1	1	0	FALSE
B7	0	0	2	0	1	FALSE	1	1	1	1	1	TRUE	1	0	1	0	1	FALSE
B8	0	0	2	0	1	FALSE	1	1	1	1	1	TRUE	1	0	1	1	0	FALSE
B9	0	0	2	0	1	FALSE	0	0	2	0	1	FALSE	1	0	1	0	0	FALSE
B10	1	1	1	1	1	TRUE	1	1	1	1	1	TRUE	1	0	0	0	0	FALSE
B11	1	1	1	1	1	TRUE	1	1	1	1	1	TRUE	1	1	1	0	0	FALSE
B12	1	1	1	1	1	TRUE	1	1	1	1	1	TRUE	1	0	0	0	0	FALSE
B13	1	1	1	1	1	TRUE	1	1	1	1	1	TRUE	1	1	1	1	0	FALSE
B15	2	2	2	0	1	FALSE	1	1	1	1	1	TRUE	1	1	1	1	1	TRUE
B16	1	1	1	0	1	FALSE	1	1	0	1	1	TRUE	1	0	0	1	0	FALSE

Treatment 3 Running total and reflection data

Suffix ‘Cal’ standards for calculation, so ‘T3Q2Cal’ means “Treatment 3 Question 2 calculation right?” and ‘T3Q2i’ means “Treatment 3 Question 2 reflective query i discover it?”. Besides right (0), wrong (1) or no response intentionally (2), subject can skip the queries because his/her answer in corresponding pre-test question was right (3). Value 4 indicates that there is no response probably because of the helper’s problematic checking.

Subject	T3Q1Cal	T3Q1i	T3Q1ii	T3Q2Cal	T3Q2i	T3Q2ii	T3Q3Cal	T3Q3i	T3Q3ii
C2	0	1	2	1	1	2	0	2	2
C4	0	1	0	0	3	3	0	1	1
C5	1	3	3	1	3	3	1	3	3
C6	1	3	3	1	3	3	1	1	1
C8	0	0	0	0	0	0	0	0	0
C9	1	1	1	1	1	1	0	0	1
C10	1	3	3	1	1	1	1	1	1
C11	1	1	1	0	1	1	1	1	1

C12	1	1	1	1	1	1	1	1	1
C14	1	1	2	1	1	2	1	1	2
C17	1	4	4	0	1	0	1	4	4
C18	0	1	1	1	1	1	0	1	1
C19	0	1	1	1	1	1	1	1	0
C20	1	1	0	1	1	0	1	1	1
C21	1	3	3	0	1	1	1	1	1
C22	1	3	3	1	3	3	1	3	3