Table of Contents

Go Back

Can Students Learn Stock-Flow-Thinking?

An Empirical Investigation

Daniel Kainz
Karl-Franzens-University Graz, Austria
Department of Differential Psychology
Universitätsplatz 2
A-8010 Graz Austria / Europe
phone: +43-316-380-8514
daniel.kainz@kfunigraz.ac.at

Guenther Ossimitz
University of Klagenfurt, Austria
Department of Mathematics Education
Universitätsstr. 65
A-9020 Klagenfurt Austria / Europe
Phone: +43-463-2700-3132
Fax: +43-463-2700-3199
guenther.ossimitz@uni-klu.ac.at
http://www.uni-klu.ac.at/~gossimit/home.htm

Preliminary Version January 2002

Submitted for the 2002 System Dynamics Conference, Palermo, Italy

Key words: Stock-Flow-Thinking, Bath Tub Dynamics, Pretest-Posttest design, Crash Course in Stock-Flow-Thinking.

Can Students Learn Stock-Flow-Thinking?

An Empirical Investigation

Daniel Kainz
Guenther Ossimitz

ABSTRACT

Linda Booth Sweeney and John D. Sterman (2000) were able to show in their remarkable study that the ability of MIT-students to discern between stocks and flows in practical situations is rather low. In their paper they coined the term “Bath-Tub-Dynamics”, indicating that things as simple as the inflow and outflow of a bathtub might cause considerable problems for students to correctly determine the amount of water inside a bath tub.

In a sequel study Guenther Ossimitz (2001) showed that the performance of Austrian Business Administration students on “Bath Tub Dynamics tasks” is even worse than in Sweeney and Sterman. The results also gave first indications that some of the observed problems might be due to deficiencies in correctly interpreting flow-related functional graphs.

In the present paper we report about another study investigating students’ abilities in stock-flow-thinking thereby introducing for the first time a pretest-posttest design with a 90-minute crash course introducing basic stock-flow concepts between the two times of testing. We were able to demonstrate that even an intervention as makeshift as our crash course was suitable to bring about an improvement of performanc

1 Preface - Introduction

(by Guenther Ossimitz)

When John D. Sterman presented the results of his "Bath Tub-Dynamics" study (Sweeney and Sterman 2000) in a spectacular plenary lecture at the 2000 conference of the System Dynamics Society in Bergen, the international SD-community was baffled. It seemed incredible that educated (mostly postgraduate) students at MIT, one of the world's leading educational institutions, could exhibit such severe problems in discerning between stocks and flows.

In the fall semester of 2000 I reproduced these findings with a sample of 154 Business Administration students of different Austrian universities (Ossimitz 2001). In that study the original Bath Tub task devised by Sweeney and Sterman was combined with some other tasks designed to assess the extent performance depended upon different versions of cover stories and whether performance would vary according to tabular versus graphical presentation of flow data. The overall results were even more alarming than the results obtained by Sweeney and Sterman. Students displayed severe deficits in grasping the relationships between stocks and flows. Also, it could be shown that when flow data was given in tabular form students' performance was somewhat better than when the information had been presented in a flow chart.

The disastrous findings of my first study concerning this issue had motivated me to conduct a sequel study during spring semester 2001 using a pretest-posttest design with a short treatment introducing basic stock-flow concepts between the two times of testing. The pretest was thereby presented to a group of Business Administration majors attending an introductory course of Applied Statistics in March 2001, followed by a 90-minute crash course one month later and the posttest carried out in June 2001. With the treatment accommodated in the framework of regular class hours and hence not allowing for a strict clinical design providing a control group, the conditions of this investigation were far from optimal. Nevertheless, a significant increase in performance could be observed for most items.

However, this study sought not only to evaluate students'abilities in stock-flow-thinking but was also intended to function as a preliminary test run providing us with valuable insight regarding the usefulness of our tasks and test design. Unlike my first study, the tasks and cover stories used in this study proved to work quite well. Only in a few cases some special problems of interpretation occurred – as in any complex empirical investigation.

A number of people have made this study possible. First of all I would like to thank a series of enthusiastic Business Administration students at the University of Klagenfurt, who did a great job in managing the whole investigation and who were also responsible for preliminary coding and evaluation of data. Sandra Galli, Eva Gratzer, Susanne Kotzent, Benjamin Kreisler, Martin Waiguny and Melanie Zoltan wrote a German project report[1] about the main findings of this study. Part of funding was provided by the Research Commission of the University of Klagenfurt, Austria.

I am very grateful to Daniel Kainz from the Department of Psychology at the University of Graz for his thorough statistical in-depth analysis of the data. Also, the main part of this paper has been contributed by him.

[1] See http://www.uni-klu.ac.at/~gossimit/pap/bf_endber.pdf

2 Method

A pretest-posttest design was chosen in order to address the question whether a one and a half hour introduction to stock-flow-thinking would be suitable to improve students‘ performance on various tasks pertaining to stock-flow-thinking abilities.
To that end, five different tasks had been created to assess students‘ abilities in stock-flow-thinking as well as in reading and interpreting graphs which were to be administered to students at one time before and some time after the crash course.
For every item presented to students in the pretest one corresponding item very similar in structure and tapping the same stock-flow-thinking capabilities clad in by a different cover story was given.

A. The Water Butt Flow / Maier's Bath Tub Flow (WBF/MBTF) Task

The Water Butt Flow / Maier's Bath Tub Flow (WBF/BTF) Task was designed to ascertain whether subjects would be able to ascribe the information given in the reading text as belonging to either the stock or flow and whether they were subsequently capable of using this information to correctly map the in- and outflow onto two blank charts.

The two cover stories are highly similar in content: The Water Butt Flow (WBF) condition describes a water butt with rain water flowing in and excess water slopping over the edge (Figure 1) whereas Maier‘s Bathtub Flow (MBTF) condition describes a bathtub with water flowing in from a faucet and draining out through the outlet (Figure 2). Round numbers were used for inflow and outflow rates in both stories to facilitate plotting.

B. The Tabular Hospital / Tabular Parking Lot (THP/TPL) Task

In this task subjects were shown a table reporting the flows in and out of a stock. The two conditions were intended to test subjects‘ basic understanding of accumulations; in particular, whether participants would be able to correctly compute the net flow and consequently to infer from that information the quantity in a stock at a particular point in time. Again, the two cover stories used are very similar in content. The Tabular Hospital (THP) condition includes a cover story depicting a hospital setting with skiing-accident victims arriving and recovered patients departing over a period of two weeks (Figure 3). By analogy, the Tabular Parking Lot (TPL) condition features a college parking lot with cars arriving and departing during the course of a day (Figure 4). Participants in both conditions were asked to determine the exact moment (day and hour respectively) when the stock reaches its maximum. In order to assess whether subjects had difficulties in switching back and forth in stock vs. flow thinking, they were also asked to specify when the outflow reaches its maximum. In contrast to the TPL condition where this question was left out, participants in the THP condition were also asked to name the day on which the net flow reaches its negative maximum. This was done to test once more whether candidates had difficulties in redirecting their attention from stocks to flows.

C. The Graphic Parking Lot / Graphic Hospital (GPL / GHP) Task

For this task the two cover stories already used in Task B were taken up again and redesigned to create two corresponding conditions in which subjects were presented with a graph depicting the flows in and out of a stock. The identical data points employed in Task B were taken to generate graphs showing the inflow (arrivals) and outflow (departures) over time. Naturally, to avoid any effects brought about by the order of item-presentation the two conditions were balanced such that participants previously having received the Tabular Hospital (THP) condition (Figure 5) were now given the Graphic Parking Lot (GPL) condition (Figure 6) and vice versa. As in Task B, the two conditions were intended to determine whether participants were indeed able to transpose flow information to correctly infer the behavior of a stock over a given time period. This time however, not requiring subjects to compute the net flow from a table but rather, to determine the point in time when the stock reaches its maximum by looking at the two graphs shown in the chart. The two graphs in each condition were thereby constructed in such a way that the inflow was clearly exceeding the outflow for about halfway through the specified time period, and then, upon the net flow crossing zero, the outflow exceeding the inflow for the rest of the given time. Subjects were expected to glean the information necessary for determining the stock maximum by simply looking at the point of intersection of the inflow and outflow curves.

D. The Surge Tank / Bath Tub (ST/BT) Task

Here, the concept of the Bath Tub (BT) Task as originally used by Sweeney & Sterman (2000) has been adopted to assess participants‘ ability in reading the behavior of a stock over time from a flow chart presenting the graphs of flow into and out of a stock. The Bath Tub (BT) condition used in this study being completely identical to the Bath Tub (BT) condition described by Sweeney & Sterman, has been, in a second step, slightly altered by changing the cover story introducing the task. This was done to allow for two corresponding versions to be given to subjects in the pre- and post test. The Surge Tank (ST) condition describes a water surge tank being driven by a permanent pump as well as a second short time high performance pump thus rationalizing the alternating levels of inflow (Figure 7). In the Bath Tub (BT) condition, describing a bathtub with water flowing in and draining out, no account for the varying inflow rates was given (Figure 8). In both conditions students were subsequently asked to draw the time path for the quantity of the stock.

E. Maier’s Bath Tub Stock / Water Butt Stock (MBTS/WBS) Task

This task was derived from the Maier’s Bath Tub Flow / Water Butt Flow (MBTF/WBF) Task described in Point A. This version, like the original, was designed to check people’s ability to classify the information given in a text as distinctly flow vs. stock information. This time however, not asking subjects to use the information to sketch the correct inflow and outflow rates but to transpose the net rate of flow into a stock trajectory. The Maier’s Bath Tub Stock (MBTS) condition states the same cover story as the Maier’s Bath Tub Flow (MBTF) condition although different time points as well as inflow and outflow rates have been used (Figure 9). The Water Butt Stock (WBS) condition on the other hand coincides with the Water Butt Flow (WBF) cover story even in the choice of its datapoints (Figure 10). The wording for both conditions MBTS and WBS have been changed to require subjects to line in the behavior of the stock over time. Additionally, both conditions were balanced in order to avoid contextual confusion. Participants in the pretest were given the Water Butt Flow (MBTF) and the Maier’s Bath Tub

Stock (WBS) conditions whereas the same subjects received the Bath Tub Flow (BTF) and Water Butt Stock (WBS) conditions on the posttest.
Table 1 gives an overview of succession for tasks and conditions presented in the pre- and posttest.

Table 1
Succession of tasks and conditions of the pre- and posttest

	Condition Pretest		**Condition Posttest**	
Task A	Water Butt Flow	WBF	Maier's Bath Tub Flow	MBTF
Task B	Tabular Hospital	THP	Tabular Parking Lot	TPL
Task C	Graphic Parking Lot	GPL	Graphic Hospital	GHP
Task D	Surge Tank	ST	Bath Tub	BT
Task E	Maier's Bath Tub Stock	MBTS	Water Butt Stock	WBS

3 Subjects and Procedure

The tasks above were administered to second semester business students enrolled in a compulsory statistics course titled Applied Statistics for Business Majors during summer semester 2001 at the University of Klagenfurt (Carinthia). 94 students taking part in the pretest received the pretest conditions Task A through Task E presented in Table 1. Students were told that the survey intended to illustrate important statistical concepts they were about to learn during the following few months. Participation was voluntary and neither monetary incentives nor grades were given on the pretest nor the posttest. Four weeks after the pretest a 90-minute introduction to stock-flow concepts was held during regular class hours. During the crash course fundamental properties and common errors in handling stock and flow concepts were discussed and examples given to illustrate the application of stocks and flows in everyday contexts. A translation of the German manuscript used during the crash-course will be provided in the final version.
Another four weeks after the crash course had taken place the five posttest conditions were administered to 64 remaining students who chose to continue participation. Solutions to tasks on the pretest had not been given in order to enable investigation of the effects of treatment on later performance.
To protect student privacy, subjects were free to choose a random ID-code or ID-name which had to be used at both times of testing. Participants were given approximately 20 minutes on the pre- and posttest to complete the test form with extra time provided if necessary.
On the pretest form a background data sheet was included requesting students to fill in age and sex demographics. To explore whether prior knowledge of systems- related concepts might have any effect on performance, people were also asked to rate their graphical reading and comprehension abilities on a 5-point scale. In addition, they were asked to indicate their final math grade on the school leaving exam and to note whether he or she had already heard about stocks and flows somewhere before. If a subject indicated a yes-answer to the latter question, it was also requested to specify the difference between stocks and flows in the blank space provided. Table 2 gives a summary of the background data for the 64 subjects who participated on both times of testing.

Table 2
Subject demographics

All entries are in %		Pretest
Age	18-21	42
	22-25	47
	26 and up	9
Sex	Female	53
	Male	47
Self-assessed reading and comprehension of graphs		
	Very good	2
	Good	27
	Average	56
	Bad	13
	Very Bad	0
Math grade on school-leaving exam		
	A	25
	B	19
	C	17
	D	17
	No Grade	22
Heard about stocks & flows before?		
Yes		6
	No	94
Specify difference between stocks and flows	Correct	100
	Not correct	0
		N = 64

As anticipated, not all 94 students taking part in the pretest could be induced to participate in the posttest two months later. This was in part due to some students dropping out of the course during the semester and some simply refusing to continue testing.
With regard to age our sample was largely comprised of students ranging in age from 22 to 25, followed by the category 18 to 21 and considerably less participants being older than 26. Gender was roughly equidistributed with 53% females and 47% males. Subjects were generally conservative in rating their reading and comprehension of graphs abilities. Most answers stated that reading and comprehension of graphs was only average. About half as many subjects rated their capabilities as being good, followed by a minor number of people indicating very good graphical abilities. No persons though assessed their skills as being very bad. Surprisingly, male subjects assessed their graphical abilities consistently better than females did. 8% of men rated their proficiency as being very good and 37% as being good as compared to only 0% and 12% in women. Similarly, only 6% and 0% of male participants indicated a bad and very bad answer as opposed to 16% and 2% of female participants. The observed difference is very significant ($X^2 = 14.3$, $p < 0.01$).
Concerning the math grades on the school leaving exam roughly one fifth of participants indicated a math grade on each of the respective 5 categories. Incidentally, 20% stated that they had not taken their school leaving exam in math.
Closer analysis revealed an unforeseen cohort effect: Subjects between the age of 18 to 21 had been doing significantly better on their school leaving exams than subjects over 26. 39% vs. 0% indicated an A on their final exam and 24% vs. 18% indicated a B. Similarly, only 6% of subjects in the first age group versus 36% of subjects over 26 had received a C and 9% vs. 36% had obtained a D on the school leaving exam ($X^2 = 18.2$, $p < 0.05$). These findings might be startling but do not however, disclose any information about possible factors generating these results.

To the question whether subjects had already heard about stocks and flows only five students indicated a positive answer. These five subjects were also able to specify correctly the difference between stocks and flows.

4 Results

For reasons of comparability and statistical correctness all subsequent statistical analyses have been conducted with only those 64 subjects who could be prevailed on to participate on both times of testing.

4.1 The Water Butt Flow/Maier's Bath Tub Flow (WBF/MBTF) Task

The WBF/MBTF task was designed to test whether subjects would be capable of processing stock and flow information given in a reading text to correctly sketch the corresponding inflow and outflow graphs. Subjects were given the Water Butt Flow (WBF) condition on the pretest and received Maier's Bath Tub Flow (MBTF) condition for the posttest.

Pretest - The Water Butt Flow(WBF) condition:

Subjects who understood that water was flowing into the water butt at a constant rate and slopping over after the brim had been reached were expected to draw the correct inflow and outflow graphs onto two blank charts (Figure 1). The two graphs were analyzed independently from each other because it was reckoned with the possibility of a test person doing well on one chart and finding it harder to do well on the other.
Also, in order to help identify potential sources of confusion and uncertainty for participants, additional coding criteria were devised to categorize typically erroneous answers.
Table 3 breaks down performance on the inflow chart accordingly.

Table 3
Performance on the WBF condition - the inflow

All entries are %	total
Completely right	3
Right stock	47
Right stock but beyond	27
Severe mistakes on stock	23

N = 64

Generally, performance on this task was poor. Only 3% were able to draw the correct inflow graph. The rest of the sample chose to depict the behavior of the quantity of water in the butt. Almost every second person decided on drawing the graph of the stock and succeeded in doing so correctly. More than a fourth drew the graph of the stock over time but disregarded the fact that at 6 p.m. the constant inflow exceeded the capacity of the butt thus slopping over the edge. Consequently, these subjects wrongly sketched the stock graph even beyond the 100 liter mark. Likewise, nearly another fourth opted for the stock graph but exhibited considerable mistakes. Subjects of this category commonly erred in one or more of the following: Choosing the wrong time of commencement for the inflow, delineating a false stock slope, disregarding the 100 liter mark and/or indicating a false endpoint.
Compared to the inflow chart subjects did somewhat better on the outflow chart. Table 4 again breaks down performance by individual coding criteria. Since there was greater variance in the

types of errors coding criteria differ in type and number from the criteria used on the inflow chart.

Table 4
Performance on the WBF condition - the outflow

All entries are %	total
Completely right	11
Right stock	41
Descending stock	16
Serious mistakes in stock	25
No answer/random	8

N= 64

Here, 11% of subjects correctly understood that they were required to draw the flow curve as opposed to sketching the behavior of the stock. Interestingly, as on the inflow chart, all participants who indeed did opt for the flow chart did also manage to draw the graph without any discernible mistakes.

82% on the other hand did not follow the instruction correctly and delineated the curve for the stock of excess water with varying amounts of errors. 41% accomplished to draw an accurate stock curve for the water spilled compared to almost one fourth who exposed serious mistakes in their stock graphing whereby participants commonly showed one or more of the following errors: Choosing the wrong time of commencement for the inflow, delineating a false stock slope and/or indicating a false endpoint.

Another class of error was exhibited by 16% of subjects who confounded the flow of excess water draining away with a decrease in the stock of water thus falsely sketching a decreasing graph with the outflow rate determining the slope.

8% either gave no response at all or gave a response that was so bizarre in its use of chaotic graphing that it was evident that the person had clearly not understood what was required.

Notwithstanding the fact that slightly differing coding criteria were applied to the inflow and outflow graphs it was still expected that subjects doing well on the inflow graph would do well on the outflow graph and vice versa. To this end, even though coding criteria were designed to be nominal in nature, we assumed an order relation by weighing completely right answers to be highest in ranking and responses with severe mistakes or no answers at all to be lowest in ranking respectively. Contrary to our expectations however, bivariate analysis yielded a rather modest correlation suggesting that people performing well on sketching the inflow graph did not necessarily do equally well in drawing the outflow graph and subjects having great difficulty on the inflow graph do not stringently find it harder to give a correct response on the outflow graph ($r = 0.32$, $p<0.01$). Indeed, 8% of subjects drew the correct outflow graph without having sketched the right inflow graph.

In order to test for any possible differences in difficulty we reduced our data by recoding answers into right vs. false responses and found that subjects performed significantly better on the outflow graph than on the inflow graph (average score of 11% vs. 3%, $t= -2.31$, $p< 0.05$).

Posttest – Maier’s Bath Tub Flow (MBTF) condition:

This condition being very similar to the WBF condition except for different inflow and outflow rates as well as a different time scale required participants to sketch the inflow and outflow onto two blank charts. People were expected to line in the inflow of water starting at 6:05 until 6:11 at a constant rate of 20 liters a minute irrespective of the outlet being open or shut. Analogously, a correct response for the outflow chart consisted of indicating the outflow starting at 6:05 and

ending at 6:07 (when Mr. Maier shuts the outlet) and again starting at 6:20 and ending at 6:26 when Mr. Maier opens the outlet and the water has drained away respectively.
Again, coding criteria have been devised which seek to subsume the given responses in an appropriate manner. Table 5 gives an account of the resulting responses on the inflow chart according to the chosen coding criteria.

Table 5
Performance on the MBTF condition - the inflow

All entries are %	**total**
Completely right	31
Flow/mistakes	6
Right stock	23
Error in stock	20
Separate stock	6
Separate stock/error	3
Completely wrong/missing	9

N=64

In this condition 37% understood correctly that they were expected to draw the inflow as opposed to depicting the stock. 6% committed minor mistakes in either choosing the wrong starting time or wrong end of influx. Nearly one fourth did opt for a correct depiction of the stock compared to 20% of subjects also sketching the graph for the stock but committing one or more errors of the following types: Choosing the wrong time of commencement for the inflow, delineating a false stock slope and/or indicating a false endpoint.
6% of subjects drew a trajectory that presumably intended to resemble the behavior of the stock of water flowing in, irrespective of the water actually inside the bathtub. Another 3% followed the same idea but committed one or more errors of the following: Choosing the wrong time of commencement for the inflow, delineating a false stock slope, indicating a false endpoint and/or assuming the stock to fall off to zero once the inflow has stopped.
9% gave either no answer at all or gave a response that made it clear that they had not understood the task at hand.
Since the spectrum of responses on the outflow chart was identical to the inflow chart identical coding criteria were employed. Table 6 gives a summary of the results observed for the outflow graph.

Table 6
Performance on the MBTF condition - the outflow

All entries are %	**total**
Completely right	34
Flow/mistakes	3
Right stock	21
Error in stock	3
Separate stock	8
Separate stock/error	17
Completely wrong/missing	14

N=64

On the blank chart destined for the outflow graph subjects did equally well as on the inflow graph. 34% correctly lined in the outflow graph whereas 3% chose a wrong end of influx. Almost one fourth chose to depict the stock trajectory and managed to do so without major mistakes compared to 3% who either chose the wrong time of commencement for the outflow, indicated a false endpoint and/or assumed a false stock maximum.
Like on the inflow chart a group of people drew stock trajectories that sought to resemble the behavior of the stock of water flowing out. Considering the possibility that the waste water could again be collected in a second container and thus describing the stock of waste water, 8% did sketch an appropriate stock graph whereas 17% attempted to do so thereby committing one or more errors of the following: Choosing the wrong time of commencement for the outflow, delineating a false stock slope, indicating a false endpoint and/or assuming the stock to fall off to zero once the outflow has stopped.

In order to investigate whether subjects would do equally well on both graphing charts an order relation was assumed in the manner already adopted for analysis in the WBF condition. A completely right answer was classified as highest ranking followed by the order of categories as depicted in tables 5 and 7 and completely wrong/missing ranking lowest. This time, the interrelation between performance on the inflow graph versus performance on the outflow graph was more clear than on the pretest: Bivariate analysis yielded a high positive correlation reaching substantial significance ($r= 0.84$. $p < 0.01$) thus suggesting that subjects who did well on sketching the inflow graph did comparatively well on drawing the outflow graph and subjects having difficulties in sketching the inflow also finding it harder to correctly line in the outflow graph.
Comparison of performance on the inflow and outflow charts yielded no significant result. Subjects were almost identical in their average performance on both charts (average score of 36% and 38%, $t = -1$, $p> 0.05$).

Comparison of pretest and posttest
In order to determine whether the crash course had any positive effect on performance on the WBF/MBTF Task initial results on the inflow and outflow charts were analyzed separately for the pre- and posttest condition. In accordance with our expectations we found that subjects did significantly better both on the inflow chart ($t = -5.20$, $p<0.01$) and the outflow chart ($t = -3.73$, $p< 0.01$) after having received a one and a half hour introduction to stock and flow concepts.

4.2 The Tabular Hospital / Tabular Parking Lot (THP/TPL) Task

The aim of the THP/TPL task was to reveal whether subjects would be capable of computing the net flow from a table reporting the flows in and out of a stock and consequently to indicate the point in time when the quantity of the stock reaches its maximum. To enable retracing of possible sources for confusion and uncertainty people had to justify their answer. Additionally, in order to assess whether participants felt secure in handling stock and flow information they were asked to specify the time when the outflow exhibits its maximum. In the THP condition not in the TPL condition though it was also asked to name the point in time when the net flow reaches its negative maximum.

Pretest- The Tabular Hospital (THP) condition

Figure 3 shows the condition as it was presented on the test form including the correct responses for questions I, II and III.

I. On what day did the largest number of patients occupy the hospital?

There are two possible strategies identifiable that lead to a correct solution to question I. One can either compute the net flow for each day and arrive at the solution by summing up the net flows thus obtaining the day with the maximum number of patients accommodated. Or, one can omit summing up the net flows and derive the correct answer by simply discerning that until January 8th every day more patients were arriving at the hospital than recovered patients departing the clinic and the opposite holding true for the rest of the week.
Again, appropriate coding criteria were developed to disclose frequent sources of error. Table 7 gives a summary of the responses given to question I broken down according to practical coding criteria.
Table 7
Performance on the THP condition – question I

All entries are %	total
Correct answer	56
+/- 1 day	16
Day of largest influx	9
No, because...	19

N=64

In this condition 56% indicated correctly January 8th as their answer. Of this group 39% justified their response by stating that until this day the number of new arrivals exceeded the number of departures or gave a response that used different wording but was correct in the underlying meaning. 58% on the other hand stated that they were arriving at the right response by summarizing the net flows of each day. 3% gave no justification as to how they deduced their answer.

Another 16% gave a response one day adjacent the right date. We assumed that this group did use the correct strategy but committed errors in computing and consequently obtained a slightly different result. This is emphasized by the fact that 10% of subjects in this group did state that until the specified day the number of arrivals had exceeded the number of departures. Similarly, 50% justified their response by stating that they had simply added and subtracted the net flows of each day. 40% on the other hand, disposed a rule of thumb strategy: They did not closely consider the in- and outflow but just reckoned by following the course of arrivals and departures that the right answer must lie somewhere around January 8th.

9% of subjects confounded inflow and net flow and subsequently chose January 1st as their answer. Of this group 83% indicated that they had indeed interpreted the question as asking for the day of largest influx. 17% on the contrary indicated no answer at all.

19% of subjects on this task indicated that this question was not answerable. 67% in this category argued that no answer can be given because the quantity of the initial stock had not been specified. The rest did not indicate why they were not able to give a response.

II. On what day did the largest number of patients depart the hospital?
Question II seemingly caused less confusion than question I. 95% correctly indicated January 13th as the day when the outflow reaches its maximum. 5% chose January 31st as their answer even though inflow and outflow rates were not specified beyond January 31st. These subjects presumably theorized that patients were to be dismissed by the end of the month.

III. On what day did the largest number of beds become vacant?

In order to find the correct response to this question subjects were required to find the day on which the net flow reaches its negative maximum.
Table 8 gives an account of given responses according to appropriate criteria.

Table 8
Performance on the THP condition – question III

All entries are %	total
Correct answer	39
Same question as on II.	38
31.12.	5
Not answerable	19

N=64

39% were able to give the correct response. About the same percentage did not however discern any difference between question II and question III. Interestingly, all subjects who found question III to be the same as question II had given a correct answer to question II.
5% specified 31st of December to be the day with the largest number of beds becoming vacant. We hypothesized that this group might have understood the question differently and was led into thinking, since no initial stock had been specified, that on the first day of opening there naturally must have been the largest number of beds available.
Those 19% of subjects who indicated that question III was not answerable indicated either no reason at all or gave a vast array of reasons that mainly targeted the following three statements: The number of outpatients had not been specified, the total stock of available beds would have to be known and some patients might have died in the course of treatment.

Posttest- The Tabular Parking Lot (TPL) condition

Figure 4 shows the condition as it was presented on the test form including the correct responses for question I and question II.

I. At what hour did the largest number of cars occupy the parking lot?

Analogous to the THP condition subjects were presented a table reporting the inflow and outflow of cars of a parking lot in the course of a day. The present condition is highly similar to the THP condition although a different time scale as well as different data points have been used. Again, participants were free to employ two similar strategies to deduce the correct solution. They were expected to either compute and sum up the net rate of arriving and departing cars or inversely, to detect the correct solution by ascertaining that until 12:00 the inflow of cars has exceeded the outflow and the reverse pattern to hold true for the rest of that afternoon.
As subjects gave the same array of responses as on the pretest condition, the same coding criteria have been used. Table 9 breaks down performance according to coding criteria.

Table 9
Performance on the TPL condition – question I

All entries are %	total
Correct answer	80
+/- 1 hour	5
hour of largest influx	11
No, because...	5

N=64

More than two third (80%) of our sample was able to correctly identify the time between 11 to 12 as the hour with most cars occupying a parking space on the university parking lot. 55% of these subjects indicating the right answer did also state that they arrived at their conclusion by noticing that until that specified hour more cars were entering the parking lot than cars were leaving. Similarly, 20% indicated that they were computing the right time by summing up the net flows of every hour.
All remaining persons of this group (25%) did not specify how they arrived at the correct result.

We assumed that even more subjects were following the right strategy since another 5% gave a response adjacent to the target answer. 67% of these participants missing the solution by an inch correctly appreciated the contingency of the inflow exceeding the outflow until 12:00. On the other hand, 33% did not specify how they arrived at their conclusion.

Comparable to the pretest, some subjects (11%) mistook stock for flow and indicated the hour between 8 to 9 with the largest influx as their answer. Unexpectedly, 14% of subjects in this category nevertheless indicated that they arrived at their conclusion by beholding that until the specified time the inflow of cars was greater than the outflow of cars. Another 14% stated that they computed their solution by summing up the net flows. This finding is somewhat startling since these subjects obviously were familiar with the correct strategy for solution but did not however follow through with that method. Another 29% affirmed our assumption that they were indeed confounding the stock with the inflow. 43% did not justify their evidently wrong answer.

5% did indicate that the time with the largest number of cars occupying the parking lot was not detectable since the initial quantity of the stock of cars had not been specified.

II. At what hour did the largest number of cars depart the parking lot?

Like on the pretest this question apparently made it it easy for subjects to find the right answer. The results are identical with the pretest condition with 95% indicating the hour between 4 to 5 p.m. as the time with most cars departing the lot and 5% giving no or entirely false answers.

Comparison of pretest and posttest

A t-test for repeated measures with a 99% confidence interval revealed that subjects performed significantly better on the second time of testing after having been introduced to important concepts of stock and flow thinking. Average performance on the Tabular Hospital condition was only 56% as opposed to 80% of subjects giving the target answer on the posttest condition ($t = -3.56$, $p < 0.01$). Adopting a more tolerant manner in the choice of coding criteria thereby allowing responses adjacent to the target response to pass as a correct response yielded a similar significant result. Average performance on the pretest was likewise significantly lower than on the second time of testing (84% vs. 72%, $t = -2.05$, $p < 0.05$).

4.3 The Graphic Parking Lot / Graphic Hospital (GPL / GHP) Task

As on the previous task this exercise was devised to assess whether participants were capable of transposing flow information to correctly infer the behavior of a stock over a given period of time. To test whether participants would show performance comparable to that on the THP/TPL task when given a graphical depiction of the in- and outflows instead of a tabular presentation, the two cover stories already used in the THP/TPL task had been redesigned to accommodate a chart with the inflow and outflow graphs lined in. To this end the same data points and time scales as in the THP/TPL task had been employed. The two resulting graphical conditions were constructed in such a way that for about halfway through the specified time period the inflow was clearly exceeding the outflow and upon the net flow crossing zero the opposite pattern holding true. Subjects were asked to specify the point in time when the quantity of the stock reaches its maximum and to explain how they arrived at their conclusion. Additionally, in order to probe whether subjects would find it difficult to switch between stock and flow concepts while processing this task they were also asked to specify the time when the outflow exhibits its maximum. In the GHP condition unlike the GPL condition it was also asked to state the time when the net flow reaches its negative maximum.

Pretest-The Graphic Parking Lot (GPL) condition

Figure 5 shows the condition as it was presented on the test form including the correct responses for questions I and II.

I. At what hour did the largest number of cars occupy the parking lot?

Subjects were expected to find the right solution by looking at the intersection of the inflow and outflow graphs. From 5:00 until 12:00 the number of cars entering the parking lot was higher than the number of cars departing. Therefore, the number of cars occupying a parking space was steadily increasing until it reached a maximum at 12:00. In the afternoon the reverse pattern was observed. Every hour more cars were departing than arriving. Since the graphing was relatively coarse and data points were not clearly identifiable participants were not expected to derive their answer by adding and subtracting the net flows.
Table 10 breaks down performance according to appropriate coding criteria.

Table 10
Performance on the GPL condition – question I

All entries are %	**total**
Correct answer	14
+/- 1 hour	5
hour of largest inflow	63
hour of largest outflow	3
No, because...	16
	N=64

Only 14% of subjects identified 12:00 correctly as the hour with most cars present at the parking lot. All persons of this group further justified their response by indicating that they gleaned the information by looking at the intersection of the inflow and outflow graph.

5% chose a solution adjacent to the target answer whereby one third indicated having derived the answer by looking at the intersection, another third by computing and summing up the net flows and the rest indicated no explanation. We suspect that this group might have been given a chance to perform better if a more detailed graph chart had been provided.

The largest group in table 10 is comprised of subjects mistaking the net flow for the stock. 63% falsely chose the hour between 8:00 to 9:00 with the largest positive net flow as their answer. Three fourth consequently also affirmed their error by stating that they had indeed chosen the hour with the maximum positive difference between inflow and outflow. The rest of this group indicated no justification.

Oddly, a minor group of 3% similarly indicated that they derived their response by choosing the largest positive net flow, even though these subjects did in fact state the hour between 16:00 to 17:00 with the net flow reaching its negative maximum.

16% on the other hand indicated the question as not answerable with half of these subjects criticizing that the quantity of the initial stock had not been given and the rest giving no further explanation.

II. At what hour did the the largest number of cars depart the parking lot?

No differences in the frequency of right answers could be observed compared to the THP/TPL condition. 95% gave a correct answer opposed to 5% giving a completely wrong answer or no answer at all.

Posttest- The Graphic Hospital (GHP) condition

Figure 6 presents the Graphic Hospital condition as it was presented on the test form including the correct responses for questions I,II and III.

I. On what day did the largest number of patients occupy the hospital?

The GHP condition is almost identical with the THP condition in its use of cover story, datapoints and time-scale except for a graphical presentation of inflows and outflows being employed instead of a tabular presentation. Subjects were again expected to derive the day with the maximum quantity of the stock by regarding the point of intersection of the inflow and outflow graph.
Table 11 gives an overview of the various types of answers given to question I.

Table 11
Performance on the GHP condition – question I

All entries are %	total
Correct answer	67
Largest inflow	25
No, because...	8
	N=64

Average performance on this task was 67%. Of those subjects giving a right response 93% were also able to indicate the right strategy. The remaining representatives of this category surprisingly stated that they had computed the solution by summarizing the net flows of each day. This finding might appease our concerns that the graphical quality of our graphs was too blurry and imprecise for closer reflection.

Exactly one fourth confounded the maximum quantity of the stock with the largest positive net flow and indicated January 1st as their answer. 81% of these subjects directly affirmed this assumption and the rest indicated no justification.

8% of participants indicated that question I was not answerable whereby 40% impeached that the initial quantity of the stock had not been specified and the remaining subjects giving no explanation.

II. On what day did the largest number of patients depart the hospital?

Here, even slightly more subjects gave the correct answer than on the pretest (95% vs. 97%) The rest either gave no response at all or gave a response that appeared so random that these subjects were conjectured to have guessed their answer.

III. On what day did the largest number of beds become vacant?

Subjects were again required to find the day with the largest positive difference between the inflow and outflow curve. Table 12 breaks down performance according to coding criteria deemed most appropriate.

Table 12
Performance on the GHP condition – question III

All entries are %	total
Correct answer	47
Same question as on II.	31
Wrong answer/random	3
No answer	19
	N=64

All 47% of subjects giving a right answer to question III had already given the target answer to question II. 31% on the other hand could not discern any difference between question II and question III whereby 10% of this category had already chosen a wrong answer to question II and consequently responded in the same manner to question III.
3% indicated a response that could only be classified as randomly wrong.
19% indicated no answer at all without further specifying their difficulties. We surmise that a large proportion of these subjects thought question III to be identical to question II since all subjects of this category had well given the correct response to question II.

Comparison of pretest and posttest

Closer analysis of data revealed a highly significant difference in performance for question I on the GPL/GHP task between the two times of testing. Subjects did significantly better on the posttest after having been introduced to stock-flow thinking than on the pretest where no such measures had been provided (average score of 67% vs. 19%, $t = -7.69$, $p < 0.01$)
In contrast, no differences in performance could be detected between both times of testing for question II ($t = 0.44$, $p > 0.05$).

We further attempted to probe for possible differences in performance between the tabular versus the graphic presentation of inflow and outflow rates. Contrary to our expectations comparison between the Tabular Hospital and the Graphic Parking Lot conditions of the pretest yielded a highly significant result. Subjects did significantly better in determining the maximum of a stock from a table reporting the flow information than in gleaning the information from a graph chart portraying the inflow and outflow graphs (56% vs. 14%, $t = 6.05$, $p < 0.01$). For the posttest the same difference could be observed. Even though subjects had received an extra lesson on stocks and flows they nevertheless showed significant better performance on the Tabular Parking Lot condition than in the Graphic Hospital condition (80% vs. 67%, $t = 2.20$, $p < 0.05$).

4.4 The Surge Tank / Bath Tub (ST/BT) Task

This task was adopted with permission from the Sweeney and Sterman study (2000). As in the original, subjects were asked to describe the behavior of a stock over time by using the inflow and outflow curves provided in a flow chart. According to Sweeney and Sterman this task was designed to target subjects' graphical integration ability, a skill essential in understanding the dynamics of complex systems. In order to derive two corresponding conditions for the present study we used the original version of Sweeney and Sterman's Bath Tub task for the posttest (BT), and designed another version for the pretest. This was done solely by changing the cover story to describe a surge tank being driven by one permanent and one high performance water pump (ST) to account for varying inflow rates. Except for these minor changes in the ST cover story both conditions were kept identical.
The ST and BT conditions as they were presented on the test forms including their target answers can be gathered from Figure 7 and Figure 8.

The responses given by our sample were coded according to the same criteria already employed by Sweeney and Sterman. Readers wishing a more detailed description of the criteria in use are kindly referred to the original. Average scores for each of the 7 criteria as well as corresponding results of t-testing are summarized for both measures in Table 13.

Table 13
Performance on the ST and BT conditions

Criterion	Pretest Ave	Posttest Ave	t	p
1 When the inflow exceeds the outflow, the stock is rising.	0.38	0.59	- 2.90	< 0.01
2 When the outflow exceeds the inflow, the stock is falling.	0.36	0.59	- 3.21	< 0.01
3 The peaks and troughs of the stock occur when the net flow crosses zero (i.e. at t = 4,8,12,16)	0.39	0.59	- 2,86	< 0.01
4 The stock should not show any discontinuous jumps.	0.75	0.78	- 0.44	> 0.05
5 During each segment the net flow is constant so the stock must be rising (falling) linearly.	0.34	0.63	- 4.10	< 0.01
6 The slope of the stock during each segment is the net rate (i.e. $\pm$ 25 units/time period)	0.16	0.31	- 2.61	< 0.05
7 The quantity added to (removed from) the stock during each segment is 100 units, so the stock peaks at 200 units and falls to a minimum of 100 units.	0.16	0.31	- 2.61	< 0.05
Mean for all items	**0.36**	**0.54**	**- 3.21**	**< 0.01**

Except for criterion 4 which yielded no discernible difference subjects performed significantly better on all other criteria after the crash course had taken place. The results clearly speak for a beneficial effect of the stock and flow intervention.

Additionally, in order to conduct bivariate analysis we pooled the responses given on both times of testing for every criterion. We found that criterion 1 and criterion 2 as well as criterion 6 and 7 exhibit a high degree of interdependency ($r = 0.96$, $p < 0.01$ for pair 1 and 2; $r = 1$, $p< 0.01$ for pair 6 and 7). Also, for criteria 1 and 2 we performed cross-tabulation and found that subjects who correctly showed them in their drawing had a higher chance of consequently fulfilling the other criteria. Inversely, hardly any person who violated these two fundamental aspects had a chance on succeeding in other criteria. These findings suggest that this pair of criteria could essentially be considered as a single variable. For all other intercorrelations however, results were not as clear.

Since the Surge Tank condition of the pretest was, except for changes in cover story, directly derived from the original Bathtub Task devised by Sweeney and Sterman this task lends itself for comparison regarding average performance of the MIT students investigated in Sweeney and Sterman (2000) and the sample comprised of mostly business administration students in Ossimitz (2001). Table 14 presents average performance on the seven criteria for all three research studies. The results show that subjects in this study performed considerably poorer than the MIT students investigated by Sweeney and Sterman and surprisingly also less successfully than the students tested by Ossimitz.

Table 14
Comparison of performance on the BT task with results in Ossimitz (2001) and Sweeney and Sterman (2000)

Criterion	Pretest Ave	Ossimitz	Sweeney &Sterman
1 When the inflow exceeds the outflow, the stock is rising.	0.38	0.42	0.87
2 When the outflow exceeds the inflow, the stock is falling.	0.36	0.43	0.86
3 The peaks and troughs of the stock occur when the net flow crosses zero.	0.39	0.56	0.89
4 The stock should not show any discontinuous jumps.	0.75	0.64	0.96
5 During each segment the net flow is constant so the stock must be rising (falling) linearly.	0.34	0.38	0.84
6 The slope of the stock during each segment is the net rate	0.16	0.26	0.73
7 The quantity added to (removed from) the stock during each segment is the area enclosed by the net rate.	0.16	0.27	0.68
Mean for all items	**0.36**	**0.42**	**0.83**
	N = 64	N = 154	

4.5 The Maier's Bath Tub Stock / Water Butt Stock (MBTS/WBS) Task

These two variations of the original Maier's Bath Tub Flow / Water Butt Flow (MBTF/WBF) conditions were likewise designed to check subjects' understanding of stock and flow relationships. In contrast to the original where participants had to sketch the inflow and outflow graphs, subjects were now asked to transpose the net rate of flow into a stock trajectory. The Maier's Bath Tub Stock (MBTS) condition employs the same cover story already used for the MBTF condition although different time points as well as inflow and outflow rates have been used. For the Water Butt Stock (WBS) condition no such changes have been made, except of course, for slight changes in wording to require participants to draw the stock trajectory.

As the reader will soon notice, the two conditions might be problematic to compare in a pretest-posttest design because the resulting stock trajectory in the WBS condition differs from the MBTS stock trajectory to the effect that it does not exhibit a negative slope and consequently might be easier for subjects to complete. Further, no outflow had to be subtracted from the inflow in the WBS condition in order to compute the net flow since the inflow simply halted after reaching the maximum. We are aware of that problem but still hope this task to be useful in demonstrating that, when given stock and flow information in a reading text, subjects had considerably more difficulties in sketching the inflow and outflow graphs than in plotting the stock trajectory.

Pretest- The Maier's Bath Tub Stock (MBTS) condition

The correct answer to the MBTS condition is depicted in Figure 9.

Subjects were expected to gather from the instruction that the net flow undergoes change during four different segments: During segment 1 ($0 < t \leq 4$) the net inflow is 5 liters/minute (l/min). After the outlet is shut the net inflow increases to 14 l/min in segment 2 ($4 < t \leq 9$). In segment 3 ($9 < t \leq 15$) when Mr. Maier turns off the faucet the net flow drops to 0 l/min and falls to – 9 l/min when he reopens the outlet in segment 4 ($15 < t \leq 25$). In order to deduce the behavior of the stock subjects had to calculate the total added to the stock by the end of every segment, given by the area bounded by the net rate curve. Since the net flow remains constant during every segment the stock also moves forward steadily and subjects simply had to draw straight lines connecting the stock at the beginning and the end of every segment. Average performance for this condition was surprisingly good. Table 15 summarizes average scores for each of the seven coding criteria adopted from Sweeney and Sterman.

Participants in this condition did best showing that the stock is dropping when the outflow is greater than the inflow. A fraction less was able to show correctly that the stock is rising when the inflow exceeds the outflow. As on the ST/BT task people had most difficulties in showing that the net rate is the slope of the stock and that the area enclosed by the net rate in any interval is the quantity added to the stock during the interval.

In order to determine any possible differences in performance between plotting a stock trajectory versus plotting the flow graphs we somewhat arbitrarily assigned the category [correct response] to subjects who correctly incorporated all 7 criteria in their stock graphing in the MBTS condition. We then contrasted the results of performance on the MBTS condition results of the Water Butt Flow (MBF) condition also given on the pretest. Interestingly, participants performed significantly better in drawing the correct stock trajectory than in plotting the correct inflow graph (59% vs. 3%, $t = -8.48$, $p < 0.01$). A similar result was found when we compared the stock graphing with the outflow graphing (59% vs. 11%, $t = -6.88$, $p < 0.01$).

Figure 10 shows the WBS condition as it was given on the posttest including the correct response.

Table 15
Performance on the MBTS and the WBS conditions

Criterion	Pretest Ave	Posttest Ave	t	p
1 When the inflow exceeds the outflow, the stock is rising.	0.83	1.00	- 3.62	< .01
2 When the outflow exceeds the inflow, the stock is falling.	0.89	----		
3 The peaks and troughs of the stock occur when the net flow crosses zero	0.81	0.87	- 0.94	> .05
4 The stock should not show any discontinuous jumps.	0.86	1.00	- 3.22	< .01
5 During each segment the net flow is constant so the stock must be rising (falling) linearly or remain constant.	0.81	0.94	- 2.20	<.05
6 The slope of the stock during each segment is the net rate	0.61	0.86	- 3.38	<.01
7 The quantity added to (removed from) the stock during each segment is the area enclosed by the net rate.	0.59	0.86	- 3.56	<.01
Mean for all items	**0.77**	**0.92**	**-3.39**	**<.01**

Posttest-The Water Butt Stock (WBS) condition

Here, the challenge for participants lay in comprehending that the stock of water in the water butt remained constant after the capacity of 100 liters had been reached. During segment 1 (14:00 < t $\leq$ 18:00) the net inflow is 25 l/hr and the stock rises at a constant rate with the slope of 25 units per time period. After full capacity has been reached the net inflow drops to 0 l/hr as the rain water simply slops over and the stock subsequently remains constant in segment 2 (18:00 < t $\leq$ 00:00). Average scores for six of the seven criteria devised by Sweeney and Sterman are depicted in table 15.

Closer analysis of the WBS condition and the MBTF condition yielded similar results as for the pretest. Despite the extra lesson on stocks and flows participants still had considerably fewer difficulties in determining the stock trajectory than in delineating the inflow graph (86% vs. 37%, t = - 7.10, p< 0.01). Analogously, they did also perform substantially better on the stock graphing task than in drawing the outflow graph (86% vs. 37%, t = - 6.88, p< 0.01).

Taking into account the weaknesses of comparability of the pretest and posttest conditions, the results of t-testing are nevertheless summarized in table 15. Results exhibit the tendency that the crash course might have had some beneficial effects on posttest performance.

Impact of Subject Demographics: We also attempted to explore whether variables such as age, sex, self-assessed reading and comprehension of graphs ability as well as math grade on the school leaving exam might be able to differentiate between persons performing well and not so well on our five tasks. (Since only four persons stated on the pretest sheet that they had already heard about stocks and flows before we did not consider this item for further analysis).We particularly hypothesized that subjects receiving a better grade on their school leaving exam would outperform those who were only average. Similarly we deemed it legitimate to expect participants rating their reading and comprehension of graphs ability to be good to perform better than persons indicating their abilities as bad or only average. In order to test this assumption we conducted several analyses of variance relating performance on the pre- and posttest conditions to the four variables mentioned above. Unexpectedly, neither age and math grade nor self-assessed reading and comprehension of graphs ability were able to account for participants' performance on the pre- and posttest. Merely gender appeared to play a role for some of the items presented on the test forms. In the GPL condition men were consistently better in determining the point in time when the stock reaches its maximum by looking at a flow chart depicting the flows into and out of a stock ($t = -2.84$, $p < 0.01$). Men similarly outperformed their counterparts in the ST condition where participants not only had to glean the stock information from the inflow and outflow graphs but subsequently had to sketch the time path for the quantity of the stock ($t = -3.05$, $p < 0.01$). Surprisingly, this gender difference persisted even after participants had received the one and a half hour intervention. Men continued to do better on the GHP condition ($t = -3.41$, $p < 0.01$) as well as on the BT condition ($t = -3.61$, $p < 0.01$). The MBTS condition which required subjects to transpose the flow information presented in the reading text into a stock trajectory exhibited a similar gender difference ($t = -2.61$, $p < 0.05$). This time however, the effect of gender did not carry over to the WBS condition of the posttest ($t = -0.22$, $p > 0.05$).

5 Discussion

The present study sought to pursue two main goals. First, we attempted to assess whether difficulties in solving tasks pertaining to stock-flow-thinking were due to deficiencies within the ability to discern between stocks and flows or conversely, due to poor reading and interpreting of graphs. Second, we pursued to evaluate whether a short treatment introducing the properties and fundamental differences of stocks and flows would proof suitable to bring about an improvement in performance in a pretest-posttest design.
We are aware at this point that our design exhibits a number of shortcomings which should be taken into account when interpreting the initial results of our study. Particularly the absence of an adequate control group allows for legitimate scepticism that the observed changes in performance might in part be attributable to a mere training effect. Also, in order to determine whether each of the five pretest-posttest pairs of conditions were indeed analogous in their level of difficulty it would have been necessary to cross-balance all 10 conditions such that half of the subjects would be presented with the pretest condition and the other half be presented with the putative posttest conditions first. Despite these insufficiencies to be addressed in future investigations and the preliminary character of our study we nevertheless feel to have detected a number of interesting findings.

Regarding our first goal we found that subjects found it considerably easier to transpose the flow information given in a reading text into a stock trajectory than sketching the corresponding inflow and outflow graphs. 59% of subjects managed to draw the correct stock trajectory in the MBTS condition compared to only 3% and 11% depicting the correct inflow and outflow curves in condition WBF. On the other hand 47% decided to sketch the behavior of the amount of water flowing into the water butt in the WBF condition and did so correctly by drawing a stock trajectory with the net rate determining the slope.
Analogously, 41% drew a stock trajectory for the amount of water spilling over the water butt thereby equally using the net flow for the slope of the trajectory. This fact supports the idea that subjects had not per se difficulties in differentiating stock and flow information but rather, found it hard to represent flow information by means of a flow chart. They were well able to discern that water was flowing into the butt at a rate of 25 l/hr and that the water would accumulate to 100 l by 6 p.m. thereby reaching full capacity of the container. However, when asked to graph the inflow and outflow onto two blank charts they apparently felt that linear curves with zero slope would not be appropriate to resemble water flowing in and out of a butt. There was also

some slight indication that subjects had more difficulties in sketching the correct inflow graph than in sketching the outflow graph. 8% of subjects managed to draw the correct outflow graph after having sketched a stock trajectory in the chart destined for the inflow graph. The only explanation we can offer at this point is that possibly, subjects found it more plausible that water flowing out of a container should be presented in a flow chart since the water indeed is not contained in a vessel. There might be alternative explanations however for the observed reservation to draw flow graphs which should be addressed in future research. Especially interviews following testing could provide a good method for clarifying the underlying motives beneath this tendency.

Another striking result shows itself in the observed differences in performance between tabular versus graphic presentation of flow information. In the THP condition a total of 56% were able to correctly infer from a table presenting arrivals and departures of patients the day with the maximum number of patients accommodated in the hospital. In comparison, when given the inflow and outflow chart depicting similar flow information in the GPL condition only 14% were able to find the target answer. One might object that some individuals in the THP condition adopted the strategy of computing and summing up the net flow of each day to arrive at the right solution. Considering the possibility that a person might have added up net flows without having understood at some level the relationship between inflow and outflow rates and the corresponding behavior of the stock, the difference in performance remains nevertheless poignant. 22% who arrived at their answer by strictly making out that until the 9th of January every day the number of arrivals exceeded the number of departures, can be contrasted to 14% who succeeded in adopting a similar strategy when looking at the flow chart in the GPL condition.
With regard to question II in the THP and the GPL conditions asking to specify the point in time when the outflow reaches its maximum, average performance was 95% on both tasks. Despite the extreme simplicity of this question we are all the same surprised that no difference in performance was detectable. We conjecture that different modes of flow presentation do not induce differences in performance until a certain threshold of difficulty has been overstepped. Question III in the THP condition was similarly intended to assess whether subjects would be able to switch back into thinking in terms of flow information. Subjects were expected to indicate the day exhibiting the negative maximum in net flow. We valued this item to be somewhat more difficult to answer than question II since participants had to realize that the stock of available beds would increase when departures exceed arrivals rather than simply looking at the outflow rates. Indeed average performance was 39% and thus considerably lower than performance on question II. However, the observed difference is put into perspective when we acknowledge that 38% thought this question to target the same answer as question II and 19% giving no response at all. We surmise that the difficulty in differentiating between question II and question III might for the most part be due to a semantic inadequacy of question III. For want of alternative wording we chose question III to ask for the day on which the largest number of beds became vacant but having previously required participants to process information concerning the stock of patients this subtle switch in stock terminology might have been confusing to many. We are not quite sure how to tackle this issue at this point since rewording the question to address the stock of patients instead of stock of vacant beds would render question III essentially more complex and confusing.

Even though the Surge Tank condition differs from the original Bath Tub task devised by Sweeney and Sterman in the usage of its cover story which accounts for varying rates of inflow and thus potentially forfeits some of its comparability, we nevertheless seized the opportunity for comparison and a conservative assessment of preliminary results. Table 14 illustrates that participants in the present study performed far poorer on all seven criteria than the highly educated sample investigated by Sweeney and Sterman. One might argue that naturally, since the

MIT sample was mainly comprised of students enrolled in various graduate programs and having received advanced training in calculus, these participants should perform better than first and second semester business administration students. Despite the notion that calculus is also an integral part of all Austrian mathematics curricula during the last two years of secondary education we feel obliged to point out that Sweeney and Sterman repeatedly stressed the fact that the BT task can be answered without use of any mathematics beyond simple arithmetic. The finding that students nevertheless performed that poor on this task emphasizes once more the stringent necessity to take deliberate measures to overcome these alarming deficits in the ability to understand the basic concepts of accumulation.
Students of the present study not only performed poorer than MIT students but seemingly also a trifle less successfully than the students investigated by Ossimitz 2001. The difference is noticeable but small and may very likely reflect an artefact brought about by the comparably modest size of our sample which counted less than half as many students than investigated by Ossimitz (N = 64 vs. N = 154).

The suggestion of a gender effect was first elicited by Sweeney and Sterman who were able to ascertain that males repeatedly outperformed females on all of the three tasks comprising their systems thinking inventory. The effect was only marginally significant but consistent. Ossimitz similarly observed men performing significantly better than females on five of his six tasks designed to target stock-flow-thinking and the ability of reading and interpreting graphs. In the present study we were able to show that males performed significantly better than their female counterparts in the GPL, the ST as well as the MBTS conditions presented on the pretest. Considering that women did not differ from men with respect to the math grade obtained on the school leaving exam this finding appears even more puzzling. We are not sure how to account for this unexpected gender effect and wish to abstain from unqualified speculation. However, it seems pressing to us that the issue of gender interacting with performance be addressed in future systems thinking research.

With regard to our second goal we were able to show that an intervention as short as 90 minutes in length following pretesting and introducing only the basic properties and fundamental differences between stocks and flows, was suitable to significantly improve students‘ performance on subsequent tasks targeting stock-flow-thinking and reading and interpreting graphs abilities. We found a significant improvement in performance for all our five tasks (except for question II and III of the THP/TPL task and question II and III of the GPL/GHP task respectively). Certainly, these findings are preliminary and further investigation and optimization of intervention measures is needed, but the results seem to allow for the optimistic view that interventions designed to develop systems thinking abilities might well be able to overcome fundamental deficiencies in these capabilities.

References

Ossimitz, G.(2001). Stock-Flow-Thinking and Reading stock-flow-related Graphs: An Empirical Investigation in Dynamic Thinking Abilities. Paper submitted for the 2002 SDS-Conference, Palermo.

Sweeney, L. B. & Sterman, J. D. (2000). Bathtub Dynamics: Initial Results of a Systems Thinking Inventory. System Dynamics Review, 16 (4), 249 –286.

Figure 1: The Water Butt Flow (WBF) condition

A water butt with a capacity of 100 liters is empty at 12:00. At exactly 14:00 rain sets in and the water from the gutter flows into the butt at a rate of 25 l/hr until midnight. The butt has no outlet but fills up until the water slops over. The slopping water is regarded as outflow. Draw the inflow and outflow between 12:00 and 00:00 in the charts below!

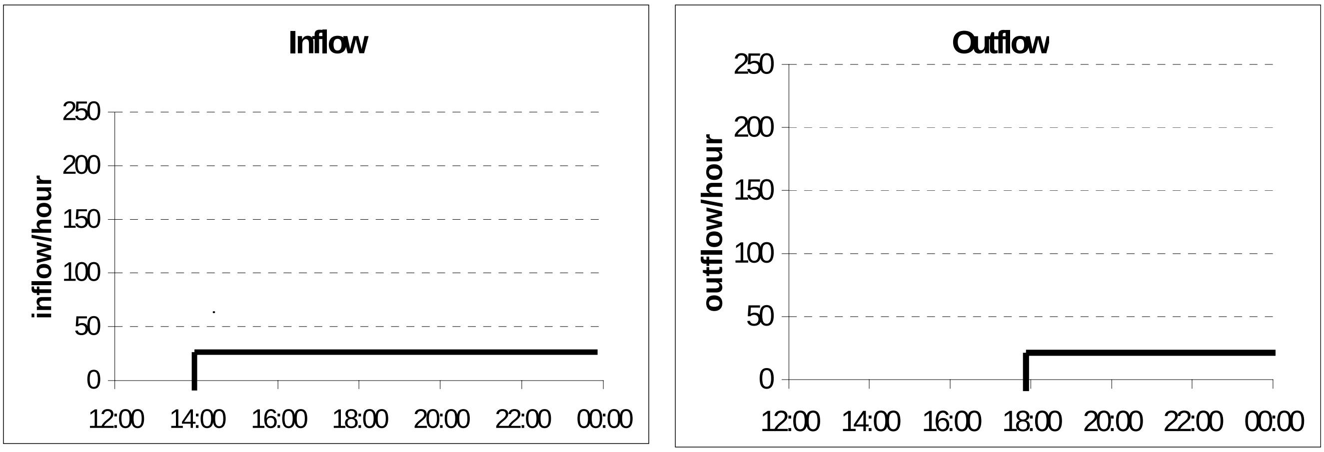
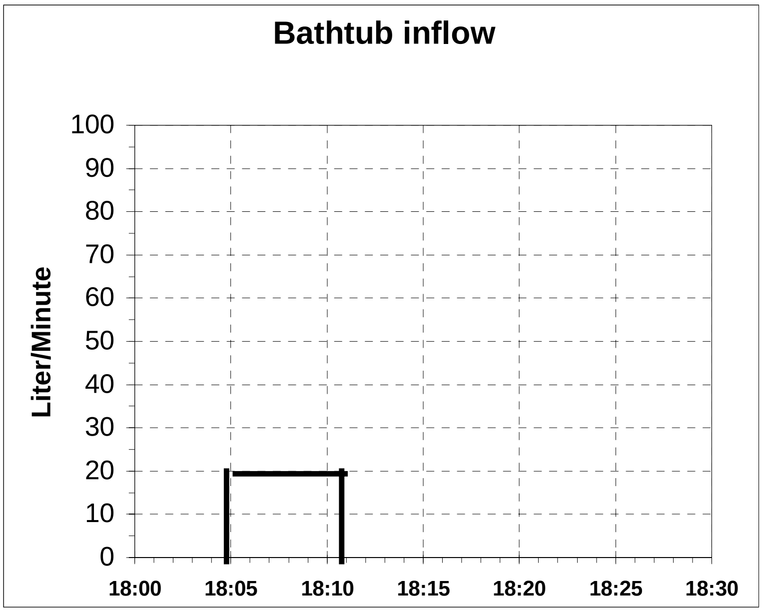


Figure 2: The Maier's Bath Tub Flow (MBTF) condition

At exactly 18:05 Mr. Maier turns on the faucet and starts filling his bath tub. Water flows in at a constant rate of 20 liters/min. At exactly 18:07 Mr. Maier notices that the outlet had not been shut, so he closes the outlet, but leaves the faucet turned on. Through the outlet water flows out at a constant rate of 15 liters/min. At exactly 18:11 Mr. Maier turns off the faucet and enjoys his bath until 18:20. At exactly 18:20 he opens the outlet and lets the whole 90 liters of waters drain away.

Sketch the behavior of the inflow and outflow during the specified time period in the charts below !



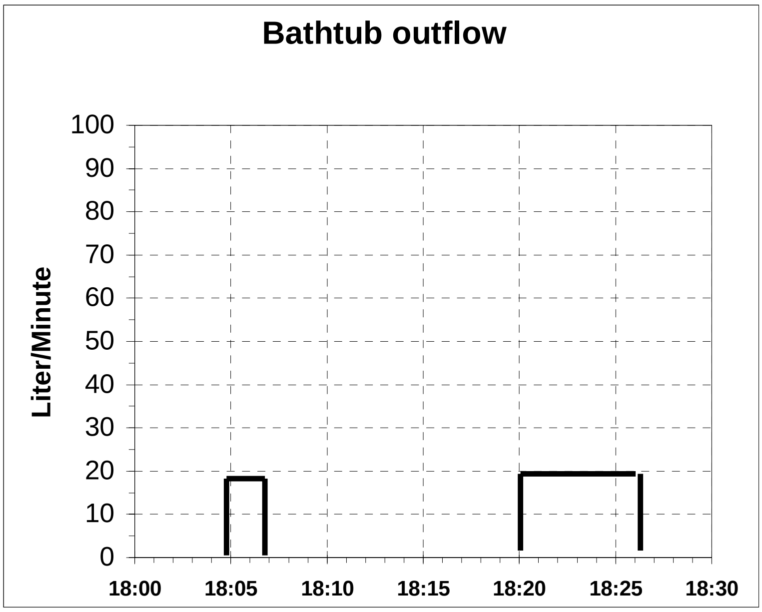


Figure 3: The Tabular Hospital (THP) condition

During skiing season many injured persons are being brought to the hospital Schwarzach-St.Veit from surrounding areas. On peak skiing days most new patients arrive at the hospital. Departures usually take place during the week. Patients receive in patient-treatment for an average of 5 to 12 days. The following table reports the number of arriving and departing patients for every day during a period of two weeks

	Date	Arrivals	Departures	Date	Arrivals	Departures
So	31.12.2000	13	0	07.01.2001	19	3
Mo	01.01.2001	18	0	08.01.2001	11	9
Di	02.01.2001	10	4	09.01.2001	12	15
Mi	03.01.2001	7	5	10.01.2001	8	17
Do	04.01.2001	10	7	11.01.2001	10	25
Fr	05.01.2001	10	8	12.01.2001	8	28
Sa	06.01.2001	12	0	13.01.2001	14	30

All following questions refer to the time period depicted in the table!

I. a) On what day did the largest number of patients occupy the hospital?

Date: _January 8th___ **O question is not answerable because...**

____.

b) How did you arrive at your answer?

Until January 8th arrivals exceed departures and therefore the number of patients at the end of every day is larger than in the morning of that day

From the January 9th onwards departures exceed arrivals and therefore the number of patients at the end of each day is smaller than at the beginning of that day .

II. On what day did the largest number of patients depart the hospital?

Date: _January 13th ___ **O question is not answerable because....**

____.

III. On what day did the largest number of beds become vacant?

Date: _January 12th _ **O question is not answerable because....**

____.

Figure 4: The Tabular Parking Lot (TPL) condition

The table below reports the amount of cars arriving and departing a university parking lot during the course of a day.

Hour	Arrivals	Departures	Hour	Arrivals	Departures
5 -6	0	0	14-15	24	48
6 -7	11	1	15-16	15	59
7-8	44	2	16-17	10	98
8 -9	92	11	17-18	16	35
9-10	53	17	18-19	9	22
10-11	42	12	19-20	4	8
11-12	37	25	20-21	3	5
12 -13	24	31	21-22	1	2
13-14	47	55	22-23	0	0

All following questions refer to the time period depicted in the table!

I. **Can you tell by looking at the table at what hour the largest number of cars did occupy the parking lot?**

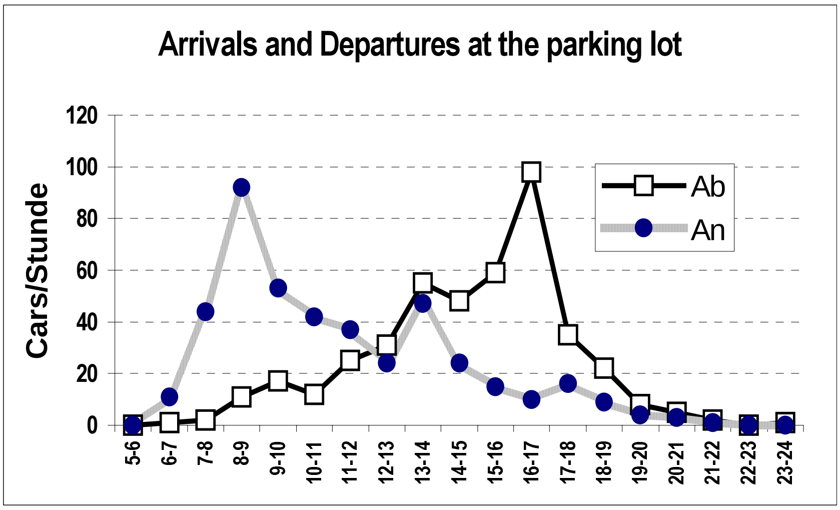
0 Yes, at _12:00_**,** **by using the following method:**
Until 12:00 arrivals exceed departures, but after 12 departures exceed arrivals, therefore most of the cars are at 12 o´clock at the parking lot.

0 No, because…

II. **At what hour did the largest number of cars depart the parking lot?**
Between 16:00 and 17:00**.**

Figure 5: The Graphic Parking Lot (GPL) condition

The following chart reports how many cars arrived at and departed from a college parking lot during the course of a day (before 5 a.m. the parking lot is empty):



I. **Can you tell by looking at the chart at what hour the largest number of cars did occupy the parking lot?**

O Yes, at 12:00 **by using following method:** Before the two graphs intersect at 12:00 the number of cars arriving exceeds the number of cars departing. For the rest of the time the opposite is true.

.

O No, because ...

.

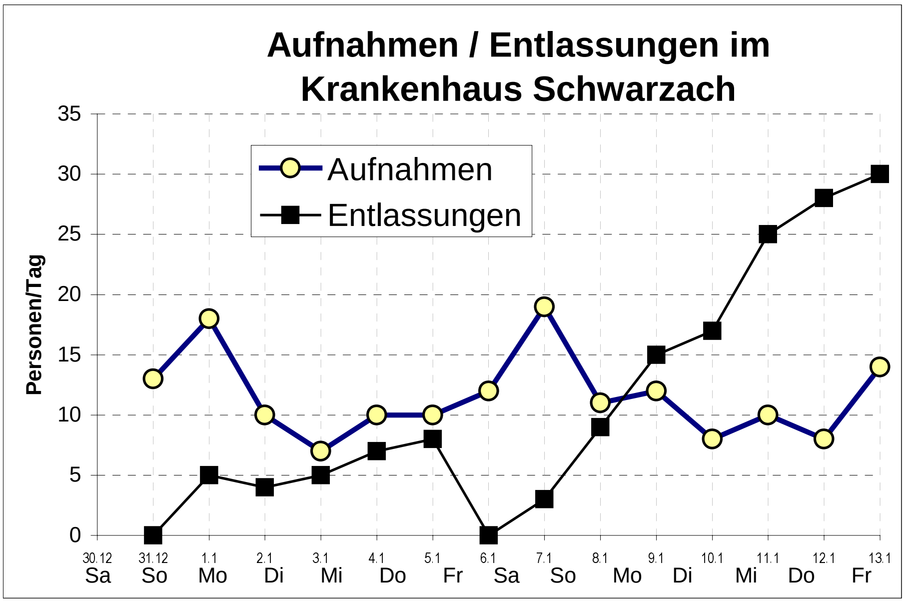
II. **At what hour did the largest number of cars depart the parking lot?**
Between 16-17h.

.

Figure 6: The Graphic Hospital Task (GHP) condition

During skiing season many injured persons are being brought to the hospital Schwarzach-St.Veit from surrounding areas. On peak skiing days most new patients arrive at the hospital. Departures usually take place during the week. Patients receive in patient-treatment for an average of 5 to12 days. The following chart reports for each day the number of arriving and departing patients during the course of two weeks

All following questions refer to the time period depicted in the table!



I. a) On what day did the largest number of patients occupy the hospital?

Date: _January 8th___ **O question is not answerable because...**

____.

b) How did you arrive at your answer?

Until January 8th arrivals exceed departures and therefore the number of patients at the end of every day is larger than in the morning of that day. From January 9th onwards departures exceed arrivals and therefore the number of patients at the end of each day is smaller than at the beginning of that day.

II. On what day did the largest number of patients depart the hospital?

Date: _January 13th__ **O question is not answerable because...**

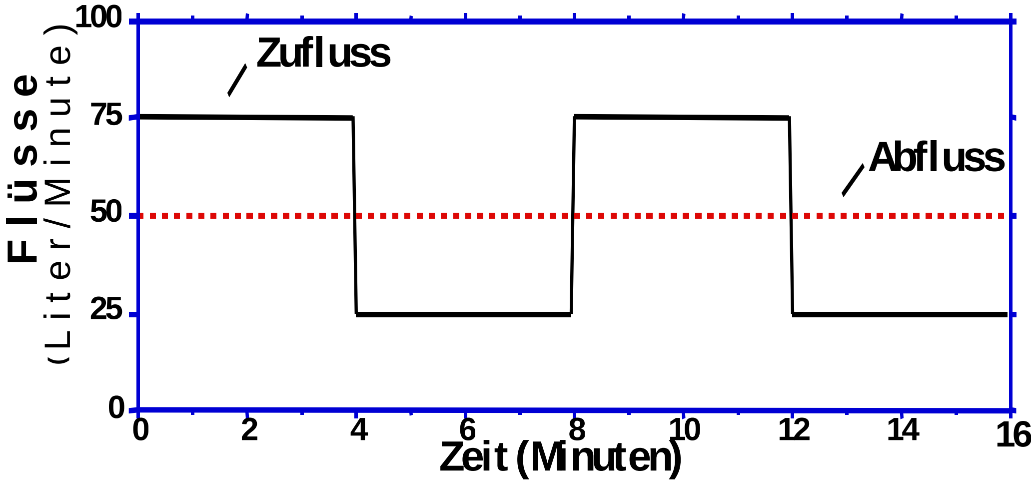
____

III. On what day did the largest number of beds become vacant?

Date: _January 12th___ **O question is not answerable because...**

Figure 7: The Surge Tank (ST)

A water surge tank shall evenly balance 50 liters of water per minute. The tank is driven by two pumps. One permanent pump delivers evenly a certain amount of water per minute. The second inflow pump is a short time high performance pump which evenly delivers water for four minutes and is then turned off for the next four minutes. This is done in order for the well that is tapped by the high performance pump to regenerate. The graphic below depicts the outflow (dotted line) and the total inflow (continuous line) of the surge tank.



Assuming the inflow and outflow conditions from above sketch in the chart below how the quantity of water in the surge tank behaves during the specified 16 minutes. Assume the initial quantity in the container to be 100 liters.

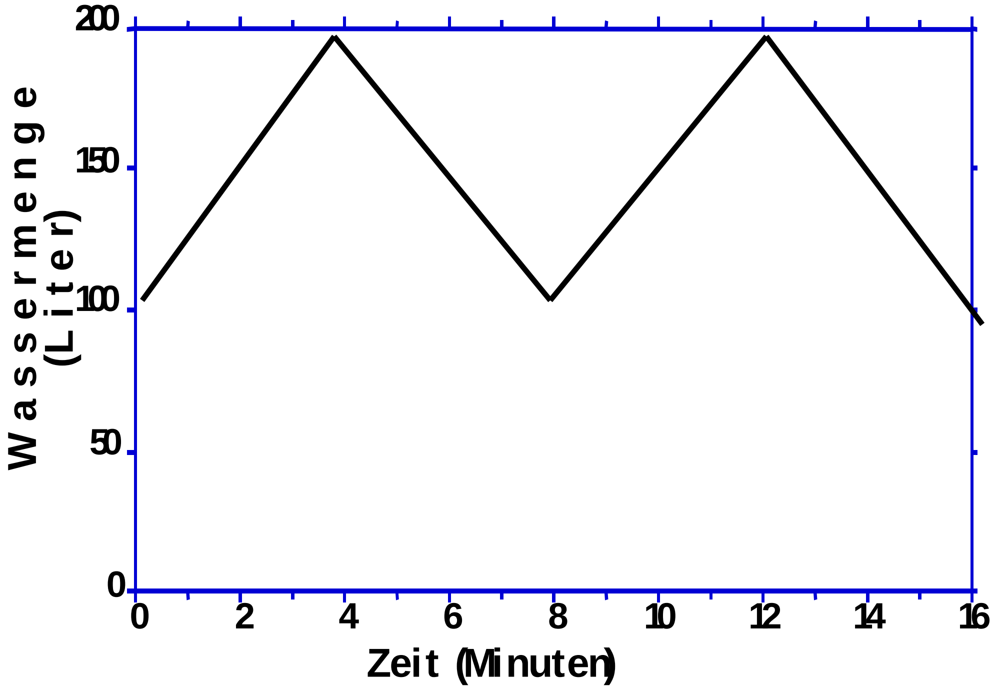
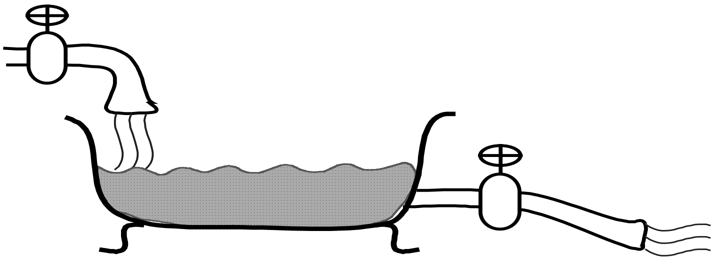
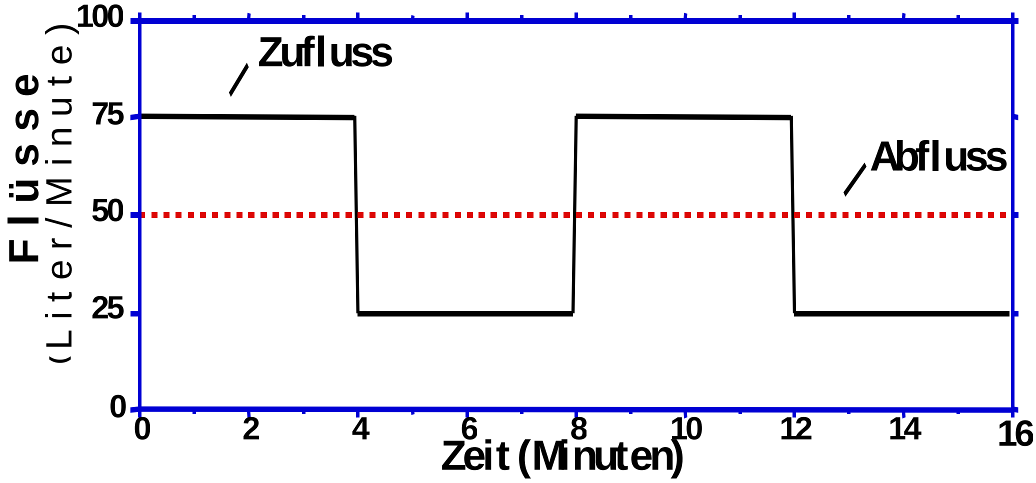


Figure 8: The Bath Tub Task (BT) condition

Consider the bath tub shown below. Water flows in at a certain rate, and exits through the drain at another rate:



The graph below shows the hypothetical behavior of the inflow and outflow rates for the bath Tub. From that information, draw the behavior of the quantity of water in the tub on the second Graph below. Assume the initial quantity in the tub (at time zero) is 100 liters.



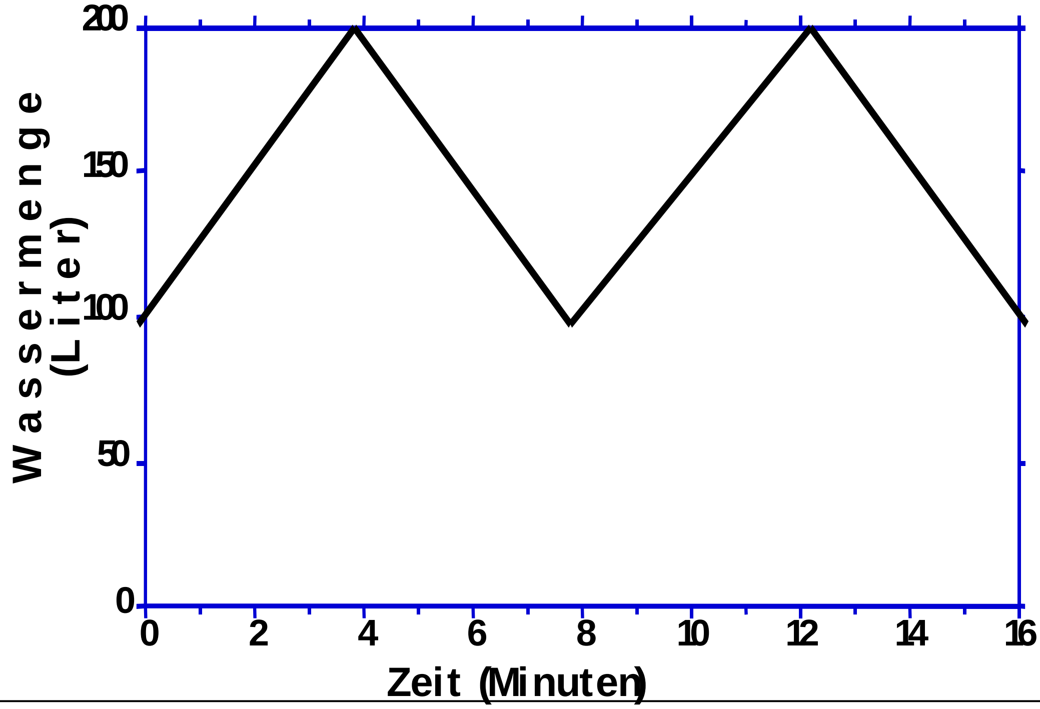
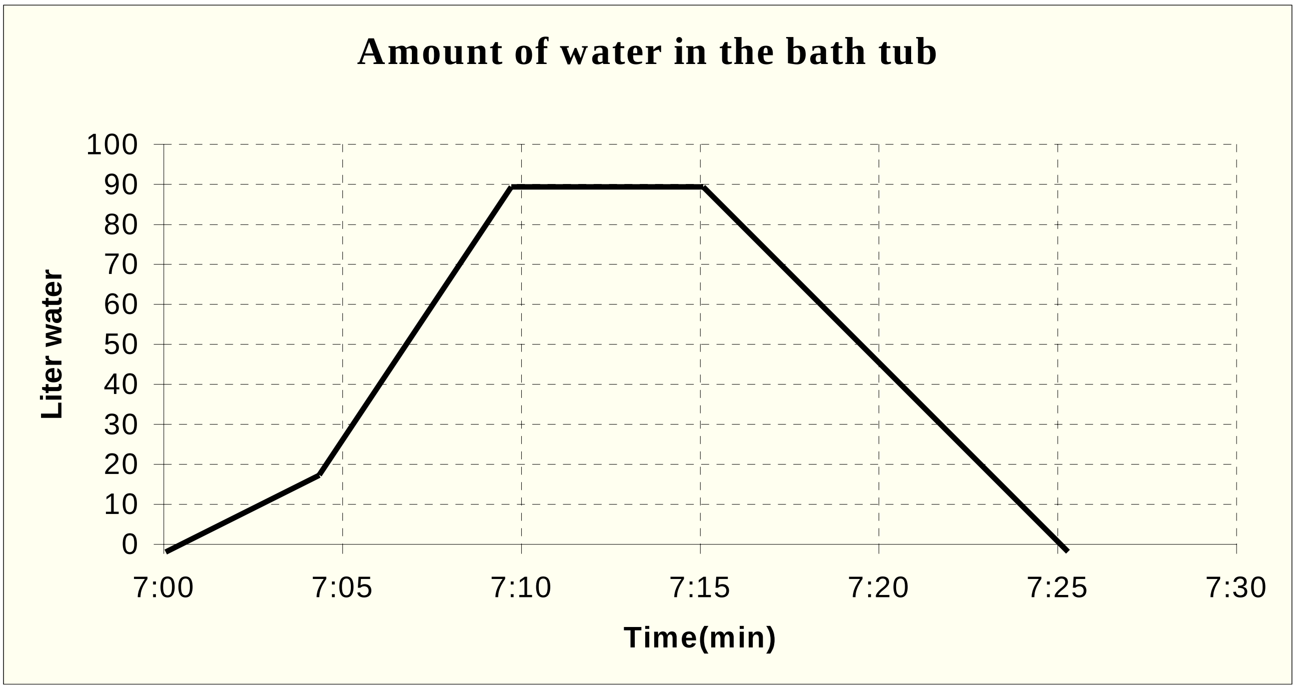


Figure 9: Maier´s Bath Tub Stock (MBTS) condition

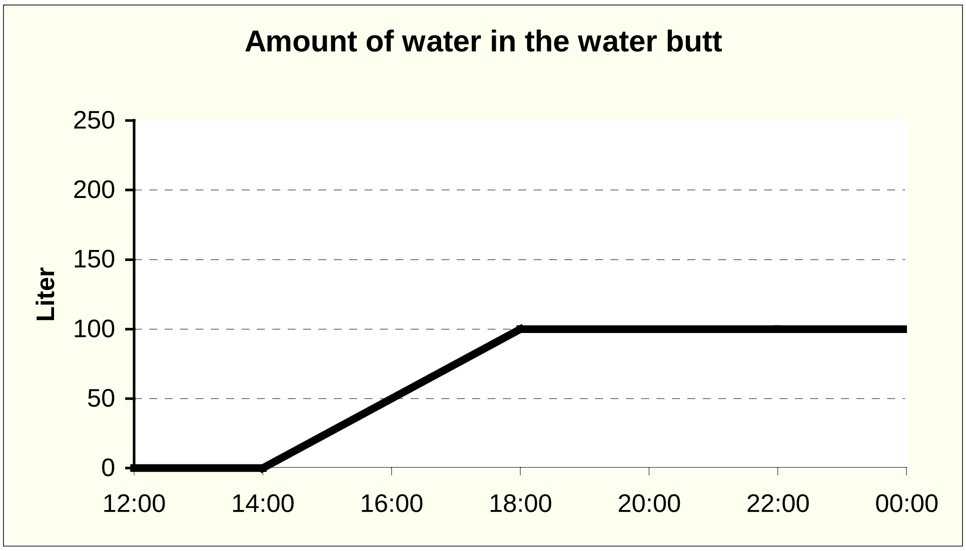
At exactly 19:00 Mr. Maier turns on the faucet and starts filling his bath tub. Water flows in at a constant rate of 14 liters/min. At exactly 19:04 Mr. Maier notices that the outlet had not been shut, so he closes the outlet but leaves the faucet turned on. Through the outlet water flows out at a constant rate of 9 liters/min.At exactly 19:09 Mr. Maier turns off the faucet and enjoys his bath until 19:15. At exactly 19:15 he opens the outlet and lets the whole water drain away.



Draw the behavior of the quantity of water in the tub above!

Figure 10: **The Maier´s Bath Tub Flow (MBTF) condition**

A water butt with a capacity of 100 liters is empty at 12:00. At exactly 14:00 rain sets in and the water from the gutter flows into the butt at a rate of 25 l/hr until midnight. The butt has no outlet but fills up until the water slops over. The slopping water is regarded as outflow. Draw the behavior of the quantity of water in the water butt between 12:00 and 00:00 in the chart below!



Back to the Top