ICAM - International Conference on Advances in Management
July 20-23, Washington D.C.

Strategy Lab Experiences: Bathtub Dynamics

Carlos Capelo
PhD Student
ISCTE
Ald. Vale S.Gião, 50, 2665-381 Milharado, Portugal
Phone: +351962831740
carlos.capelo@galpenergia.com

João Ferreira Dias
Associate Professor
ADETTI/ISCTE
Av das Forças Armadas, 1649-026 Lisboa, Portugal
Phone: +351938450825
fdias@iscte.pt

1. Introduction

As the world changes ever faster, managers increasingly recognize the complexity and turbulence of the business systems in which they are embedded. The management problems are dynamic, while the dynamic complexity comes frequently from few variables with circle and delays interrelations that introduce nonlinearities.

Previous behavioural research in dynamic decision-making has found that decision makers are very susceptible to the dynamic complexity of the decision context (for example, Sterman, 1989; Paich and Sterman, 1993; Diehl and Sterman, 1995). One explanation for the poor performance found in these studies is based on bounded rationality of human decision-making (Sterman, 2000). Due to limitations of cognitive capabilities, first - the mental models people use to make their decisions are deficient – second – even people form adequate mental models, are unable to correctly infer the dynamic behaviour of the system.

To understand complex systems, people must firstly understand some basic systems thinking concepts such as feedback, stocks and flows, time delays and nonlinearities (Sterman, 2002). These building blocks are essential for effective systems thinking and the first step is to understand the misunderstandings.

Following previous similar studies (for example: Sweeney and Sterman, 2000; Sterman, 2002; Ossimitz, 2002; Kainz and Ossimitz, 2002; Lyneis and Lyneis, 2003; Zarara, 2003;.Fisher, 2003; Kapmeier, 2004; Armenia et al., 2004), our work aims to experimentally confirm the managers' misunderstanding of some basic systems thinking concepts.

The present paper describes a research conducted in Portugal with two different groups - one, academic; the other, professional - where we explored the subjects' understanding of some basic systems thinking concepts such as stock-flow relationship, feedback processes and time delays.

2. Method

In this section we describe the method we used to conduct the research. We explored the participants' understanding of some basic systems thinking concepts such as stock-flow relationship, feedback processes and time delays. With this purpose, four simple tasks were applied to two different groups: at ISCTE (a business graduate school in Lisbon) and at Galp Energia, one of the biggest Portuguese firms (the Portuguese oil company).

In order to compare them with similar studies previously done, the research was based on the same four dynamics tasks. Tasks 1, 2 and 3 were based on "Bathtub", "Cash Flow" and "Manufacturing" (Sweeney and Sterman, 2000). Task 4 was based on "Department Store" (Sterman, 2002).

These tasks were designed to be simple and can be answered by people with no special mathematic skills as they do not require subjects to apply any analytic tool of mathematic calculus but simple arithmetic operations and the ability, known as graphical integration, to read and interpret the graphics.

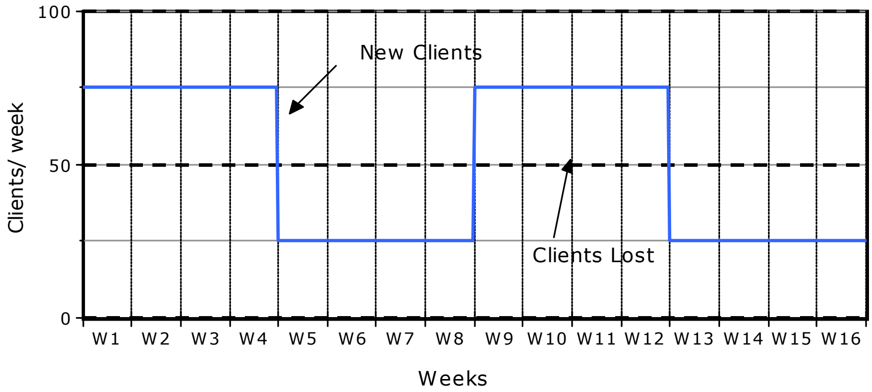
The structure and the main content of the original tasks were maintained. The only difference was that we decided to replace the stock of Task 1. We used "number of clients" rather than "bathtub" and the time is measured in weeks rather than minutes. The tasks were translated into Portuguese.

We compare the results with those of other universities, Tor Vergata University in Rome (Armenia et al., 2004); Stuttgart Institute of Management and Technology (Kapmeier, 2004) and MIT (Sweeney and Sterman, 2000; Sterman, 2002).

Consider a client base of a company. Clients are won in a certain rate, and are lost at another rate.

The graph below shows the hypothetical behavior of the inflow and outflow rates for the client base. From that information, draw the behavior of the client base on the second graph below.

Assume that the initial number of clients is 100.



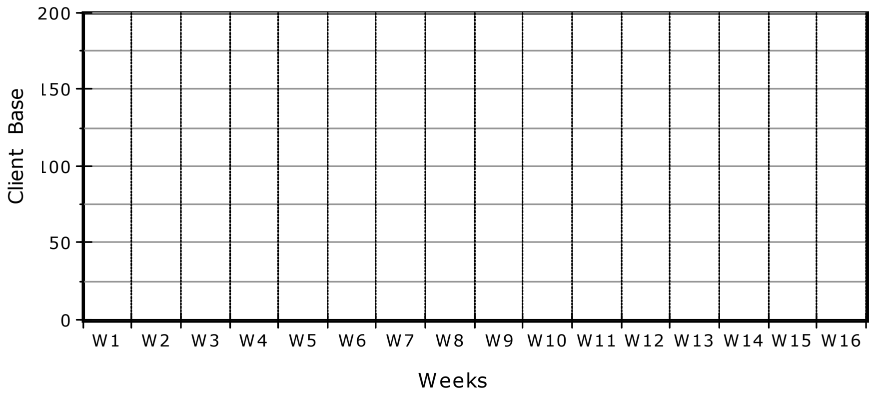


Figure 1 – Task 1 based on “Bathtub Task” (Sweeney and Sterman, 2000)

In the first two tasks (figure 1 and figure 3), we explore the understanding of stock-flow relationship. Stocks and flows are fundamental to the dynamics of systems (Forrester, 1961). In these tasks a stock, already filled with an initial amount, is filled through an inflow and emptied through an outflow at the same time. The subjects are asked to draw the expected behaviour of stock over time.

There are no feedback processes, no time delays, no nonlinearities involved in these tasks. There is only one stock. The outflow is constant, and the inflows follow simple patterns. The inflow patterns differ: square wave pattern for Task 1 and sawtooth pattern for Task 2. Therefore both tasks are simple examples of stock-and-flow thinking. Round numbers are used, so it is easy to calculate the net flow and quantity added to (removed from) the stock.

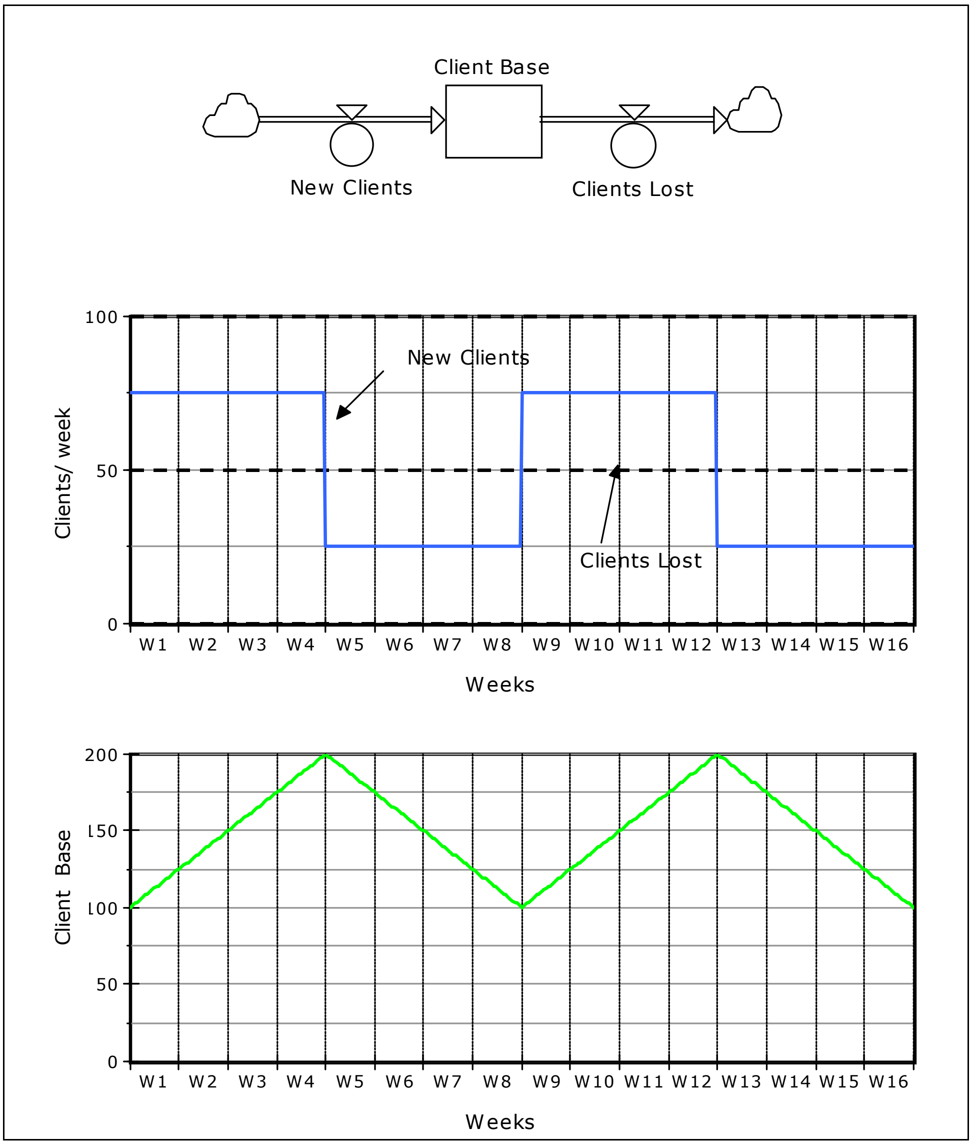


Figure 2 – Solution for Task 1

Task 1 (figure 1), based on "Bathtub Task" (Sweeney and Sterman, 2000), describes a company with an initial number of clients. New clients are being won at a certain rate and, at the same time, existing clients are being lost at another rate. Participants are asked to indicate the number of clients over time. Figure 2 shows the correct answer to task 1. Both inflow and outflow are constant during each segment. The answer involves a simple graphical integration task; as the net flow is constant, the stock increases or decreases linearly.

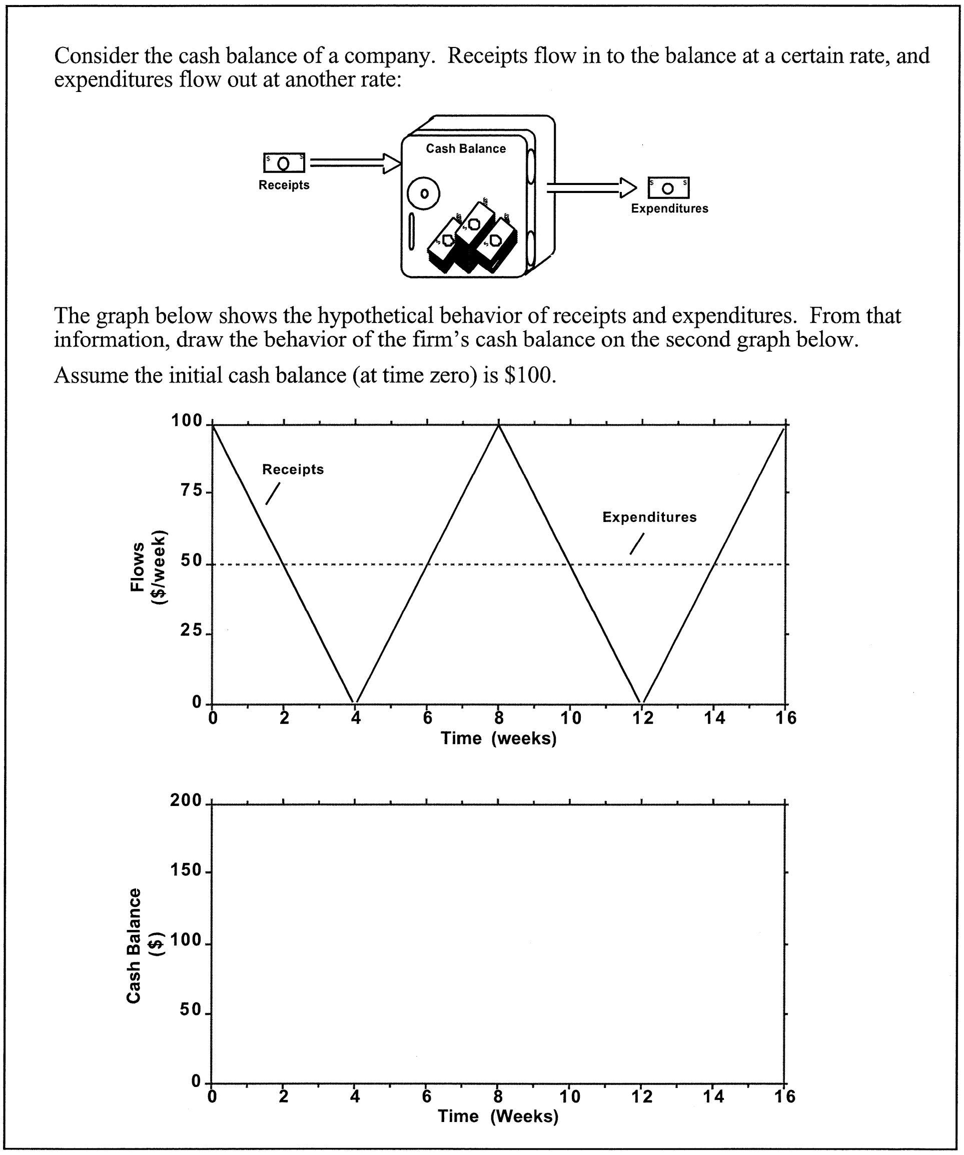


Figure 3 – Task 2

Source: Sweeney and Sterman, 2000, p 254

Subject's performance on task 1 was evaluated by using the following criterions (Sweeney and Sterman, 2000):

C1- When the inflow exceeds the outflow, the stock is rising

C2- When the outflow exceeds the inflow, the stock is falling

C3- The peaks and troughs of the stock occur when the net flow crosses zero

C4- The stock should not show any discontinuous jumps

C5- During each segment the net flow is constant so the stock must be rising (falling) linearly

C6- The slope of the stock during each segment is the net rate

C7- The quantity added to (removed from) the stock during each segment of four periods is the area enclosed by the net rate (25 units/time x 4 time periods = 100 units, so the stock peaks at 200 units and falls to a minimum of 100 units).

In Task 2 (figure 3), based on “Cash Flow Task” (Sweeney and Sterman, 2000), the outflow is also constant while the inflow follows a sawtooth wave, rising and falling linearly. The participants are asked to draw the cash balance of a company, given the initial balance, rate of receipts (inflow) and rate of expenditures (outflow). Figure 4 shows the correct answer to this task.

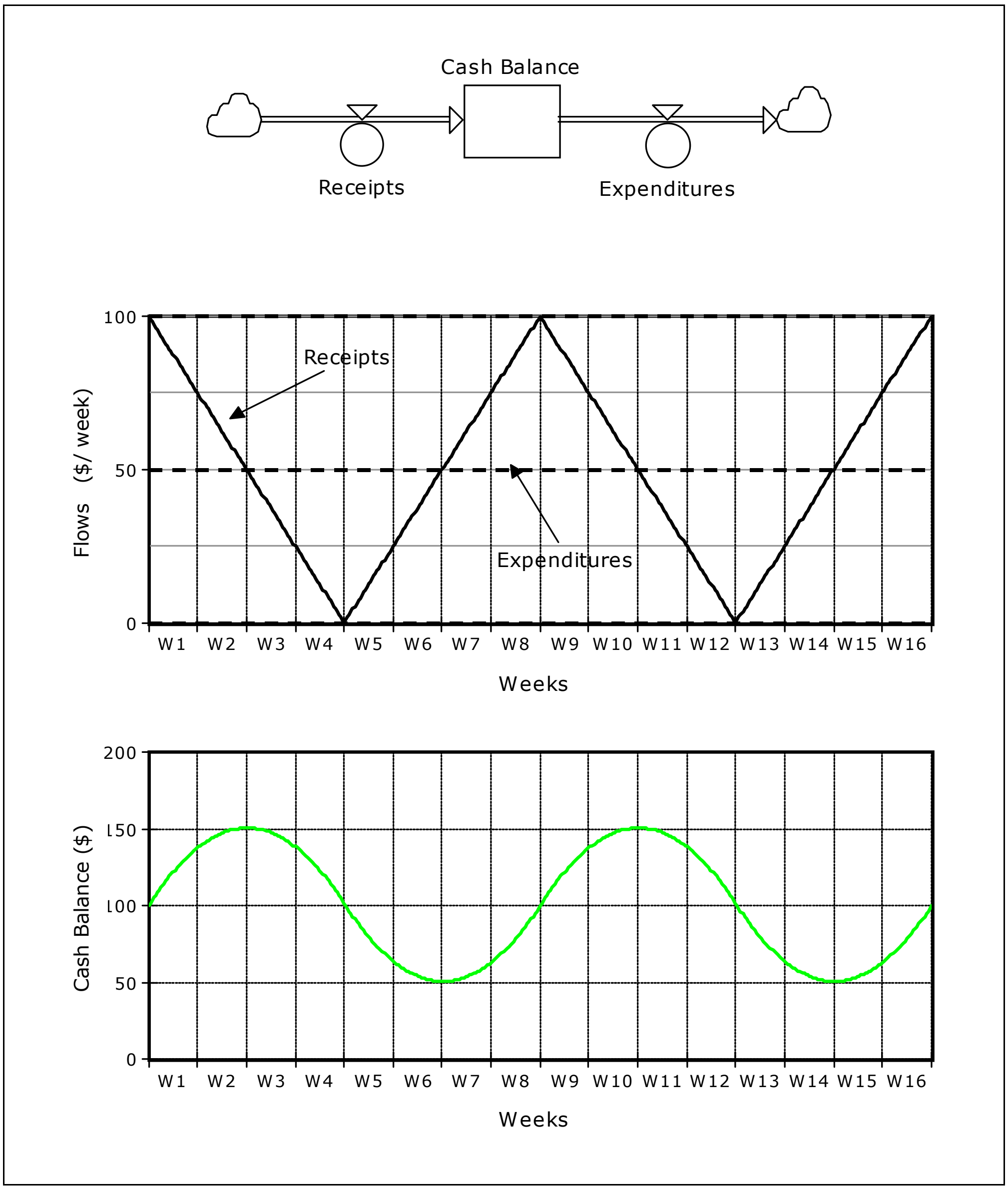


Figure 4 – Solution for task 2

Subject’s performance was evaluated by using the following criterions (Sweeney and Sterman, 2000) that are similar to those used in task 1:

C1- When the inflow exceeds the outflow, the stock is rising
C2- When the outflow exceeds the inflow, the stock is falling
C3- The peaks and troughs of the stock occur when the net flow crosses zero
C4- The stock should not show any discontinuous jumps
C5- The slope of the net stock at any time is the net rate. Therefore:

a) when the net flow is positive and falling, the stock is rising at a decreasing rate $(0<t<2;8<t<10)$;
b) when the net flow is negative and falling, the stock is falling at a increasing rate $(2<t<4;10<t<12)$;
c) when the net flow is negative and rising, the stock is falling at a decreasing rate $(4<t<6;12<t<14)$;
d) when the net flow is positive and rising, the stock is rising at a increasing rate $(6<t<8;14<t<16)$;

C6- The slope of the stock when the net rate is at its maximum is 50 units/ period (t= 0. 8, 16);
C7- The slope of the stock when the net rate is at its minimum is -50 units/ period (t= 4, 12);
C8- The quantity added to (removed from) the stock during each segment of two periods is the area enclosed by the net rate (triangle with area 1/2 x 50 units/time x 2 periods = 50 units). So the stock peaks at 150 units and falls to a minimum of 50 units.

Tasks 1 and 2 explore participants' understanding of the concept of stock accumulation, without any feedbacks or time delays. Complex systems often include feedback processes and time delays and they have a significant effect on their dynamics. For instance, time delays embedded in negative feedback loops can cause instability and oscillation. Task 3 that is based on "Manufacturing Task" (Sweeney and Sterman, 2000), differs from previous tasks as it includes both a negative feedback loop and time delays.

Task 3 is an example of a simple stock management task. In this task, the stock manager's objective is to keep the inventory stable at 50.000 units, by regulating the inflow to the stock in a negative feedback loop. In stock management systems often there are delays between the initiation of a control action and its effect. Here the stock manager faces a lag (four weeks) between a change in production schedule and the actual production rate.

Customer orders are constant for the first five weeks (10.000 units/ week). Then they suddenly increase to 11.000 units/ week. Participants are asked to draw the production rate in order to readjust inventory to its desired level (50.000), and to draw the behavior of the stock.

As there is a time delay (four weeks), production cannot be increased immediately. So, after the step increase in orders at week 5, inventory decreases linearly until week 9. The stock manager must not only equals the production rate to the new orders rate, but also rebuilt its inventory to the desired level. Therefore, production rate must overshoot orders rate and stay there until inventory is readjusted to its desirable level. After that point, production rate equals the orders rate (11.000).

Consider a manufacturing firm. The firm maintains an inventory of finished product. The firm uses this inventory to fill customer orders as they come in. Historically, orders have averaged 10,000 units per week. Because customer orders are quite variable, the firm strives to maintain an inventory of 50,000 units to provide excellent customer service (that is, to be able to fill essentially 100% of every order), and they adjust production schedules to close any gap between the desired and actual level. Although the firm has ample capacity to handle variations in demand, it takes time to adjust the production schedule, and to make the product – a total lag of four weeks.

Now imagine that the order rate for the firm's products suddenly and unexpectedly rises by 10%, and remains at the new, higher rate indefinitely, as shown in the graph below. Before the change in demand, production was equal to orders at 10,000 units/week, and inventory was equal to the desired level of 50,000 units.

Sketch the likely path of production and inventory on the graphs below. Provide an appropriate scale for the graph of inventory.

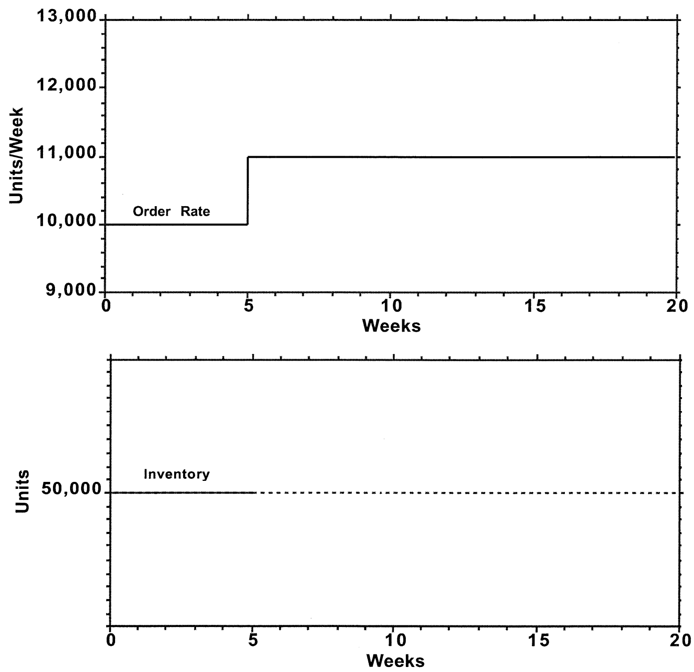


Figure 5 – Task 3

Source: Sweeney and Sterman, 2000, p 258

There is no unique correct answer to this task. The quantity of production above customer orders must equal the quantity of inventory lost during the period when orders exceed production. Figure 6 shows a possible response that is increasing production rate from 10.000 to 12.000 units/week for four weeks. As we can see, the area bounded by the production overshoot (quantity of production above customer orders) equals the area between orders and production (quantity of inventory lost during the period when orders exceed production). Participants are asked to draw the behavior of the stock as well. After week 5, inventory falls linearly in first 4 weeks to 46.000 units to increase linearly back to 50.000 units in week 13.

Answer from figure 6 assumes a pipeline delay (4 weeks) between Production and Desirable Production. In turn, Desirable Production is leading by Orders and is done by assuming the same pipeline delay. Production starts in equilibrium with Orders. When Orders change, Production will remain as its original level for 4 weeks. So, when Orders increase, 4 weeks later, Production raises to compensate the quantity lost four weeks ago due to Orders change.

Production (t + 4) = Orders (t) + "quantity lost four weeks ago due to Orders change"
or
Production (t + 4) = Orders (t) + (Orders (t) – Orders (t - 4))

Figure 6 presents a simple system dynamics model where the variables Desirable Production and Productions have the following equations:
Desirable Production (t) = Orders (t) + (Orders (t) – Orders (t - 4))
Production (t) = Desirable Production (t - 4)

Subject's performance was evaluated by using the following criterions (Sweeney and Sterman, 2000):

C1- Production must start in equilibrium with orders.
C2- Production must be constant prior to week 5 and indicate a lag of 4 weeks in the response to the step increase in orders.
C3- Production must overshoot orders to replenish the inventory lost during the initial period when orders exceed production. Production should return to the equilibrium rate of 11.000 units/week.
C4- Conservation of material: The area enclosed by production and orders during the overshoot of production (when production > orders) must equal the area enclosed by orders less production (when production < orders).
C5- Inventory must initially decline (because production < orders).
C6- Inventory must recover after dropping initially.
C7- Inventory must be consistent with the trajectory of production and orders.

Task 4 (see figure 7) is based on “Department Store Task” (Sterman, 2002) which in turn was inspired by a task developed by Ossimitz (2002). According to Sterman (2002), the reason people can perform poorly on Tasks 1 and 2 is not that they do not understand stocks and flows, but perhaps that they can not read graphs, or can not do graphical integration, or can not do the arithmetic, or are not given enough time. To do so, Sterman (2002) developed this task as an even simpler challenge (comparing to tasks 1 and 2) to test the understanding of stock-flow relationship.

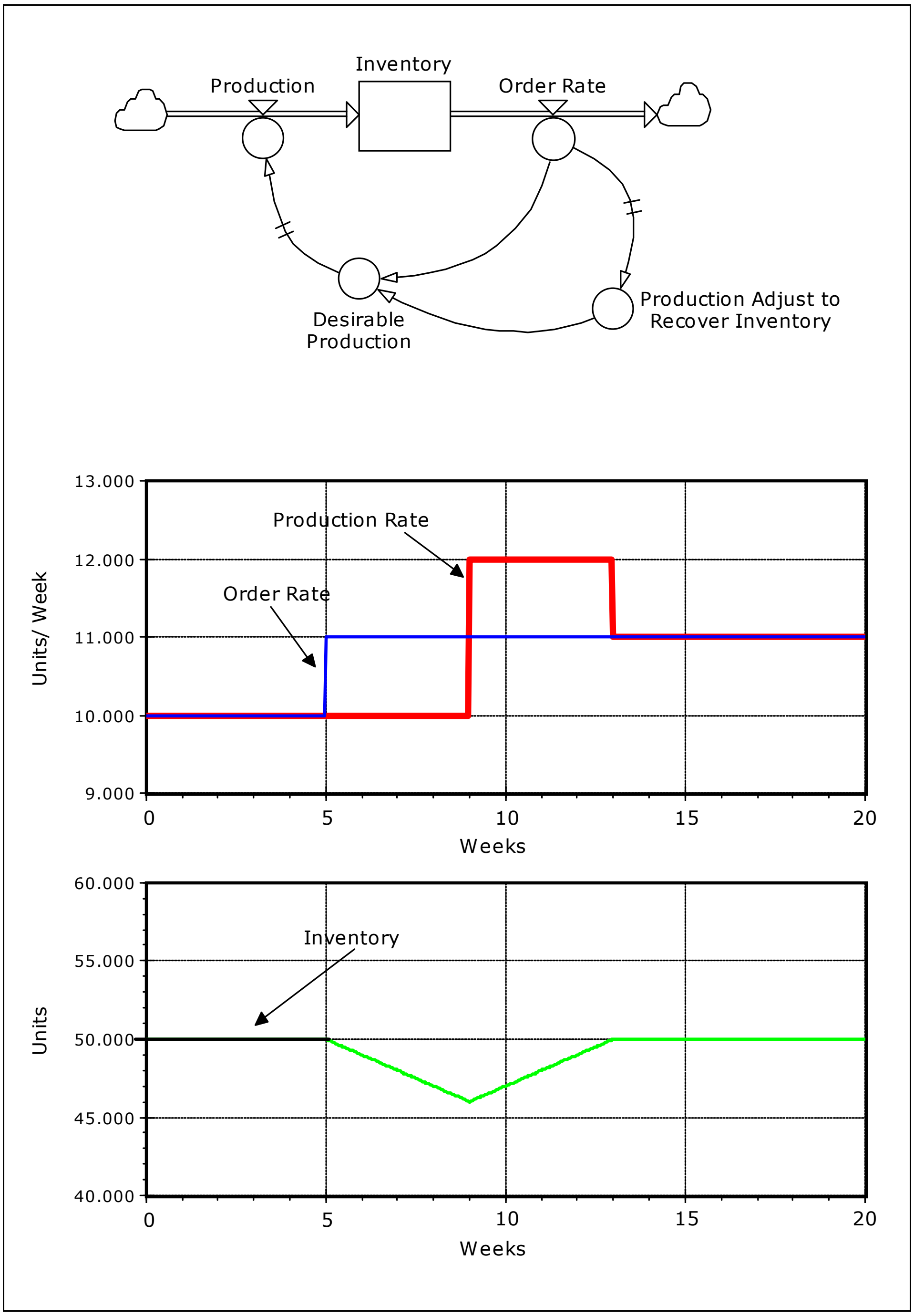
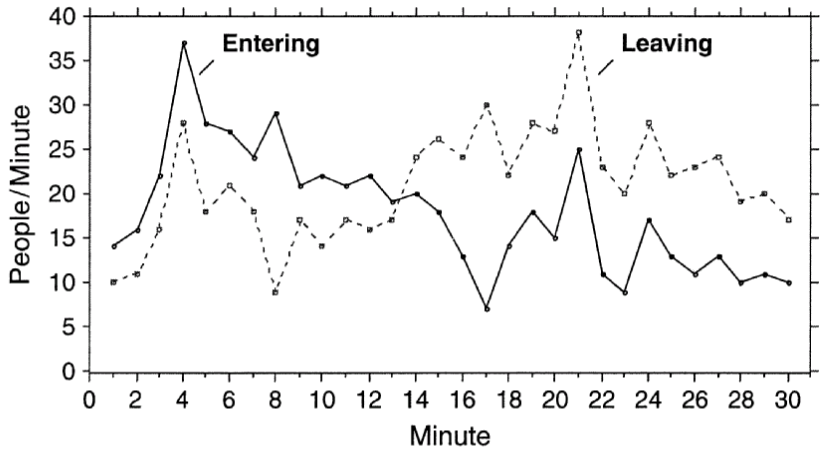


Figure 6 – A correct response to Task 3

A graph shows an inflow and an outflow indicating the number of people entering and leaving a department store per minute over a period of 30 minutes. Participants are asked to answer four questions. The first two questions refer to the flows ("during which minute did the most people enter/leave the store?"). The last two questions refer to the accumulation of the two flows ("during which minute were the most/fewest people in the store?").

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



Please answer the following questions.

Check the box if the answer cannot be determined from the information provided.

1. During which minute did the most people enter the store?

   Minute ___4___

2. During which minute did the most people leave the store?

   Minute ___21___

3. During which minute were the most people in the store?

   Minute ___13___

4. During which minute were the fewest people in the store?

   Minute ___30___

Figure 7 – Task 4 and its solution

Source: Sterman, 2002, p 510

The number of people in the store accumulates the flow of people entering less the flow of people leaving. From looking at the graph we can see that until minute 13 (where the two curves cross) the inflow exceeds the outflow. Therefore, the number of people in the store increases until minute 13 and afterwards decreases until minute 30. Thus the most people are in the store at minute 13 (question 3). To answer to last question, participants must determine whether the area between the inflow and the outflow curves up to minute 13 is greater or smaller than the area between the curves afterwards. Looking at the graph we can state that the area between the two curves from minute 13 on is larger than the area between curves through minute 13. Thus, more people left after minute 13 than were added up that point and so there must be less people in the store at the end than in the beginning. Therefore the fewest people are in the store at minute 30 (question 4).

3. Participants

This research was conducted with two very different groups: at ISCTE (a business graduate school in Lisbon) and at Galp Energia, one of the biggest Portuguese firms (the Portuguese oil company). At the ISCTE the group consisted of 43 undergraduate students in their last year of Business Degree. Their age ranged from 21 to 25. This group of subjects was relatively homogeneous. At Galp Energia the tasks were performed by a more diversified group of 31 managers and engineers. None of the subjects had any prior experience with systems thinking or system dynamics.

Total Number of Participants	31
Gender	
Female	6
Male	26
Age	
< 30	10
30 – 40	6
40 – 50	8
> 50	7
Prior Field of Study	
Engineering/Math/Science	16
Management/Social Sciences	15
Highest Prior Degree	
Five Years Degree	24
MBA/ MS	7
Job Level	
1 (high)	9
2 (middle)	16
3 (low)	6

Table 1 – Participants demographics - Galpenergia

As these two groups of subjects are different regarding to demographical data, the results are presented separately in order to provide the comparability of the two samples.

The participants at Galpenergia were asked to fill out a form about their gender, age, prior field of study (Engineering/Math/Science or Management/Social Sciences), highest prior degree and job level (table 1).

As can be seen from the table 1, there was a clear minority of female participants (19%). The participants were fairly distributed on "age" and "prior field of study". A great part of the participants had a five years degree (77%) and there was a predominance of them in middle job positions. None of the participants had any prior experience with system dynamics.

4. Procedure

At the ISCTE, the students went through two tests. The first test considered the tasks 1, 2 and 3 and 43 students were present. They had nearly 40 minutes in order to complete the test. In the second test 33 students were given the task 4 "Department Store Task" and they had nearly 15 minutes time.

At Galp Energia the 31 participants handled the tasks in their own work place. This research was carried out in two sessions. In the first session the participants received a demographics information form, the Task 1 and the Task 2. Some days later, the same participants received the Task 3 and the Task 4. The participants were given approximately 30 minutes in each session.

Before answering the tests, the participants of both groups were told that the purpose of the tasks was to participate in a survey to gain insight into people's understanding of some systems thinking concepts. The tests were anonymous and without rewards. As the participants from Galp Energia did not answer the tasks simultaneously, they were told that it was an individual research and so that they must not share with other participants.

5. Results

Statistical results on the four tasks for the two groups of subjects are provided and some of the interesting solutions as well. We also compare the results with those of other universities, Tor Vergata University in Rome (Armenia et al., 2004); Stuttgart Institute of Management and Technology (Kapmeier, 2004) and MIT (Sweeney and Sterman, 2000; Sterman, 2002).

5.1 Task 1 "Bathtub Task"

Regarding the Task 1 "Bathtub task" it can be seen in table 2 that performance in both groups were poor (tables in appendix A2.1 and B2.1 show complete data). Average performance on this task was 28% at ISCTE and 58% at Galp Energia.

Taking into consideration every item, the participants at Galp Energia performed much better than the participants at ISCTE. The performance of both groups in general terms was worse than with other groups (TV, SIMT and MIT).

Criterion	ISCTE	GALP	TV	SIMT	MIT
C1- When the inflow exceeds the outflow, the stock is rising	0.28	0.65	0.70	0.77	0.80
C2- When the outflow exceeds the inflow, the stock is falling	0.30	0.61	0.70	0.68	0.80
C3- The peaks and troughs of the stock occur when the net flow crosses zero	0.30	0.61	0.64	0.64	0.86
C4- The stock should not show any discontinuous jumps	0.33	0.45	0.73	0.82	0.89
C5- During each segment the net flow is constant so the stock must be rising (falling) linearly	0.28	0.61	0.74	0.59	0.78
C6- The slope of the stock during each segment is the net rate	0.23	0.55	0.56	0.50	0.66
C7- The quantity added to (removed from) the stock during each segment of four periods is the area enclosed by the net rate (25 units/time x 4 time periods = 100 units, so the stock peaks at 200 units and falls to a minimum of 100 units).	0.23	0.55	0.47	0.55	0.63
Mean for all items	0.28	0.58	0.65	0.65	0.77
Standard Deviation	0.42	0.46	0.10	-	0.34

Table 2 –Performance on the Task 1 "Bathtub task"

The first five items describe qualitative and intuitive features of the stock accumulation process and do not require any calculation. Items 6 and 7 describe the trajectory of the stock quantitatively.

At Galp Energia, participants did best on item 1 (When the inflow exceeds the outflow, the stock is rising) and they did worst on item 4 (The stock should not show any discontinuous jumps). Just like it was observed in the other three researches, participants at ISCTE did better on item 4. Criterion 6 (The slope of the stock during each segment is the net rate) and Criterion 7 (The quantity added to/removed from the stock during each segment of four periods is the area enclosed by the net rate) were ISCTE participants' worst items.

The first 3 items are fundamental to the concept of stock accumulation. More than 70% of the ISCTE group and more than 35% of the Galp Energia group failed on items 1, 2, 3. The results observed with other groups were better as this failure occurred from 20% (MIT) to 32% (SIMT) of the subjects.

As can be seen in appendix A2.1 and B2.1, all of the participants that correctly response to items 1 and 2 also gave correct answers to item 3 (the peaks and troughs of the stock occur when the net flow crosses zero).

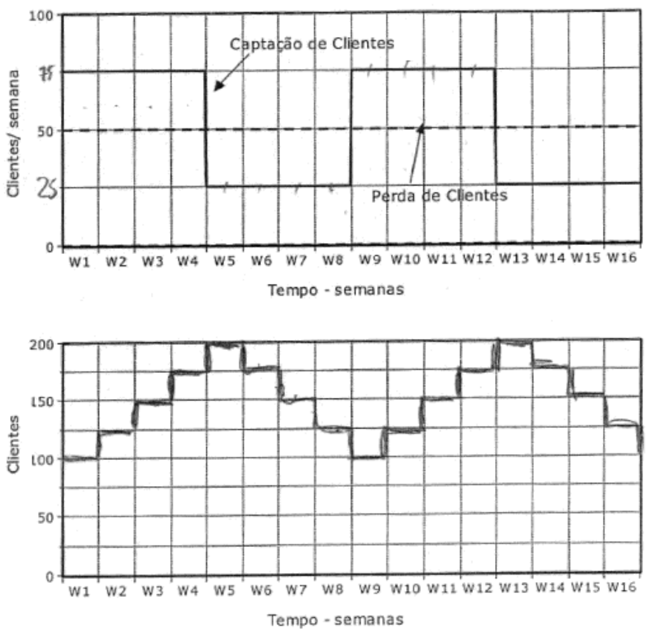


Figure 8 – Error "spreadsheet thinking"

As we stated before, item 4 was Galp Energia participants' worst item. Figure 8 shows an error that was observed in 20% of this group. The participants drew the stock trajectory changing suddenly at the end of every week. They assume that the flows are discrete, with inflows and outflows only occurring at the end of every period. Sweeney and Sterman (2000, p 267) reported the same error. According to the authors those subjects suffered from "spreadsheet thinking". None of the ISCTE' participants exhibited "spreadsheet thinking". It suggest that, due to their professional experience, the participants at Galp Energia tend to suffer more from "spreadsheet thinking" than the ISCTE' participants.

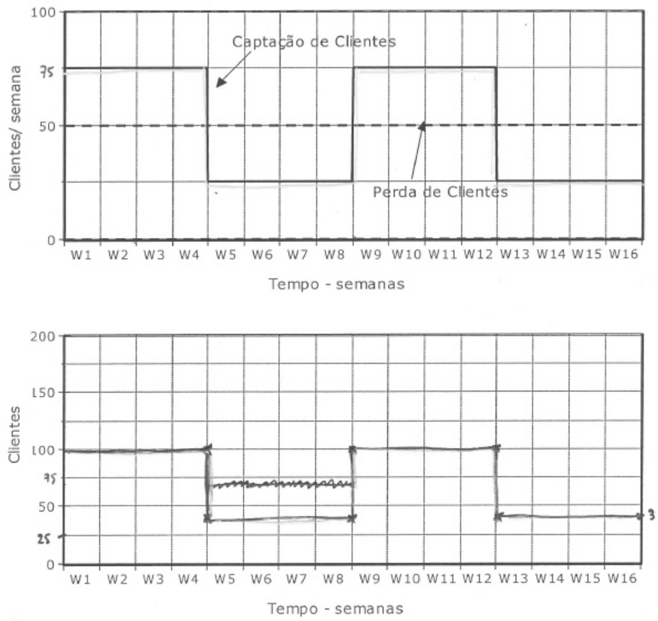


Figure 9 – Error "matching of stock and net inflow trajectories"

Figure 9 illustrates the most typical error of both groups (exhibited by 35% of the participants at ISCTE and 29% of the participants at Galp Energia). The participants drew the stock by matching its pattern to the net inflow. The stock trajectory changed discontinuously, jumping up and down in phase with the net flow rate.

The participants' responses suggest that their poor performance arose not from numerical errors in the required calculations, but due to the participants' misunderstanding of stock-flow relationship. In fact, the responses suggest that participants were confused about the definitions of stocks and flows and did not understand the basic relationship between a net flow and its impact on stock accumulation, in particular, that the change in the stock over a period is the area bounded by the net rate in that period.

5.2 Task 2 “Cash Flow Task”

As in the other surveys both groups found this task with the sawtooth pattern more difficult than the square wave pattern from previous task. Overall performance was poorer than in the task 1. According to Table 3, average performance was 44% for the Galp Energia participants and only 22% for the ISCTE participants (complete data in appendix A2.2 and B2.2).

When compared with the other three surveys, the ISCTE group revealed the worst performance and the Galp Energia group exhibited an average performance (poorer than MIT and better than TV and SIMT).

Criterion	ISCTE	GALP	TV	SIMT	MIT
C1- When the inflow exceeds the outflow, the stock is rising	0.14	0.58	0.39	0.27	0.47
C2- When the outflow exceeds the inflow, the stock is falling	0.14	0.55	0.36	0.32	0.44
C3- The peaks and troughs of the stock occur when the net flow crosses zero	0.14	0.55	0.52	0.32	0.40
C4- The stock should not show any discontinuous jumps	1,00	1,00	0.96	0.86	0.99
C5- The slope of the net stock at any time is the net rate. Therefore: a) when the net flow is positive and falling, the stock is rising at a decreasing rate (0<t<2;8<t<10); b) when the net flow is negative and falling, the stock is falling at a increasing rate (2<t<4;10<t<12); c) when the net flow is negative and rising, the stock is falling at a decreasing rate (4<t<6;12<t<14); d) when the net flow is positive and rising, the stock is rising at a increasing rate (6<t<8;14<t<16).	0.12	0.26	0.35	0.23	0.28
C6- The slope of the stock when the net rate is at its maximum is 50 units/ period (t= 0, 8, 16);	0.05	0.19	0.21	0.23	0.47
C7- The slope of the stock when the net rate is at its minimum is -50 units/ period (t= 4, 12);	0.09	0.23	0.20	0.14	0.45
C8- The quantity added to (removed from) the stock during each segment of two periods is the area enclosed by the net rate (triangle with area 1/2 x 50 units/time x 2 periods = 50 units). So the stock peaks at 150 units and falls to a minimum of 50 units.	0.07	0.13	0.32	0.32	0.37
Mean for all items	0.22	0.44	0.41	0.34	0.48
Standard Deviation	0.24	0.30	0.26	-	0.35

Table 3 –Performance on the Task 2 “Casf Flow Task”

As in Task 1, the first five items describe qualitative and intuitive features of the stock accumulation process and do not require any calculation. Here, the first 3 items are also fundamental to the concept of stock accumulation. A low number of ISCTE participants (14%) answered correctly to items 1,2,3. At Galp Energia about 45% of the participants failed to do so.

As in Task 1, all of the participants that correctly response to items 1 and 2 also gave correct answers to item 3 (see appendix A2.2 and B2.2).

All participants from both groups answered correctly to criterion 4 “The stock should not show any discontinuous jumps”. This is the only item where participants did better in Task 2 than in Task 1. As the net flow rate in Task 2 is continuous while in Task 1 is discontinuous, it suggests that many participants failed this item in Task 1 because they drew stock trajectories that matched the pattern of the net flow rate (Sweeney and Sterman, 2000, p 269).

Criterion 5 evaluates if participants correctly related the net flow rate to the stock. Only 12% of ISCTE participants and 26% of Galp Energia participants succeeded in this item.

Items 6,7,8 describe the trajectory of the stock quantitatively, but the arithmetic required is trivial. Answering to these items also requires participants to read and interpret the graph of the flow rates. Participants must also know the formula for the area of a triangle (to apply graphical integration) to calculate the maximum and the minimum of the stock (for item 8) and be able to construct a line for the stock trajectory showing the slope of the stock properly at the inflection points t = 0. 4, 8, 12 and 16 (for items 6,7). Less than 10% of the ISCTE participants and less than 20% of the Galp Energia participants responded to these items correctly.

As in Task 1, a great part of the participants (28% at ISCTE and 20% at Galp Energia) drew the stock trajectory by matching its pattern to the net inflow. Figure 10 shows an example of this error.

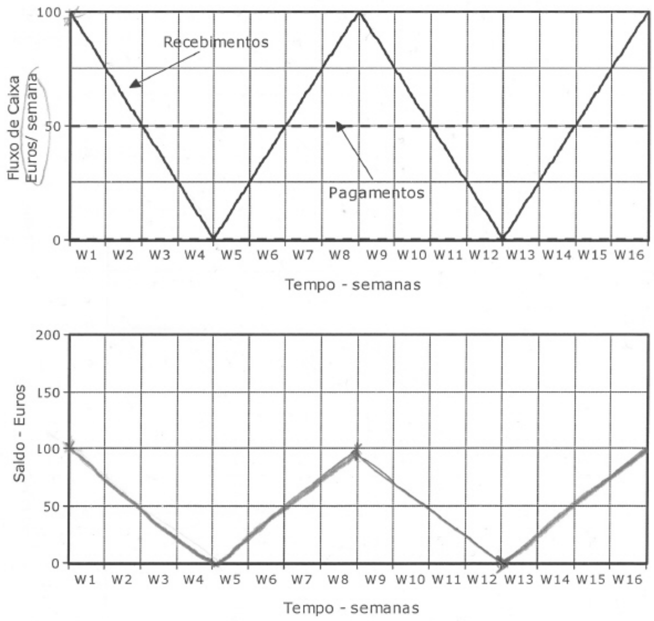


Figure 10 – Error “matching of stock and net inflow trajectories”

A great part of the Galp Energia participants (23%), even though they answered correctly to items 1,2,3,4, they drew the stock trajectory linearly and so that they failed the items 6,7 (example in figure 11).

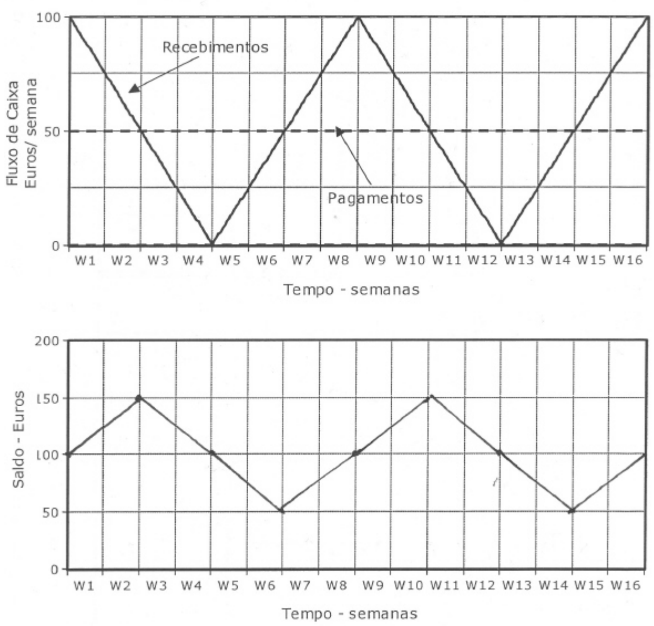


Figure 11 – Error “linear trajectory of stock”

As reported by Sweeney and Sterman (2000, p 271), a participant at Galp Energia attempted to solve the problem analytically by writing the formulas for the integrals of the net flows.

5.3 Task 3 “Manufacturing Task”

Even though this task is more difficult than the two previous tasks, it is still simple in content and structure, involving only one stock, one time delay, and one negative feedback loop. Overall performance was poor. The Galp Energia group performed much better than the ISCTE group. According to Table 4, average performance was 53% for the Galp Energia participants and only 14% for the ISCTE participants (complete data in appendix A2.3 and B2.3). The average performance of the ISCTE participants was worse than in the Task 1 and Task 2. Participants at Galp Energia did worse than in Task 1 and better than in Task 2.

When compared with the other three surveys, the ISCTE group revealed the worst performance and the Galp Energia group exhibited an average performance (poorer than TV and better than SIMT and MIT).

Only 2% of the ISCTE participants and 16% of the Galp Energia participants showed correct solutions to this problem.

Criterion	ISCTE	GALP	TV	SIMT	MIT
C1- Production must start in equilibrium with orders (=10.000 units/ week)	0.23	0.71	0.74	0.45	0.53
C2- Production must be constant prior to week 5 and indicate a lag of 4 weeks in the response to the step increase in orders	0.05	0.61	0.63	0.36	0.44
C3- Production must overshoot orders to replenish the inventory lost during the initial period when orders exceed production. Production should return to the equilibrium rate of 11.000 units/week	0.05	0.39	0.46	0.23	0.44
C4- Conservation of material: The area enclosed by production and orders during the overshoot of production (when production > orders) must equal the area enclosed by orders less production ((when production < orders)	0.05	0.32	0.37	0.09	0.11
C5- Inventory must initially decline (because production < orders)	0.42	0.74	0.72	0.45	0.68
C6- Inventory must recover after dropping initially	0.14	0.71	0.59	0.36	0.56
C7- Inventory must be consistent with the trajectory of production and orders	0.02	0.19	0.51	0.18	0.10
Mean for all items	0.14	0.53	0.57	0.30	0.41
Standard Deviation	0.22	0.36	0.40	-	0.28

Table 4 –Performance on the Task 3 “Manufacturing Task”

Nearly 70% of the Galp Energia participants determined correctly that production (criterion 1) starts in equilibrium with orders and only 23% at ISCTE succeeded on this issue. Only 5% of the ISCTE participants answered to item 2 by considering the production adjustment delay (61% at Galp Energia). More than 60% of the Galp Energia participants and 95% of ISCTE participants failed on item 3 (production trajectory must overshoot orders in order to replenish the inventory). Only 32% of the Galp Energia participants and 5% of ISCTE participants understood the conservation of material and drew correct production trajectories (criterion 4). Both groups did best on criterion 5 "Inventory must initially decline because production < orders" (42% at ISCTE and 74% at Galp Energia). The two groups showed a quite different performance on item 6 "Inventory must recover after dropping initially" (14% at ISCTE and 71% at Galp Energia). The great majority of the two groups participants drew inventory paths that were inconsistent with the according production paths (criterion 7), what means that they failed to relate the flows to the stock correctly (98% at ISCTE and 81% at Galp Energia).

Figure 12 illustrates an error showed by 26% of the Galp Energia participants and 9% of the ISCTE participants. They drew the inventory declining during the lag. Inventory then rose back to equilibrium (50.000 units) even though production equalled orders. They did not show the production overshoot in order to replenish the inventory.

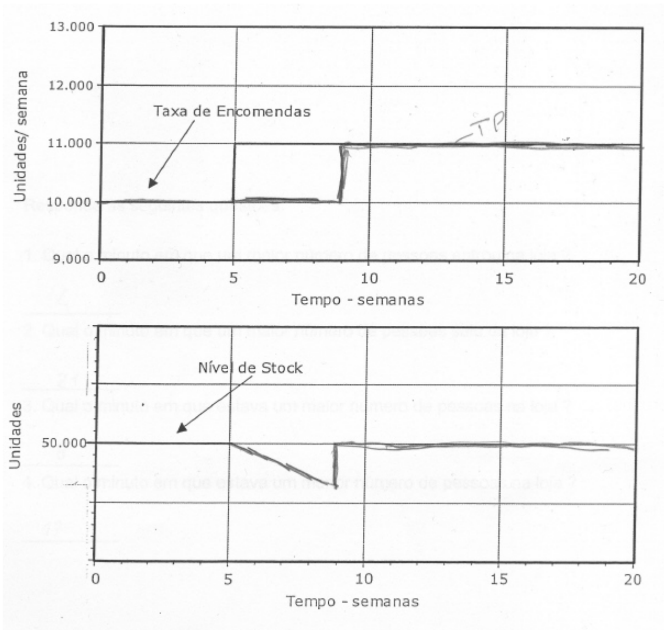


Figure 12 – Error "stock recover without production overshoot"

Figures 13 and 14 show other two common errors. At ISCTE, 16% of the participants drew a trajectory for inventory that just matched the production path. Another error consisted in showing stock path by matching its pattern to the inverse pattern of the production (14% at ISCTE). Notice that a great majority of these participants (87%) also showed this error in Tasks 1 and 2.

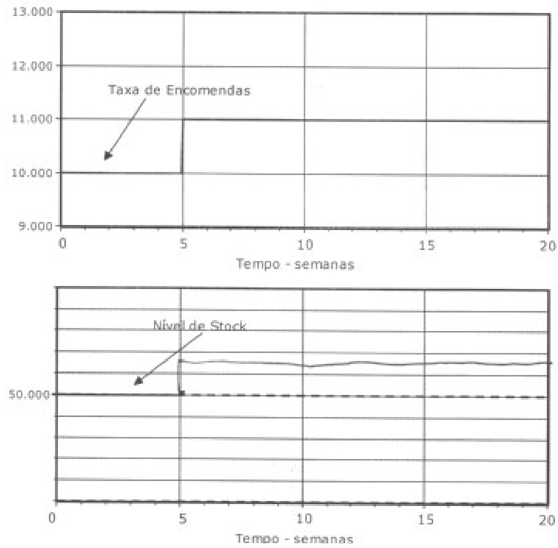


Figure 13 – Error regarding to pattern matching (a)

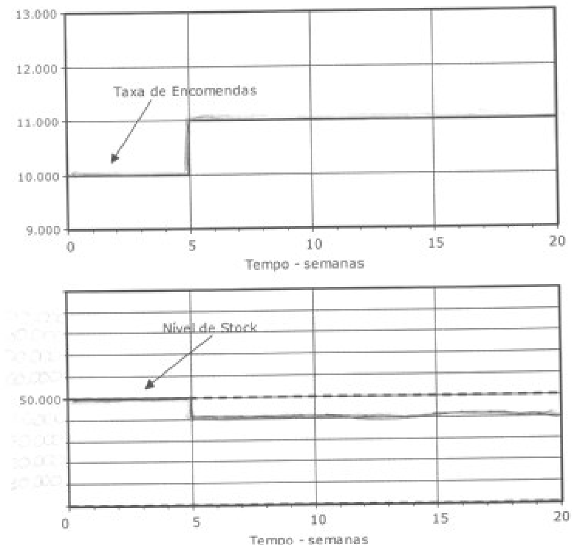


Figure 14 – Error regarding to pattern matching (b)

A frequent erroneous response the participants at Galp Energia (26%) showed was what Sweeney and Sterman (2000) called “spreadsheet thinking”: participants drew short paths with steps at the end of each week for inventory (example in figure 15). Interestingly, none of ISCTE participants applied “spreadsheet thinking” in this task.

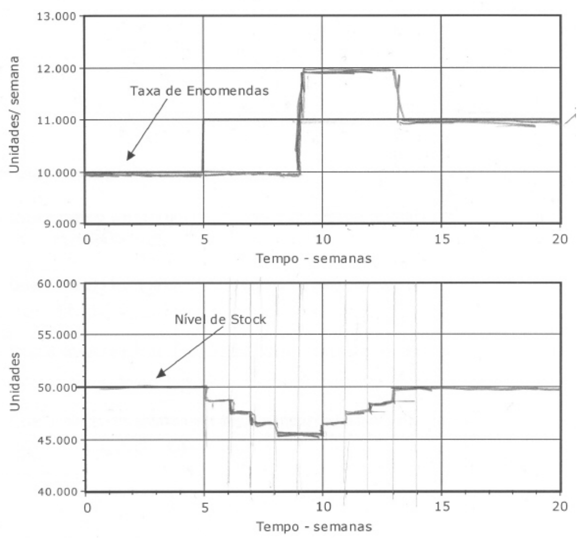


Figure 15 – Error “spreadsheet thinking”

5.4 Task 4 “Department Store Task”

According to Sterman (2002), as this task does not require participants to apply graphical integration nor arithmetic, but only the ability to read graphics, it is a more proper challenge to test the intuitive understanding of stock-flow relationship.

Table 5 shows the performance on this task for both groups and the complete data can be seen in appendix A2.4 and B2.4. All participants of the Galp Energia group and a great majority of the ISCTE participants stated correctly when the most people enter and leave the store (Q1=4, Q=21). As participants had to understand the flows of people drawn on the graph to answer to these questions correctly, one can conclude that those participants can indeed read and interpret graphs properly.

The questions 3 and 4 test the participants’ understanding of stocks and flows. As in other three surveys, the participants did worst on questions 3 and 4 regarding to the accumulations of people entering and leaving. As stated before, to answer to question 3, participants had to look at the graph and see that until minute 13 (where the two curves cross) the inflow exceeds the outflow and so that the number of people in the store increases until minute 13 and afterwards decreases until minute 30. Thus the most people are in the store at minute 13 (question 3). Only 12% of the ISCTE participants and 55% of the Galp Energia participants respectively answered correctly.

To answer to question 4, participants must judge whether more people enter up to minute 13 than leave afterwards, that is, whether the area between the flow curves up to minute 13 is greater or smaller than the area between the curves afterwards. As the area from minute 13 on is larger than the area through minute 13, the fewest people are in the store at minute 30 (question 4). Only 3% of the ISCTE participants and 35% of the Galp Energia participants succeeded.

Question	ISCTE	GALP	TV	SIMT	MIT
Q1- During which minute did the most people enter the store?	0.97	1,00	0.98	1,00	0.94
Q2- During which minute did the most people leave the store?	0.91	1,00	0.90	0.95	0.94
Q3- During which minute were the most people in the store?	0.12	0.55	0.33	0.42	0.42
Q4- During which minute were the fewest people in the store?	0.03	0.35	0.24	0.26	0.30
Mean for all questions	0.51	0.73	0.61	0.66	0.65
Standard Deviation	0.12	0.23	0.38	-	0.34

Table 5 – Performance on the Task 4 "Department Store Task"

Indeed, answering to questions 3 and 4, does not require any calculation – only read the graph and understand that a stock rises when its inflow exceeds its outflow and falls when its outflow exceeds its inflows. As stated before, participants succeeded on questions 1 and 2, then one can conclude that they can read and interpret graphs properly. It suggests that the participants, who failed on questions 3 and 4, had no problem in reading graphs, but difficulty with the concept of stock-flow relationship.

Nearly 40% of the ISCTE participants and 30% of the Galp Energia participants answered that the most people were in the store in minute 8 (question 3) and the fewest people were there in minute 17 (question 4). Notice that the minute 8 is the moment when the difference between the inflow and outflow is at its maximum (about 21 people/min) and minute 17 is the moment when the difference between the two flows is at its minimum (about – 23 people/min). Minutes 8 and 17 are the moments where the net flow is at its maximum and minimum respectively. Those participants failed to differentiate between an accumulation and the net flow rate. It suggests that those participants were confused and they did not understand the concept of stock-flow relationship. Interestingly, at Galp Energia, nearly 50% of those participants answered to Task 1 correctly. It suggests that those participants solved Task 1 by using graphical integration, but as it was revealed in Task 4 they did not understand intuitively the concept of stock-flow relationship.

5.5 Correlation Between Tasks

In this section, we present the analysis of the correlation between the tasks results (mean for all items in each task) for the two groups of participants. As stated before, the ISCTE participants went through two tests. The first test considered the tasks 1, 2 and 3, and in second test the task 4 was given. As we did not keep any reference between the first and the second test, we are not able to implement the correlation analysis between task 4 and any other task.

	Task 1		Task 2		Task 3	
	Pearson Correlation	Significance	Pearson Correlation	Significance	Pearson Correlation	Significance
Task 1			0.555**	0.000	0.312*	0.041
Task 2					0.363*	0.016

**Correlation is significant at the level ρ<0.001; *Correlation is significant at the level ρ<0.05

Table 6 – Correlations between tasks for ISCTE group

Regarding the ISCTE group, table 6 shows the correlation (Pearson) matrix for the regression analysis between task results. As can be seen, the results of the three tasks are significantly correlated.

Table 7 shows the correlation (Pearson) matrix for the Galp Energia group. We did not find significant correlations between tasks, this is, the participant's performance in a given task, did not imply a similar result in other tasks. The only exception is for the pair Task 3 - Task 4 (correlation =0.516; ρ=0.003).

	Task 1		Task 2		Task 3	
	Pearson Correlation	Significance	Pearson Correlation	Significance	Pearson Correlation	Significance
Task 1						
Task 2	-0.046	0.804				
Task 3	-0.052	0.783	0.234	0.205		
Task 4	0.151	0.416	0.053	0.779	0.516*	0.003

*Correlation is significant at the level ρ<0.01

Table 7 – Correlations between tasks for Galp Energia group

As all the three tasks 1, 2, and 4 consider the accumulation process of a single stock through an inflow and an outflow, we expected to see significant correlations between task 4 and task1/2 results. As stated before in section "task 4 results", it suggests that some participants answered Task 1 by using graphical integration correctly, but in fact they did not intuitively understand the concept of stock accumulation as it was revealed in their Task 4 results.

6. Impact of Participant Demographics

The Galp Energia participants filled out a demographics information form about their gender, age, highest prior degree, prior field of study, and job level (table 1). Table 8 shows the mean and standard deviation for performance according to different demographic data.

		Task 1		Task 2		Task 3		Task 4	
Gender	**N**	**μ**	**σ**	**μ**	**σ**	**μ**	**σ**	**μ**	**σ**
Female	6	0,17	0,41	0,56	0,37	0,48	0,40	0,71	0,25
Male	26	0,67	0,42	0,41	0,28	0,54	0,35	0,73	0,23
Age									
< 30	10	0,47	0,50	0,64	0,23	0,60	0,37	0,75	0,24
30 – 40	6	0,90	0,12	0,40	0,22	0,57	0,33	0,63	0,21
40 – 50	8	0,39	0,48	0,25	0,27	0,45	0,40	0,72	0,25
> 50	7	0,65	0,46	0,39	0,36	0,47	0,37	0,79	0,22
Highest Prior Degree									
Five Years Degree	24	0,58	0,46	0,48	0,31	0,51	0,35	0,71	0,22
MS, MBA	7	0,57	0,46	0,29	0,24	0,59	0,38	0,79	0,27
Prior Field Study									
Engineering/ Science	16	0,59	0,45	0,41	0,31	0,49	0,40	0,66	0,22
Management/S. Sciences	15	0,56	0,48	0,46	0,30	0,56	0,32	0,80	0,22
Job Level									
1	9	0,51	0,49	0,54	0,31	0,63	0,36	0,75	0,25
2	16	0,53	0,49	0,44	0,30	0,51	0,30	0,72	0,22
3	6	0,69	0,28	0,25	0,23	0,43	0,49	0,75	0,25

Table 8 - Mean and standard deviation for performance according to different demographic data – Galp Energia group

Due to the small number of participants in the present study, it is not reasonable to look for significance. However, to test the hypotheses that demographic variables would affect performance, we ran a variety of regression linear models relating performance on the different tasks to the various demographic variables participants reported.

Table 9 reports regression models in which performance on each task is explained by all demographic variables. We refined the regression models by performing a stepwise regression procedure in order to maximize R^2adjusted, to exclude the variables that do not seem to significantly explain the performance on the different tasks and to keep the most explanatory variables (table 10). The regressions were run on standardized values for all variables to be able to directly compare the relative effect of each variable on the performance.

As shown in table 10, it is possible to observe some impact of participant demographics on performance in some tasks. However, there is no consistent pattern. While some variables are significant in some tasks they are far from significant in the other tasks. Performance appears to depend on the participants' gender in Task 1, age and highest prior degree in the Task 2, no variable in Task 3 and prior field of study in Task 4.

	Task 1		**Task 2**		**Task 3**		**Task 4**	
Variable	Standardized β	Significance ρ	Standardized β	Significance ρ	Standardized β	Significance ρ	Standardized β	Significance ρ
Gender	0.507**	0.015	-0.058	0.768	0.201	0.353	0.134	0.500
Age	-0.226	0.352	-0.428*	0.087	-0.029	0.913	0.412	0.099
H. Degree	-0.150	0.449	-0.334	0.102	0.168	0.436	0.263	0.193
Field Study	-0.135	0.473	-0.021	0.910	-0.134	0.512	-0.388**	0.048
Job Level	0.253	0.309	0.067	0.788	-0.311	0.250	-0.413	0.104
R^2 adjusted	0.108		0.096		-0.055		0.094	
ρ	0.165		0.187		0.366		0.190	

** Significant at the level ρ<0.05; * Significant at the level ρ<0.1

Table 9 - Regression results for all variables

Gender is significant for Task 1 (β=0.448, ρ=0.012), suggesting that males outperformed females. However, due to the uneven distribution of males and females in this study, it is not reasonable to make any suggestions concerning the gender effect.

The results (β=-0.408, ρ=0.02) show that the variable "age" did have an impact on the performance in Task 2. Participants performed better in that task the younger they were. The variable 2 "highest prior degree" also appears to influence the performance in that task (β=-0.314, ρ=0.068). Participants with five years degree performed better than those with master degree. However, as in variable "gender", due to the small number of participants with master degree, it is not reasonable to conclude on this variable effect.

	Task 1		Task 2		Task 3		Task 4	
Variable	Standardized β	Significance ρ	Standardized β	Significance ρ	Standardized β	Significance ρ	Standardized β	Significance ρ
Gender	0.448**	0.012	-	-	-	-	-	-
Age	-	-	-0.408**	0.020	-	-	-	-
H. Degree	-	-	-0.314*	0.068*	-	-	-	-
Field Study	-	-	-	-	-	-	-0.322*	0.077
Job Level	-	-	-	-	-	-	-	-
R^2 adjusted	0.173		0.186		<0.019		0.073	
ρ	0.012		0.021		>0.2		0.077	

** Significant at the level $\rho<0.05$; * Significant at the level $\rho<0.1$

Table 10 - Regression results with most explanatory variables remaining from a stepwise regression.

Sweeney and Sterman (2000) and Kapmeier (2004) observed a significant impact of subjects prior academic field on performance in some tasks. Subjects with technical backgrounds performed better than those in social sciences. We also expected that participants with more training in mathematics or engineering would perform better than those in management/ social sciences. In the present study, we only found some effect of the variable "prior field of study" on the performance in Task 4 (β=-0.322, ρ=0.077). Interestingly, we did find the opposite. The results indicate that participants in management/ social sciences performed better than those in engineering and sciences.

Considering that Task 4 consists in the most proper problem to test stock-and-flow thinking, it suggests that a more training in mathematics do not help participants to better understand the concept of stock accumulation.

The results show that the variable "job level" appears to not have any impact on the performance in all Tasks. Interestingly, none of the demographic variables seems to influence the performance in the most difficult task (Task 3).

As we did not find any variable that have a consistent impact on the performance in all tasks, it is clear that the results from the present study (Galp Energia group) do not provide full support for the hypotheses that demographic variables affect participants' performance.

7. Discussion

The tasks that were applied in the present survey do not require participants to use any of the analytic tools of calculus and can be answered without use of any mathematics beyond simple arithmetic. Task 4, in particular, only requires participants to read and interpret the graph. Furthermore, as the tasks consider simple situations that participants are confronted with in their everyday lives, it is reasonable to assume that the reason for the poor results was not the unfamiliarity of the situations in the tasks.

As can be seen in table 11, both groups showed an overall poor performance. The ISCTE participants showed a worse overall performance than those obtained at TV, SIMT and MIT. Even though overall performance was poor, comparing to other studies, the Galp Energia participants did perform relatively well. Overall, the performance of the Galp Energia participants was similar to that of the other studies with an average on Task 1 of 0.58, on Task 2 of 0.44, on Task 3 of 0.53 and on Task 4 of 0.73. For example, MIT participants achieved 0.77, 0.48, 0.41 and 0.65 respectively.

	ISCTE		GALP		TV		SIMT		MIT	
	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
Task 1	0.28	0.42	0.58	0.46	0.65	0.10	0.65	-	0.77	0.34
Task 2	0.22	0.24	0.44	0.30	0.41	0.26	0.34	-	0.48	0.35
Task 3	0.14	0.22	0.53	0.36	0.57	0.40	0.30	-	0.41	0.28
Task 4	0.51	0.12	0.73	0.23	0.61	0.38	0.66	-	0.65	0.34

Table 11 – Mean and standard deviation for performance in all tasks

Comparing the results at Galp Energia to other studies, it appears that participants improved their performance progressively. In fact, they did worse in Task 1, they exhibited an average performance in tasks 2 and 3, and they performed better in Task 4. One explanation could be that as the tasks were carried out in two sessions (the participants performed Task 1 and Task 2 and some days later they did Task 3 and Task 4), perhaps the participants learnt from session 1 and started session 2 with an improved understanding of the concepts that were being tested.

There is a remarkable difference between the performance of the ISCTE and Galp Energia groups. The better performance found at Galp Energia might be explained through the differing professional experience of the group participants. Whereas the ISCTE participants had no professional experience, participants at Galp Energia had an average 14 years of work experience, which suggests that professional experience may influence the system thinking skills.

Other explanation could be that the Galp Energia participants did give more attention to the tasks than ISCTE students. It could be assumed that the ISCTE students were not motivated enough and hence they did not put much effort into working on the tasks, as they did not receive any incentives. It is possible that additional incentives in the form of grades or monetary payment would improve the results. However, the students in the other studies did not receive any incentive either. On the other hand, the Galp Energia participants consisted of a subset of a larger group of subjects that were invited to participate in this survey. It could be assumed that the participants that agreed to collaborate in the present survey were indeed motivated. Furthermore, the Galp Energia participants solved the problems in a familiar environment (their work place in company).

The results show that many mistakes made are not calculations errors. Participants did profoundly violate fundamental stock-and-flow relationships. Many participants violated the conservation of matter as they drew trajectories for the stock that were inconsistent with the net rate. This failure occurred in 77% in Task 1, 93% in Task 2 and 98% in Task 3 of the ISCTE participants; at Galp Energia it was exhibited in 45%, 87% and 81% respectively.

Pattern matching was one of the main errors. Many participants appear to believe that the stock trajectory should have the same qualitative shape as the net rate as they drew the stock by matching its pattern to the net flow. In Task 1, for example, more than 30% of the participants copied the inflow to draw the stock.

Other frequent failure was the “spreadsheet thinking” error (named by Sweeney and Sterman, 2000, p 267). Participants assumed that the stock changes suddenly at the end of every week, this is the flows are discrete, with inflows and outflows only occurring at the end of every period. This failure occurred in 20% in Task 1 and in 26% in Task 3 of the Galp Energia participants. Interestingly, none of the ISCTE’ participants exhibited this error. It suggests that due to their professional experience, the participants at Galp Energia tend to suffer more from “spreadsheet thinking” than the ISCTE’ participants.

Task 3 introduced the time delay and feedback concepts. Only 5% of the ISCTE participants considered the time delay correctly (61% at Galp Energia). More than 60% of the Galp Energia participants and 95% of the ISCTE participants failed on considering the feedback effect (production must overshoot orders to replenish the inventory lost and return to the equilibrium rate).

Task 4 (Department Store Task) emphasises on testing the participant intuition of the stock-flow concept, as it does not require participants to apply any mathematics but only to interpret the graph. Only 3% of the ISCTE participants and 35% of the Galp Energia participants succeeded in question 4, which is fundamental for testing the stock-flow concept. As these participants succeeded in questions 1 and 2, it can be stated that they did understand how to read graphs but failed to accumulate flows. Furthermore, as some of the participants did answer Task 1 correctly, one can conclude that these participants solved Task 1 by using graphical integration but indeed they did not intuitively understand the concept of stock-flow.

The sample from Galp Energia group is too small to look for significant influence of demographic data on the performance. Still, we performed a regression linear analysis relating task performance to the demographic variables. According to the results, overall performance did not appear to vary systematically with demographic variables (field in prior education, highest prior degree, age, gender or job level).

As we did not find that participants with more training and experience in mathematics or engineering performed better, it suggests that these skills did not translate into a better intuitive understanding of the systems thinking concepts.

8. Conclusions and Future Work

As it was pointed out by previous studies (for example :Sweeney and Sterman, 2000; Sterman, 2002), our experiences confirm a generally poor intuitive understanding of the basic system thinking, even when subjects have training or background in mathematics and sciences.

Participants interpreted the dynamic behaviour of stocks and flows relationships badly. It appears they had a weak intuitive sense of how a flow accumulates a stock over time. Some errors are systematic and violate basic system principles like conservation of matter and consistency between net flow rate and stock trajectories.

Overall performance did not appear to vary systematically with demographic variables (field in prior education, highest prior degree, age, gender or job level). However the senior professionals at Galp Energia performed better than the students of ISCTE, which suggests that the experience may influence the system thinking skills.

Improving manager ability to understand complex systems requires understanding of some basic system dynamics concepts such as feedback, stocks and flows, time delays and nonlinearity, and how to deal with them. These basic concepts are of great importance in real management tasks. For example, managers should have a feeling for flows and their related accumulations for understanding the impacts of decisions on the development of the business resources.

Assuming that reasoning is based on mental models that have analog structures to those of the external systems they imitate, strategic learning can be seen as a dynamic feedback process that emerges from the interaction of environment cues and the mind of the decision maker and results in a change of the decision maker's mental model.

In that sense, our future work will explore the influence of the structure of performance measurement systems in the formation of decision makers' mental models and how the quality of those mental models affect organizational decisions and performance in dynamically complex environments. To do this we will conduct a simulation-based research using a system dynamics-based micro world that capture the dynamic complexity of real management roles in a controlled environment.

9. References

ARMENIA, S., R. Onori and A.Bertini (2004). *Bathtub Dynamics at the Tor Vergata University in Rome, Italy*. Proceedings of the 22th International Conference of the System Dynamics Society, Oxford, England.

DIEHL, E and J. Sterman, (1995). *Effects of Feedback Complexity on Dynamic Decision Making*. Organizational Behaviour and Human Decision Process, 62 (2), pp 198-215.

FISHER, M. (2003). *Student Performance in The Bathtub and Cash Flow problems.* Proceedings of the 21th International Conference of the System Dynamics Society, New York, USA.

FORRESTER, Jay W., (1961). *Industrial Dynamics*. Pegasus Communications.

KAINZ, D. and G. OSSIMITZ (2002). *Can Students Learn Stock-Flow-Thinking ?: An Empirical Investigation.* Proceedings of the 20th International Conference of the System Dynamics Society, Palermo, Italy.

KAPMEIER, F. (2004). *Findings From Four Years of Bathtub Dynamics at Higher Education Institutions in Stuttgart.* Proceedings of the 22th International Conference of the System Dynamics Society, Oxford, England.

LYNEIS, J., D. Lyneis (2003). *Bahtub Dynamics at WPI.* Proceedings of the 21th International Conference of the System Dynamics Society, New York, USA.

OSSIMITZ, G. (2002). *Stock-flow thinking and reading stock-flow-related graphs: an empirical investigation in dynamics thinking abilities.* Proceedings of the 20th International Conference of the System Dynamics Society, Palermo, Italy.

PAICH, M. and J. Sterman, (1993). *Boom, Bust, and Failures to Learn in Experimental Markets.* Management Science, 39 (12), pp 1439-1458.

POWERSIM. Powersim Software AS. http://www.powersim.com/

STERMAN, J.D., (1989). *Modeling Managerial Behaviour: Misperceptions of Feedback in a Dynamic Decision Making Experiment.* Management Sciente, 35, 321-339.

STERMAN, J., (2000). *Business Dynamics: System Thinking and Modeling for a Complex World.* New York, Irwin Mcgraw-Hill.

STERMAN, J. (2002). *All models are wrong: reflections on becoming a systems scientist.* System Dynamics Review, 18 (4), 501-531.

SWEENEY, L. and J. Sterman (2000). *Bathtub Dynamics: Initial Results of a Systems Thinking Inventory.* System Dynamics Review, 16 (2), 249-294.

ZARARA, R. (2003). *Bahtub Dynamics in Portlans at SyMFEST.* Proceedings of the 21th International Conference of the System Dynamics Society, New York, USA.

Appendix A1 – Average performance of the ISCTE participants

	T1	T2	T3	T4
	0,00	0,13	0,00	0,50
	0,00	0,13	0,00	0,50
	0,00	0,13	0,00	0,50
	0,00	0,13	0,14	0,50
	0,00	0,13	0,00	0,50
	0,14	0,13	0,00	0,50
	0,71	0,13	0,14	0,50
	0,00	0,13	0,14	0,50
	1,00	1,00	0,43	0,50
	0,00	0,13	0,29	0,50
	0,00	0,13	0,29	0,50
	0,00	0,13	0,29	0,50
	0,00	0,13	0,29	0,50
	0,00	0,13	0,29	0,50
	0,00	0,13	0,29	0,50
	0,00	0,13	0,14	0,50
	0,71	0,13	0,00	0,50
	1,00	0,13	0,29	0,50
	0,00	0,13	0,29	0,50
	0,71	0,13	0,00	0,50
	1,00	0,13	0,00	0,50
	1,00	0,13	0,86	0,75
	1,00	1,00	1,00	1,00
	0,00	0,13	0,00	0,50
	0,14	0,13	0,14	0,50
	1,00	0,63	0,00	0,50
	1,00	0,75	0,00	0,50
	0,00	0,13	0,00	0,50
	0,00	0,13	0,00	0,50
	0,00	0,13	0,00	0,50
	0,00	0,13	0,00	0,25
	1,00	0,13	0,00	0,25
	0,00	0,13	0,00	0,50
	0,00	0,13	0,00	
	0,00	0,13	0,14	
	1,00	0,75	0,14	
	0,00	0,13	0,00	
	0,00	0,13	0,00	
	0,00	0,13	0,14	
	0,00	0,13	0,14	
	0,00	0,13	0,00	
	0,57	0,13	0,00	
	0,00	0,63	0,00	
Mean	0,28	0,22	0,14	0,51
St. Dev.	0,42	0,24	0,22	0,12

Appendix A2.1 – Performance Evaluation in Task 1 - ISCTE

C1	C2	C3	C4	C5	C6	C7	Mean
0	0	0	0	0	0	0	0,00
0	0	0	0	0	0	0	0,00
0	0	0	0	0	0	0	0,00
0	0	0	0	0	0	0	0,00
0	0	0	0	0	0	0	0,00
0	0	0	1	0	0	0	0,14
1	1	1	1	1	0	0	0,71
0	0	0	0	0	0	0	0,00
1	1	1	1	1	1	1	1,00
0	0	0	0	0	0	0	0,00
0	0	0	0	0	0	0	0,00
0	0	0	0	0	0	0	0,00
0	0	0	0	0	0	0	0,00
0	0	0	0	0	0	0	0,00
0	0	0	0	0	0	0	0,00
0	0	0	0	0	0	0	0,00
1	1	1	1	1	0	0	0,71
1	1	1	1	1	1	1	1,00
0	0	0	0	0	0	0	0,00
1	1	1	0	1	1	0	0,71
1	1	1	1	1	1	1	1,00
1	1	1	1	1	1	1	1,00
1	1	1	1	1	1	1	1,00
0	0	0	0	0	0	0	0,00
0	0	0	1	0	0	0	0,14
1	1	1	1	1	1	1	1,00
1	1	1	1	1	1	1	1,00
0	0	0	0	0	0	0	0,00
0	0	0	0	0	0	0	0,00
0	0	0	0	0	0	0	0,00
0	0	0	0	0	0	0	0,00
1	1	1	1	1	1	1	1,00
0	0	0	0	0	0	0	0,00
0	0	0	0	0	0	0	0,00
0	0	0	0	0	0	0	0,00
1	1	1	1	1	1	1	1,00
0	0	0	0	0	0	0	0,00
0	0	0	0	0	0	0	0,00
0	0	0	0	0	0	0	0,00
0	0	0	0	0	0	0	0,00
0	0	0	0	0	0	0	0,00
0	1	1	1	0	0	1	0,57
0	0	0	0	0	0	0	0,00

Mean	0,28	0,30	0,30	0,33	0,28	0,23	0,23	0,28
St. Dev.	0,45	0,46	0,46	0,47	0,45	0,43	0,43	0,42

Appendix A2.2 – Performance Evaluation in Task 2 - ISCTE

	C1	C2	C3	C4	C5	C6	C7	C8	Mean
	0	0	0	1	0	0	0	0	0,13
	0	0	0	1	0	0	0	0	0,13
	0	0	0	1	0	0	0	0	0,13
	0	0	0	1	0	0	0	0	0,13
	0	0	0	1	0	0	0	0	0,13
	0	0	0	1	0	0	0	0	0,13
	0	0	0	1	0	0	0	0	0,13
	0	0	0	1	0	0	0	0	0,13
	1	1	1	1	1	1	1	1	1,00
	0	0	0	1	0	0	0	0	0,13
	0	0	0	1	0	0	0	0	0,13
	0	0	0	1	0	0	0	0	0,13
	0	0	0	1	0	0	0	0	0,13
	0	0	0	1	0	0	0	0	0,13
	0	0	0	1	0	0	0	0	0,13
	0	0	0	1	0	0	0	0	0,13
	0	0	0	1	0	0	0	0	0,13
	0	0	0	1	0	0	0	0	0,13
	0	0	0	1	0	0	0	0	0,13
	0	0	0	1	0	0	0	0	0,13
	0	0	0	1	0	0	0	0	0,13
	0	0	0	1	0	0	0	0	0,13
	1	1	1	1	1	1	1	1	1,00
	0	0	0	1	0	0	0	0	0,13
	0	0	0	1	0	0	0	0	0,13
	1	1	1	1	0	0	0	1	0,63
	1	1	1	1	1	0	1	0	0,75
	0	0	0	1	0	0	0	0	0,13
	0	0	0	1	0	0	0	0	0,13
	0	0	0	1	0	0	0	0	0,13
	0	0	0	1	0	0	0	0	0,13
	0	0	0	1	0	0	0	0	0,13
	0	0	0	1	0	0	0	0	0,13
	0	0	0	1	0	0	0	0	0,13
	0	0	0	1	0	0	0	0	0,13
	1	1	1	1	1	0	1	0	0,75
	0	0	0	1	0	0	0	0	0,13
	0	0	0	1	0	0	0	0	0,13
	0	0	0	1	0	0	0	0	0,13
	0	0	0	1	0	0	0	0	0,13
	0	0	0	1	0	0	0	0	0,13
	0	0	0	1	0	0	0	0	0,13
	1	1	1	1	1	0	0	0	0,63
Mean	0,14	0,14	0,14	1,00	0,12	0,05	0,09	0,07	0,22
St. Dev.	0,35	0,35	0,35	0,00	0,32	0,21	0,29	0,26	0,24

Appendix A2.3 – Performance Evaluation in Task 3 - ISCTE

C1	C2	C3	C4	C5	C6	C7	Mean
0	0	0	0	0	0	0	0,00
0	0	0	0	0	0	0	0,00
0	0	0	0	0	0	0	0,00
0	0	0	0	1	0	0	0,14
0	0	0	0	0	0	0	0,00
0	0	0	0	0	0	0	0,00
0	0	0	0	1	0	0	0,14
0	0	0	0	1	0	0	0,14
1	0	0	0	1	1	0	0,43
1	0	0	0	1	0	0	0,29
1	0	0	0	1	0	0	0,29
1	0	0	0	1	0	0	0,29
1	0	0	0	1	0	0	0,29
1	0	0	0	1	0	0	0,29
0	0	0	0	1	1	0	0,29
1	0	0	0	0	0	0	0,14
0	0	0	0	0	0	0	0,00
0	0	0	0	1	1	0	0,29
0	0	0	0	1	1	0	0,29
0	0	0	0	0	0	0	0,00
0	0	0	0	0	0	0	0,00
1	1	1	1	1	1	0	0,86
1	1	1	1	1	1	1	1,00
0	0	0	0	0	0	0	0,00
0	0	0	0	1	0	0	0,14
0	0	0	0	0	0	0	0,00
0	0	0	0	0	0	0	0,00
0	0	0	0	0	0	0	0,00
0	0	0	0	0	0	0	0,00
0	0	0	0	0	0	0	0,00
0	0	0	0	0	0	0	0,00
0	0	0	0	0	0	0	0,00
0	0	0	0	0	0	0	0,00
0	0	0	0	0	0	0	0,00
0	0	0	0	1	0	0	0,14
1	0	0	0	0	0	0	0,14
0	0	0	0	0	0	0	0,00
0	0	0	0	0	0	0	0,00
0	0	0	0	1	0	0	0,14
0	0	0	0	1	0	0	0,14
0	0	0	0	0	0	0	0,00
0	0	0	0	0	0	0	0,00
0	0	0	0	0	0	0	0,00

Mean	0,23	0,05	0,05	0,05	0,42	0,14	0,02	0,14
St. Dev.	0,43	0,21	0,21	0,21	0,50	0,35	0,15	0,22

Appendix A2.4 – Performance Evaluation in Task 4 - ISCTE

C1	C2	C3	C4	Mean
1	1	0	0	0,50
1	1	0	0	0,50
1	1	0	0	0,50
1	1	0	0	0,50
1	1	0	0	0,50
1	1	0	0	0,50
1	1	0	0	0,50
1	1	0	0	0,50
1	1	0	0	0,50
1	1	0	0	0,50
0	1	1	0	0,50
1	1	0	0	0,50
1	1	0	0	0,50
1	1	0	0	0,50
1	1	0	0	0,50
1	1	0	0	0,50
1	1	0	0	0,50
1	1	0	0	0,50
1	1	0	0	0,50
1	1	0	0	0,50
1	1	0	0	0,50
1	1	1	0	0,75
1	1	1	1	1,00
1	1	0	0	0,50
1	1	0	0	0,50
1	0	1	0	0,50
1	1	0	0	0,50
1	1	0	0	0,50
1	1	0	0	0,50
1	1	0	0	0,50
1	0	0	0	0,25
1	0	0	0	0,25
1	1	0	0	0,50

Mean	0,97	0,91	0,12	0,03	0,51

St. Dev.	0,17	0,29	0,33	0,17	0,12

Appendix B1 - Participants Demographics/ Average Performance - Galp Energia

Gender	Age	Prior Field of Study	H.P. Degree	Job Level	T1	T2	T3	T4
M	25	Management/S.Sciences	5yd	1	0,86	0,88	0,86	0,75
F	26	Management/S.Sciences	5yd	1	0,00	0,88	0,71	0,50
M	27	Engineering/Science	5yd	1	1,00	0,63	0,57	0,50
F	27	Engineering/Science	5yd	1	0,00	0,88	0,00	0,50
F	27	Management/S.Sciences	MS	1	0,00	0,13	0,86	1,00
F	28	Management/S.Sciences	5yd	2	1,00	0,50	0,43	1,00
M	29	Engineering/Science	MS	2	1,00	0,63	0,00	0,50
M	29	Engineering/Science	5yd	1	0,86	0,63	1,00	1,00
M	29	Engineering/Science	5yd	1	0,00	0,50	1,00	1,00
M	29	Management/S.Sciences	5yd	2	0,00	0,75	0,57	0,75
M	31	Engineering/Science	MS	2	0,86	0,25	0,29	0,50
M	31	Management/S.Sciences	5yd	2	0,86	0,63	0,86	0,50
M	37	Management/S.Sciences	MS	3	1,00	0,63	1,00	1,00
M	39	Engineering/Science	5yd	2	1,00	0,50	0,57	0,75
M	39	Engineering/Science	5yd	2	0,71	0,13	0,57	0,50
M	39	Management/S.Sciences	5yd	1	1,00	0,25	0,14	0,50
M	41	Engineering/Science	MS	3	0,29	0,13	1,00	0,50
M	41	Management/S.Sciences	5yd	1	0,86	0,13	0,57	1,00
F	41	Management/S.Sciences	5yd	2	0,00	0,88	0,86	0,75
M	43	Engineering/Science	5yd	2	0,00	0,13	0,00	0,50
M	43	Management/S.Sciences	MS	2	0,00	0,13	0,57	1,00
F	45	Management/S.Sciences	5yd	2	0,00	0,13	0,00	0,50
M	47	Management/S.Sciences	5yd	2	1,00	0,38	0,57	1,00
M	48	Engineering/Science	5yd	3	1,00	0,13	0,00	0,50
M	51	Engineering/Science	5yd	2	0,00	0,13	0,57	0,50
M	51	Management/S.Sciences	5yd	3	1,00	0,50	0,00	0,75
M	53	Engineering/Science	5yd	3	0,71	0,13	0,00	0,50
M	54	Engineering/Science	5yd	2	1,00	0,88	0,71	1,00
M	54	Management/S.Sciences	MS	3	0,86	0,13	0,43	1,00
M	57	Engineering/Science	5yd	2	0,00	0,88	0,57	0,75
M	57	Engineering/Science	5yd	2	1,00	0,13	1,00	1,00

	T1	T2	T3	T4
Mean	0,58	0,44	0,53	0,73
St. Dev.	0,46	0,30	0,36	0,23

Appendix B2.1 – Performance Evaluation in Task 1 – Galp Energia

	C1	C2	C3	C4	C5	C6	C7	Mean
	1	1	1	1	1	1	1	1,00
	1	1	1	1	1	0	0	0,71
	0	0	0	0	0	0	0	0,00
	1	1	1	1	1	1	1	1,00
	0	0	0	0	0	0	0	0,00
	0	0	0	0	0	0	0	0,00
	1	1	1	1	1	1	1	1,00
	1	1	1	0	1	1	1	0,86
	1	1	1	0	1	1	1	0,86
	0	0	0	0	0	0	0	0,00
	1	1	1	1	1	1	1	1,00
	1	1	1	1	1	1	1	1,00
	0	0	0	0	0	0	0	0,00
	0	0	0	0	0	0	0	0,00
	1	1	1	1	1	0	0	0,71
	1	1	1	1	1	1	1	1,00
	1	1	1	1	1	1	1	1,00
	0	0	0	0	0	0	0	0,00
	1	1	1	1	1	1	1	1,00
	0	0	0	0	0	0	0	0,00
	1	1	1	1	1	1	1	1,00
	1	1	1	0	1	1	1	0,86
	0	0	0	0	0	0	0	0,00
	1	1	1	0	1	1	1	0,86
	0	0	0	0	0	0	0	0,00
	1	1	1	1	1	1	1	1,00
	1	0	0	1	0	0	0	0,29
	1	1	1	0	1	1	1	0,86
	1	1	1	0	1	1	1	0,86
	0	0	0	0	0	0	0	0,00
	1	1	1	1	1	1	1	1,00
Mean	0,65	0,61	0,61	0,45	0,61	0,55	0,55	0,58
St. Dev.	0,49	0,50	0,50	0,51	0,50	0,51	0,51	0,46

Appendix B2.2 – Performance Evaluation in Task 2 – Galp Energia

	C1	C2	C3	C4	C5	C6	C7	C8	Mean
	1	1	1	1	0	0	0	0	0,50
	0	0	0	1	0	0	0	0	0,13
	1	1	1	1	1	1	1	0	0,88
	1	1	1	1	0	0	0	1	0,63
	0	0	0	1	0	0	0	0	0,13
	1	1	1	1	1	0	1	0	0,75
	0	1	1	1	0	0	0	0	0,38
	1	1	1	1	0	0	0	1	0,63
	1	1	1	1	0	0	0	1	0,63
	1	1	1	1	0	0	0	0	0,50
	1	1	1	1	0	0	0	0	0,50
	1	0	0	1	0	0	0	0	0,25
	0	0	0	1	0	0	0	0	0,13
	0	0	0	1	0	0	0	0	0,13
	0	0	0	1	0	0	0	0	0,13
	1	1	1	1	1	1	1	0	0,88
	0	0	0	1	0	0	0	0	0,13
	1	1	1	1	1	1	1	0	0,88
	1	1	1	1	0	0	0	1	0,63
	1	1	1	1	1	1	1	0	0,88
	1	1	1	1	1	0	0	0	0,63
	0	0	0	1	0	0	0	0	0,13
	0	0	0	1	0	0	0	0	0,13
	1	0	0	1	0	0	0	0	0,25
	0	0	0	1	0	0	0	0	0,13
	0	0	0	1	0	0	0	0	0,13
	0	0	0	1	0	0	0	0	0,13
	1	1	1	1	1	1	1	0	0,88
	0	0	0	1	0	0	0	0	0,13
	1	1	1	1	1	1	1	0	0,88
	1	1	1	1	0	0	0	0	0,50
Mean	0,58	0,55	0,55	1,00	0,26	0,19	0,23	0,13	0,44
St. Dev.	0,50	0,51	0,51	0,00	0,44	0,40	0,43	0,34	0,30

Appendix B2.3 – Performance Evaluation in Task 3 – Galp Energia

	C1	C2	C3	C4	C5	C6	C7	Mean
	1	1	0	0	1	1	0	0,57
	1	1	0	0	1	1	0	0,57
	0	0	0	0	0	0	0	0,00
	1	1	0	0	1	1	0	0,57
	1	1	0	0	1	1	0	0,57
	1	1	0	0	1	1	0	0,57
	1	1	0	0	1	1	0	0,57
	1	1	1	1	1	1	0	0,86
	1	1	1	1	1	1	1	1,00
	1	1	1	1	1	1	1	1,00
	0	0	0	0	0	0	0	0,00
	1	0	0	0	0	0	0	0,14
	1	1	0	0	1	0	1	0,57
	0	0	0	0	0	0	0	0,00
	0	0	0	0	0	0	0	0,00
	1	1	1	0	1	1	0	0,71
	0	0	0	0	0	0	0	0,00
	1	1	1	1	1	1	0	0,86
	1	1	1	1	1	1	1	1,00
	1	1	0	0	1	1	0	0,57
	0	0	0	0	0	0	0	0,00
	0	0	1	1	1	1	0	0,57
	0	0	0	0	0	0	0	0,00
	0	0	0	0	1	1	0	0,29
	1	1	1	1	1	1	0	0,86
	1	1	1	1	1	1	1	1,00
	1	1	1	1	1	1	1	1,00
	1	1	1	1	1	1	0	0,86
	1	0	0	0	1	1	0	0,43
	1	1	1	0	1	1	0	0,71
	1	0	0	0	1	1	0	0,43
Mean	0,71	0,61	0,39	0,32	0,74	0,71	0,19	0,53
St. Dev.	0,46	0,50	0,50	0,48	0,44	0,46	0,40	0,36

Appendix B2.4 – Performance Evaluation in Task 4 – Galp Energia

C1	C2	C3	C4	Mean
1	1	1	0	0,75
1	1	0	0	0,50
1	1	0	0	0,50
1	1	0	0	0,50
1	1	0	0	0,50
1	1	1	0	0,75
1	1	1	1	1,00
1	1	0	0	0,50
1	1	1	1	1,00
1	1	1	1	1,00
1	1	1	0	0,75
1	1	0	0	0,50
1	1	1	1	1,00
1	1	0	0	0,50
1	1	0	0	0,50
1	1	1	1	1,00
1	1	0	0	0,50
1	1	1	0	0,75
1	1	1	1	1,00
1	1	1	0	0,75
1	1	0	0	0,50
1	1	1	1	1,00
1	1	0	0	0,50
1	1	0	0	0,50
1	1	1	1	1,00
1	1	1	1	1,00
1	1	0	0	0,50
1	1	1	0	0,75
1	1	1	1	1,00
1	1	0	0	0,50
1	1	1	1	1,00

Mean	1,00	1,00	0,55	0,35	0,73
St. Dev.	0,00	0,00	0,51	0,49	0,23