Systems thinking: learning or driving?

CARLOS CAPELO E JOÃO FERREIRA DIAS

ABSTRACT: The ability to foresee how the behaviour of a system results from the interaction of its components over time - i.e., its dynamic complexity - is considered important in order to take effective decisions in our turbulent world. Dynamic complexity frequently emerges from interrelated simple structures, such as stocks and flows, feedbacks and delays (Forrester, 1961). Common sense assumes an intuitive understanding of their dynamic behaviour. However, recent research has pointed to a persistent and systematic error in people's understanding of those building blocks of complex systems.

This paper describes an empirical study on the innate ability to understand systems thinking concepts. Two different groups - one academic, and the other professional - undertook four tasks proposed by Sweeney and Sterman (2000) and Sterman (2002). The results confirm a poor intuitive understanding of the basic systems concepts, even when subjects have a background in mathematics and sciences.

Key words: Management Learning, Systems Thinking, System Dynamics

TÍTULO: Pensamento sistémico: mitos e factos

RESUMO: Num mundo turbulento, os gestores devem ser capazes de antecipar o comportamento dinâmico de sistemas socioeconómicos complexos para tomarem atempadamente as decisões adequadas. Ora, a complexidade dinâmica emerge frequentemente de interacções simples entre os componentes dum sistema.

Na linha de pesquisas que estão a ser feitas em diversos países, os autores analisaram a capacidade, quer de jovens estudantes, quer de gestores seniores, intuírem correctamente os comportamentos dinâmicos de variáveis integrantes de estruturas sistémicas muito simples. Para tal utilizaram os testes propostos por Sweeney and Sterman (2000) and Sterman (2002).

Os resultados confirmaram uma fraca compreensão da complexidade dinâmica, mesmo entre os que têm competência nas áreas da matemática e das ciências.

Palavras-chave: Ensino da Gestão, Pensamento Sistémico, Dinâmica de Sistemas

CARLOS CAPELO
carlos.capelo@galpenergia.com
Director at GALP SA.
Director na GALP SA.

JOÃO FERREIRA DIAS
fdias@iscte.pt
PhD in Strategic Management. Associate Professor at ISCTE and researcher at ADETTI/ISCTE.
Doutor em Gestão Estratégica. Professor Associado do ISCTE e investigador da ADETTI/ISCTE.

INTRODUCTION

In a world that changes at an ever-faster pace, managing system dynamic complexity is a significant challenge in every sector. Therefore, individuals with system thinking skills should be better prepared to deal with current problems (Senge, 1990; Sterman, 2000).

There is no clear and rigorous definition of systems thinking. However, there is consensus that it involves the ability to: use a holistic perspective; reveal a system's structure recognizing delays and feedback loops; model the system's complexity and understand how its behaviour emerges from the interaction of its components over time.

The emphasis on the ability to represent, assess and manage dynamic complexity comes from Jay Forrester who used servomechanism theory in the 1950s to deal with social problems. In his seminal book, Forrester (1961) pointed out that in real life the cause is usually remote from the effect in distance as well as in time. They are often parts of a circular relation where it is impossible to differentiate the roles of cause and effect. The understanding of feedback loops requires a holistic and dynamic perspective in order to see how things emerge from the whole over time.

Forrester's system dynamics, founded on the Von Bertalanfy's system paradigm, postulates that: (i) the behaviour of a system emerges from its structure (ii) which must be formally described to support computer simulation. Nowadays, the isomorphism rule (i.e., systems with equal structures have similar behaviours) has a practical use e.g. to get insights into the market penetration of an innovative product from the structural model of disease diffusion.

The requirement of rigorous system modelling and the use of computer simulation to foresee the behaviour of a complex system is not so widely accepted however. Forrester believes the human mind has processing limitations to deal with dynamic complexity. Following this idea of bounded rationality, Sterman (1989) sees two types of cognitive limitation: first, the mental models people use to represent dynamic complexity are deficient; second, even people who form adequate mental models are unable to infer correctly the dynamic behaviour of the system.

This handicap has important practical consequences. According to Piaget's constructivism theory, people develop conceptual models (schemas) as a method to understand the world. From this viewpoint, we can only think in accordance with our views of the world, which are mental models of the world itself. Good decisions cannot be expected if people wrongly infer dynamics from their mental models.

EMPIRICAL RESEARCH IN SYSTEMS THINKING

The widespread idea that educators must develop systems thinking skills has triggered a great development of bibliography, computer-aid system building simulation software and training courses. Unfortunately, this has not been followed by research to understand people's innate abilities in systems thinking and the effectiveness of the proposed learning methods.

System dynamic complexity frequently emerges from interrelated simple structures and interrelations, such as feedback, stocks and flows, and time delays. The improvement of the decision-maker's ability to manage complex systems therefore requires the prior understanding of those simple system concepts. Common sense assumes an intuitive understanding of their dynamic behaviour. However, experimental studies have found that decision makers are very susceptible to the dynamic complexity of the decision context (Sterman, 1989; Paich & Sterman, 1993; Diehl & Sterman, 1995, Ossimitz, 2002; Sterman, 2002).

Sweeney and Sterman (2000) developed some tasks to measure the subjects' ability to deal with the most basic system dynamic concepts: stocks and flows, time delays, and feedback. The sample comprised students at the MIT Sloan School of Management, with an unusually strong background in mathematics and sciences. It assumes that the performance of the general public is worse. The results showed a poor performance.

Since then, similar research has been conducted at other schools and universities including Tor Vergata University in Rome (Armenia et al., 2004), Stuttgart Institute of Management and Technology (Kapmeier, 2004) and again MIT (Sterman, 2002). Despite some difference, the results have consistently shown that people have a poor understanding of basic system concepts and made persistent, systematic mistakes in their understanding of the dynamic behaviour of those basic building blocks.

The aim of our research is similar to those studies, but we include professionals in their workplace as well as university students.

TASKS, PARTICIPANTS AND PROCEDURES

Tasks

In order to compare the results with similar studies, we use the "Bathtub", "Cash Flow" and "Manufacturing" tasks (Sweeney & Sterman, 2000) and "Department Store" task (Sterman, 2002), that we renamed as Task 1 to Task 4, respectively.

Figures 1 to 4 show the four tasks and their solutions. We use the same criteria proposed by their authors to evaluate subjects' performance. These are set out in Table I to Table 4 for Task 1 to Task 4 respectively.

Task 1 and Task 2 (Figure 1 and 2) explore the understanding of the stock-flow relationship. In these tasks, one stock (accumulator), already with an initial amount, is filled through an inflow and simultaneously emptied through an outflow. Participants must draw the expected behaviour of stock over time.

There are neither feedbacks nor time delays. There is only one stock. The outflow is constant and the inflows follow simple patterns: square wave pattern for Task 1; saw tooth pattern for Task 2. Round numbers are used, so it is easy to calculate the net flow and quantity added to (or removed from) the stock.

FIGURE 1

Task 1 problem and solution

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

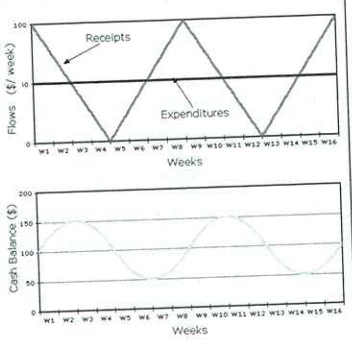
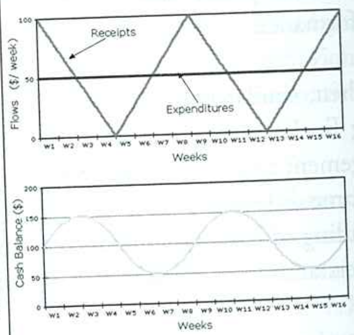


FIGURE 2

Task 1 problem and solution

Question: Consider the cash balance of a company. Receipts flow in the balance at a certain rate, and expenditures flow out at another rate. From that information, draw the behavior of the firm's cash balance.

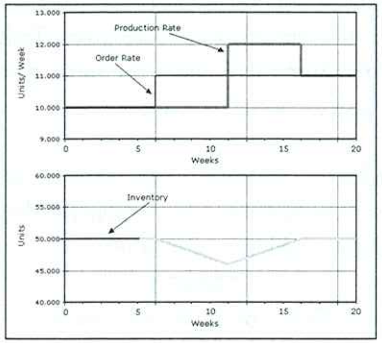


Task 3 differs from the previous tasks as it includes both a negative feedback loop and a time delay. The objective is to keep the inventory stable by regulating the inflow to the stock in a negative feedback loop. Here the stock manager faces a realistic lag between a change in production schedule and the production rate. There is no single correct answer to this task. As in the answer presented, the amount of production above customer orders must equal the quantity of inventory lost during the period when orders exceed production in all possible solutions.

FIGURE 3

Task 3 problem and solution

Question: Consider a manufacturing firm. It maintains an inventory of finished product to fill customers orders, that have averaged 10 000 units/week. The firms strives to maintain an inventory of 50 000 units to provide excellent customer service. Although it has overcapacity, it takes 4 weeks to adjust the production schedule and to make the product. Now imagine that in week 5 the order rate for unexpectedly rises by 10% and remains at the new higher rate indefinitely. 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.



Task 4 was created to eliminate the argument that people performed poorly on Tasks 1 and 2 because they do not know how to do graphical integration (Sterman, 2002). Task 4 simply shows an inflow and an outflow with the number of people entering and leaving a department store per minute over a certain period. 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?»).

Participants

Two different samples were used to conduct our research on the subjects' understanding of some basic system dynamics concepts: an academic sample from ISCTE (a business graduate school in Lisbon) and a professional sample from Galp Energia, one of the biggest Portuguese firms (the Portuguese oil company).

The ISCTE group comprised 43 undergraduate students in their last semester of a 5-year business degree. The group was homogeneous and their ages ranged from 21 to 25.

The Galp Energia group consisted of 31 people, 25 men and 6 women. Their ages varied from under 30 (10 participants), from 30 to 50 (14) and over 50 (7). All have at least a 5-year degree (seven also had an MBA/MSc degree) with background in engineering or maths (16) or in management/social sciences (15). In terms of organizational position, 9 were 1st level directors or top management senior staff, 16 were 2nd level directors or senior experts, and the others worked at the 3rd hierarchical level.

Procedure

At ISCTE, 43 students did Tasks 1, 2 and 3 in a first session of about 45 minutes. In a second session of 15 minutes, 33 (from the original 43) did Task 4. No rewards were given.

At Galp Energia, 31 participants did the tasks in two sessions at their own work place. In the first session, the participants were given Task 1 and Task 2 – demographics information forms. Some days later, the same participants received Task 3 and Task 4. Each session lasted about 30 minutes.

The objective of the Tasks, i.e., to gain insight into people understanding of some systems thinking concepts, was explained to participants before they began and they were asked not to discuss the problems with others.

TABLE 1

Task 1 performance results

Criteria	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 a nd 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 is the area enclosed by the net rate	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

Note: Comparison with similar experiments on TV=Tor Vergata University in Rome (Armenia et al., 2004), SIMT=Stuttgart Institute of Management and Technology (Kapmeier, 2004) and MIT (Sweeney & Sterman, 2000; Sterman, 2002)

RESULTS

Task 1 results

Table 1 shows the performance results for Task 1. Both groups performed poorly, on average (28% at ISCTE and 58% at Galp Energia) worse than the international groups of TV, SIMT and MIT. However, Galp Energia performance is much better than ISCTE on every item.

The first 3 items are fundamental to the concept of stock accumulation. Whereas more than 70% of ISCTE participants and more than 35% of Galp Energia participants failed these, this was only true of 20% at MIT and 32% at SIMT. We note that all participants who correctly answered items 1 and 2, also gave the right answer to item 3.

Item 4 was Galp Energia participants' worst item. Many of them (20% of total) assumed flows as discrete, with inflows and outflows only occurring at the end of every period. They drew a sudden change in stock trajectory at the end of every week. Sweeney and Sterman (2000) reported the same error and named it "spreadsheet thinking". None of the ISCTE participants exhibited «spreadsheet thinking». This might be explained by a professional bias due to the daily use of spreadsheets.

In the most typical mistake (exhibited by 35% of ISCTE and 29% of Galp Energia), the stock trajectory matches the net inflow pattern, jumping up and down in phases.

The mistakes seem to arise not from a misunderstanding of stock-flow relationship rather than numerical errors. In fact, participants did not understand the basic relationship between a net flow and its impact on stock accumulation and, in particular, that the change in the stock over a period is the area bound by the net rate in that period.

Task 2

Table 2 shows the performance results for Task 2. As it is more difficult than Task 1, it is not surprising that the performance of both groups was worse. A comparison with the other international surveys shows ISCTE to have the worst performance, while that of Galp Energia is average.

As in Task 1, the first 3 items are fundamental to the concept of stock accumulation. Few ISCTE participants (14%) answered these items correctly whereas at Galp Energia about 45% of the participants gave the incorrect answer. Once again, all the participants that correctly answered items 1 and 2 also gave the right answer to item 3.

TABLE 2

Task 2 performance results

Criteria	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	0.12	0.26	0.35	0.23	0.28
C6a- The slope of the stock is 50 units/period when the net rate is at its maximum (t= 0, 8, 16)	0.05	0.19	0.21	0.23	0.47
C6b- The slope of the stock is -50 units/period when the net rate is at its minimum (t= 4, 12)	0.09	0.23	0.20	0.14	0.45
C7- The quantity added to (removed from) the stock during each segment is the area enclosed by the net rate	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

Note: Comparison with similar experiments on TV=Tor Vergata University in Rome (Armenia et al., 2004), SIMT=Stuttgart Institute of Management and Technology (Kapmeier, 2004) and MIT (Sweeney & Sterman, 2000; Sterman, 2002)

In an improvement on Task 1, all participants from both groups answered item 4 correctly. Unlike Task 1, the net flow rate in Task 2 is continuous which suggests that many participants failed in Task 1 because they drew stock trajectories that matched the pattern of the net flow rate.

Items 5 to 7 describe the trajectory of the stock, requiring participants to read and interpret the graph showing 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 stock and be able to construct a line for the stock trajectory that shows the right slope of the stock at the inflection points. Less than 10% of the ISCTE and 20% of the Galp Energia participants answered this correctly.

As in Task 1, most of the participants (28% at ISCTE and 20% at Galp Energia) drew the stock trajectory by matching its pattern to the net inflow. Most of Galp Energia participants (23%) drew the stock trajectory linearly, even though they answered items 1 to 4 correctly.

Task 3

Table 3 shows the performance results for Task 3. Even though this 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. Average performance was poor: 53% for Galp Energia and only 14% for ISCTE.

TABLE 3

Task 3 performance results

Criteria	ISCTE	GALP	TV	SIMT	MIT
C1- Production must start in equilibrium with orders	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 [RE1]increase in orders	0.05	0.61	0.63	0.36	0.44
C3- Production must overshoot orders to replenish the inventory lost. Production should return to the equilibrium rate of 11,000 units/week	0.05	0.39	0.46	0.23	0.44
C4- The area enclosed by production and orders during the overshoot of production must equal the area enclosed by orders minus production	0.05	0.32	0.37	0.09	0.11
C5- Inventory must initially decline	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

Note: Comparison with similar experiments on TV=Tor Vergata University in Rome (Armenia et al., 2004), SIMT=Stuttgart Institute of Management and Technology (Kapmeier, 2004) and MIT (Sweeney & Sterman, 2000; Sterman, 2002)

The average performance of ISCTE was worse than in 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 international surveys, the ISCTE group had the worst performance and that of Galp Energia was average.

Only 2% of the ISCTE participants and 16% of the Galp Energia participants solved all items in this problem correctly. ISCTE students failed in most items from the departure point of production (80% of ISCTE participants), passing by the production adjustment delay (95%) and the conservation principle (95%) to the inventory path (98%). Galp Energia did better but failed in item 7 demonstrating that most of its participants did not relate the flows to the stock correctly.

If we analyse the common mistakes, we find that, when drawing the inventory by 26% of Galp Energia and 9% of ISCTE first registered a decline during the lag, then a rise back to equilibrium, even though production equalled orders.

Many ISCTE participants drew the trajectory for inventory as just matching the production path. Another mistake made was to show the 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 made this mistake in Tasks 1 and 2. At Galp Energia 26% of participants showed "spreadsheet thinking".

Task 4

Table 4 shows the performance results for Task 4. The simplicity of Task 4 makes it a good test of the intuitive understanding of the stock-flow relationship.

All participants from Galp Energia and most from ISCTE gave the correct answer to when the most people enter and leave the store (Q1=4, Q=21), showing that they read and interpret graphs properly.

To answer question 3, participants had to look at the graph and see that until minute 13 the inflow exceeds the outflow and so that the number of people in the store increases until minute 13, subsequently decreasing 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 answered this correctly.

To answer question 4, participants must judge whether more people enter until minute 13 than leave after, i.e. 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 is larger than the area through to minute 13, the fewest people

TABLE 4
Task 4 performance results

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

Note: Comparison with similar experiments on TV=Tor Vergata University in Rome (Armenia et al., 2004), SIMT=Stuttgart Institute of Management and Technology (Kapmeier, 2004) and MIT (Sweeney & Sterman, 2000; Sterman, 2002)

are in the store at minute 30 (question 4). Only 3% of the ISCTE participants and 35% of the Galp Energia participants successfully answered this.

Nearly 40% of the ISCTE participants and 30% of the Galp Energia participants gave the wrong answer 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 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. Interestingly, nearly 50% of those Galp Energia participants answered Task 1 correctly. It suggests that those participants solved Task 1 by using graphical integration, but Task 4 revealed they did not have an intuitive understanding of the concept of stock-flow relationship.

Correlation of Task Results

This section analyses the correlation between the Task results (mean for all items in each Task) for the two groups of participants – ISCTE and Galp Energia. Unfortunately, ISCTE participants that did Task 4 were not correctly identified so it was impossible to correlate these results with Tasks 1 to 3.

Table 5 shows the correlation (Pearson) matrix for the regression analysis between Task results. It indicates the relation between the expected performance in one Task and the performance in the others.

For ISCTE, Tasks 1 to 3 appear to correlate. This is not the case for Galp Energia except for the Task 3 – Task 4 pair. As Tasks 1, 2, and 4 focus on the accumulation

TABLE 5
Correlation analysis of task results

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

*Significant at the level: ** ?<0.01; * ?<0.05*

process of a single stock through an inflow and an outflow, we expected to see significant correlations between them. As stated above, this suggests that some participants answered Task 1 by using graphical integration correctly but Task 4 showed they did not really understand the concept of stock accumulation.

Impacts of Demographic variables

The Galp Energia participants filled out a form about their gender, age, highest prior degree, prior field of study, and level of job. Table 6 shows the mean and standard deviation Task results layered by demographic variables.

To test the hypotheses that demographic variables would affect performance, we ran a variety of regression linear models relating performance to the various demographic variables reported by participants. First, we used regression models relating performance to all demographic variables. After, we refined the regression models by performing a stepwise regression procedure in order to maximize R^2 adjusted, by only keeping the most explanatory variables (Table 7). We used standardized values for all variables to compare the relative effect of each one on the performance.

Although it is possible to observe some impacts of demographics on performance in some Tasks, there is no consistent pattern. Performance appears to depend on the participants' gender in Task 1, age and highest prior degree in the Task 2 and prior field of study in Task 4. None of the demographic variables influences the performance in the most difficult Task, the Task 3.

TABLE 6
Demographic impacts for Galp Energia Group

		Task 1		Task 2		Task 3		Task 4	
Gender	N	?	?	?	?	?	?	?	?
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

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 from social sciences. In the present study, we found the opposite and only for Task 4.

As no variable was found to have a consistent impact on the performance in all Tasks, it is clear that the results from the present study do not provide full support for the hypotheses that demographic variables affect participants' performance.

TABLE 7

Regression relating task performance to Demographics

	Task 1		Task 2		Task 3		Task 4	
	Standardized	Significance	Standardized	Significance	Standardized	Significance	Standardized	Significance
Variable	?	?	?	?	?	?	?	?
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	
?	0.012		0.021		>0.2		0.077	

*Significant at the level: ** ?<0.05; * ?<0.1*

DISCUSSION

No special skills are required to answer the Tasks other than the capacity to read and interpret graphs (Task 4), do simple arithmetic (Task 1 and 2) and geometric (Task 3) calculus. Furthermore, as the Tasks use current situations, unfamiliarity is not a reason for failure.

Table 8 shows the performance for both groups according to the system thinking concepts considered in each task.

ISCTE participants showed an overall poor understanding of all system concepts, beginning with the simple stock and flow relationship (only 3% answered Task 4 correctly). Not surprisingly, ISCTE had very bad results for the items related to delay and feedbacks concepts. Compared with all others groups, the ISCTE participants got the worst results (see Table 9).

Galp Energia participants understood the system concepts much better (see Table 8). Interestingly, the simple stock and flow relationship seems more difficult to understand than the impacts of feedbacks or delays. In fact, their average score in Task 3 (see Table 9) is better than those of SIMT and MIT. When we look at the evolution of the scores, we find that Galp Energia participants show a progressive improvement. Task 1 was their worst, Tasks 2 and 3 were average and Task 4 was the best. Since they did the tasks in two sessions (first Task 1 and Task 2 and some days later Task 3 and Task 4), perhaps the participants learnt from session 1 and started session 2 with a better understanding of the dynamic concepts.

TABLE 8

Task 2 performance on system thinking concepts

System Thinking Concepts	ISCTE	GALP
Task 1: Stock-flow with linear stock trajectory (a)	21%	35%
Task 2: Stock-flow with nonlinear stock trajectory (a)	5%	0%
Task 3: Delay (b)	5%	61%
Feedback (b)	5%	39%
Total (a)	2%	16%
Task 4: Stock-flow (a)	3%	35%

(a) Percentage of the participan ts with correct answers in all criteria
(b) Percentage of the participants with correct answers in criteria regarding to system thinking co ncept

TABLE 9

Task 2 overall performance in all tasks

	ISCTE		GALP		TV		SIMT		MIT	
	?	?	?	?	?	?	?	?	?	?
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

What explains the difference in performance between ISCTE and Galp Energia groups? One hypothesis is the marked difference in professional experience. Whereas the ISCTE participants had no professional experience, participants at Galp Energia had an average 14 years' work experience.

Another explanation is motivation. Perhaps ISCTE students were not motivated enough as they did not receive any incentives. However, the students in the other studies (TV, SIMT, MIT) did not receive any incentive either. On the other hand, the Galp Energia participants consisted of a subset of a larger group of subjects invited to participate in this survey. Participants that agreed to collaborate were indeed motivated. Furthermore, the Galp Energia participants solved the problems in a familiar environment (their work place in company).

CONCLUSIONS

As seen in previous studies, our experiences confirm a generally poor intuitive understanding of the basic system thinking, even when subjects have training or a background in mathematics and sciences.

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

Pattern matching was one of the main mistakes. Many participants simply drew the stock by matching its pattern to the net flow. Another curious failure at Galp Energia was the "spreadsheet thinking" error – the stock changed suddenly at the end of every period. This mistake seems to be of a professional nature, since it was not made by any of the ISCTE participants.

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

Our results emphasize the Forrester feeling (Forrester, 1971) that we are faced with a human processing limitation. Recent research conducted by Pala and Vennix (2005) on the (weak) effectiveness of system thinking learning seems to point in the same direction. To overcome this handicap, we may need computer assistance. After all, humans do not have wings but they fly aeroplanes!

REFERENCES

ARMENIA, S., ONORI, R. & BERTINI, A. (2004), «Bathtub Dynamics at the Tor Vergata University in Rome, Italy». Proceedings of the 22nd International Conference of the System Dynamics Society, Oxford, England.

DIEHL, E. & STERMAN, J. (1995), «Effects of feedback complexity on dynamic decision making». *Organizational Behaviour and Human Decision Process*, vol. 62(2), pp. 198-215.

FORRESTER, J. (1961), **Industrial Dynamics**. Pegasus Communications, Waltham MA.

FORRESTER, J. (1971), «Counterintuitive behaviour of social systems». *Technology Review*, vol. 73(3), pp. 53-68.

KAINZ, D. & OSSIMITZ, G. (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 22nd International Conference of the System Dynamics Society, Oxford, England.

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. & STERMAN, J. (1993), «Boom, bust, and failures to learn in experimental markets». *Management Science*, vol. 39(12), pp. 1439-1458.

PALA, O. & VENNIX, J. (2005), «Effect of system dynamics education on systems thinking inventory task performance». *System Dynamics Review*, vol. 21(2), pp. 147-172.

SENGE, P. (1990), **The Fifth Discipline: The Art and Practice of the Learning Organization**. Doubleday, New York.

STERMAN, J. (1989), «Modelling managerial behaviour: misperceptions of feedback in a dynamic decision making experiment». *Management Science*, vol. (35), pp. 321-339.

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

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

SWEENEY, L. & STERMAN, J. (2000), «Bathtub dynamics: initial results of a systems thinking inventory». *System Dynamics Review*, vol. 16(2), pp. 249-294.