Comparing Systems Thinking Inventory Task Performance in German Classrooms at High School and University Level

Hans Dieter Kasperidus (corresponding author)
UFZ Centre for Environmental Research Leipzig-Halle
Department Urban Ecology, Environmental Planning and Transport
Permoserstr. 15, 04318 Leipzig, Germany
Tel: 0049 341 235 2601, Fax: 0049 341 235 2534
Email: hans.kasperidus@ufz.de

Hanno Langfelder
MSc. Sustainable Resource Management
Klobensteinerstr. 27, 81545 München, Germany
Tel: 0049 89 642 70 359, Fax: 0049 89 642 70 358
Email: hanno.langfelder@greencity.de

Peter Biber
Technische Universität München
Chair of Forest Yield Science
Am Hochanger 13, 85354 Freising, Germany
Tel: 0049 8161 714708, Fax: 0049 8161 714721
Email: peter.biber@lrz.tu-muenchen.de

Abstract:

Since the introduction of systems thinking inventory tasks or "Bathtub Dynamics" tasks in 2000, an increasing number of researchers and educators have confronted various student groups, such as university students, high school students and post-graduates with these tests to assess and evaluate their understanding of basic system thinking skills and system concepts. The results showed a surprisingly poor performance. This motivated us to subjecting school and university students in Germany to the Bathtub Dynamics tests so as to broaden the international research data and to gain a better understanding of the current level of system thinking skills in selected student groups at our educational institutions. The tests groups consisted of two university and one high school group. All participants received the same tasks. The performance results were compared to each other and to other research studies that applied the same tests. The task results were assessed to discover any correlation with the subject's demographics, such as age, gender, previous degrees or university courses. As with other studies the overall performance was poor and below the level of education of the participants. The interesting aspects are the specific differences in performance between and within the individual groups and the relationship to demographic factors such as gender.

Keywords: K-12-education, university education, system thinking inventory, Bathtub Dynamics, systems thinking skills

1. Introduction

The environment is a classic example of an area where an ignorance of the dynamics and complexity of systems will inevitably lead to problems. "I was already convinced that most of the problems faced by humankind concerned our inability to grasp and manage the increasingly complex systems of our world", writes Senge (1990) in his book on learning organisations. As history demonstrated, the consequences of our ignorance and inability to act within a system's boundaries can have devastating effects. In his book *The Logic of Failure,* Dörner (2004) points out, that the Chernobyl reactor disaster had a purely human cause, because the responsible engineers – highly trained people – failed to incorporate the dynamism of the reactor into their actions and operational plans. They failed to consider delayed feedback process in response to a change to the system, which is not accounted for in the original action plan. The result was the worst civilian nuclear disaster in the world with horrendous impacts and costs to people and to the environment, not only in the country of the disaster but also in remote locations. A better understanding of system behaviour may be one possibility of overcoming such mistakes in many real life situations. The ability to understanding and handle systems is the prerequisite for such a debate.

Booth Sweeney and Sterman (2000) list basic abilities, which, in combination with scientific rational and the use of quantitative and qualitative data, can be turned into effective systems thinking. Specific skills include the ability to:

- Understand how the behaviour of a system arises from the interaction of its agents over time (= dynamic complexity);
- Discover and represent feedback processes (negative and positive) hypothesised to underlie observed patterns of system behaviour;
- Identify stock and flow relationships;
- Recognise nonlinearities;
- Recognise and challenge the boundaries of mental and formal models.

System dynamics on the other hand draws on a multitude of disciplines; applied mathematics, philosophy, social science and engineering. And is in fact "all these things and much more" (Sterman 1994). System dynamics has its beginning in control theory and nonlinear dynamics, and can be used by policy-makers as a practical tool for solving problems as well as being a method for describing real world systems through modelling (Sterman 1994). Forrester (1961) sees a clear distinction between systems thinking and system dynamics; on the one hand he sees systems thinking as a phrase for describing and talking about systems in general, but does not cover the reality of system dynamics by lacking in quantitative and dynamic analysis. Whereas system dynamics is more about the structuring and simulation of system models and it is the only way to work on deeply imbedded mental models. Sterman (2002) sees system dynamics "as being a practical tool for finding solutions to critical policy problems, as well as being a philosophy, a social science, an engineering subject, mathematics and a form of action and consulting". System dynamics is a descriptive language used for studying the structures of social, economic, organisational and ecological systems and their behaviour and the interaction of their components with each other over time. A great asset is its applicability to such a wide range of issues.

If systems thinking and system dynamics are promising methods for understanding complex real world problems and finding effective solutions for them, then it would be interesting to know how well these skills are developed in a population. The effects of poor or non-existent systems thinking skills have been described on a number of occasions. Even highly educated people and entire institutions and organisations still fail to take account of system principles (Dörner 2004, Booth Sweeney and Sterman 2000, Diamond 2005, Duffield 2004). Learning and using systems thinking and system dynamics is a matter of education, training and practice, but also developing an intuitive feeling for system behaviour. If more people are to use system dynamics they need to learn and understand the concept and methods behind it and have the continuous opportunity to apply them. Schools and universities are likely starting points. In Western societies, they are the primary institutions for the directed education of people. Within an educational context, systems thinking produce or encourage important thinking skills (Booth Sweeney and Sterman 2000).

Schools and universities can be useful for conducting research into assessing the systems thinking ability of students. A lack of data and insight into this ability led to the development of instruments called the Systems Thinking Inventory or Bathtub Dynamics tests by Linda Booth Sweeney and John Sterman (2000). These tests measure the subject's performance on selected systems thinking principles, such as stocks and flows, time delays and feedback. Their results showed that basic systems thinking principles such as stock and flow relationships were poorly understood by the participants or the required algebra could not be applied or was used in a way not applicable to the task (Booth Sweeney and Sterman 2000). The tests have been applied by other researchers to expand the available data and the understanding of people's systems thinking competences (e. g. Kapmeier and Zahn 2001, Kapmeier 2004, Kainz and Ossimitz 2002, Ossimitz 2002, Booth Sweeney and Sterman 2002, Fisher 2003, Heinbokel and Potash 2003, Kubanek 2003, Lyneis and Lyneis 2003, Quaden and Ticotsky 2003, Zaraza 2003, Capelo and Dias 2005, Pala and Vennix 2005). Of particular interest were the studies at the Massachusetts Institute of Technology (MIT) and the Universität Stuttgart conducted by Booth Sweeney and Sterman and Kapmeier respectively.

In this study, we selected three different groups for the assessment of which two were at university level education; one being a forest science diploma course (FOR) at the Technische Universität München in Freising and the other an international Master of Science in Sustainable Resource Management (SRM) at the same institute. It was considered interesting to find out whether these students had developed the ability for systems thinking through their education. By choosing these groups with their focus on environmental and natural resource management issues, we believed we were selecting a group of people more familiar with system behaviour due to the complex and interrelated nature of their areas of study. The third group, three 10^{th} grade high school classes from the Martin-Rinckart-Gymnasium in Eilenburg, provided an insight into high school level systems thinking skills. The idea was to find out what potential the test groups possessed for handling these specific systems thinking tasks and how their results compare to each other. The interest lay in discovering whether there are any difference between the groups, what they would look like and how they could have come about.

2. Procedure

2.1 The tasks

The intention was to produce results comparable with other research that used the systems thinking inventory tasks, in particular the research by Booth Sweeney and Sterman (2000) and Kapmeier (2004). The aim of the assessment is to determine the systems thinking abilities of the participants. The tasks involve simple algebra and straightforward logic. However, they also require understanding and application of principles of systems, especially stock and flow relationships and time delays as well as mathematical principles such as graphical integration (Booth Sweeney and Sterman 2000). The '*Systems Thinking Inventory Coding Guide'* document, provided by Booth Sweeney and Sterman (2001), describes in detail a set of tasks with different levels of difficulty as well as the correct solutions and coding rules. For this study, we selected three tasks, which are the following:

1. **Department Store task (DS):** the DS task shows a graphical representation of people leaving and entering a department store during a 30-minute period. The students are required to answer four questions concerning the flows (Question 1 and 2) and the accumulation of flows (Question 3 and 4). Each correctly answered question received one point.
2. **Bathtub task with square wave pattern (BT):** the cover story is a bathtub containing water, which is being filled and drained at the same time. The BT task assesses the participant's ability to track the changing quantity of a stock depending on the inflow and outflow. Here the participants could score up to 7 points. See Appendix A for a full description of the task and the solutions.
3. **Manufacturing Case task (MC):** the "Manufacturing Case" is similar to the BT task but a time delay was introduced to the stock and flow system with a single negative feedback loop. The cover story is a manufacturing firm producing widgets at 10,000 units/week with a 50,000 units inventory, which is supposed to be maintained at all times. The MC task represents a stock management situation, which occurs repeatedly in many areas of analysis and in many systems. The manager is required to maintain a certain stock regardless of the changes in stock consumption, which means that he or she will have to adjust production to compensate for the difference.

Appendix A provides a more comprehensive description of these tasks and the solutions. The idea behind these simple tasks is to see if subject responses are consistent with basic physical principles, and with the fundamental constraints imposed by the stock/flow structure (Booth Sweeney and Sterman 2001).

University students filled out a questionnaire on their demographics to be used in the statistical evaluation because other studies have found that social demographics such as prior degree, origin and academic background do play a role in the level of performance (Booth Sweeney and Sterman 2000; Kapmeier 2001 and Ossimitz 2000). The questionnaire was adapted depending on the group involved. The assumption is that subjects with previous degrees and subjects with engineering and mathematical backgrounds would perform significantly better than their colleagues.

2.2 The assessment subjects

All participants received the same tasks, they were volunteers (in the case of the university students) and neither received payment nor got any other form of reimbursement. The selection of the groups was determined by the access of the authors to subjects of these groups. The following paragraphs will briefly describe the test groups.

2.2.1 10th grade students at the Martin-Rinckart-Gymnasium

This group, called PUPIL, consisted of three classes of the Martin-Rinckart-Gymnasium in Eilenburg, Germany with 16, 19 and 19 students each, making a total of 54 participants. The age of the students was between 15 and 17 years. A “Gymnasium” in Germany is the highest school level that a pupil can visit during his or her secondary education. The mathematics teacher of the individual class undertook the test with the students. No mention was made of the background to the tests and the pupils believed they were being graded. Being a very homogenous group, the pupils did not fill out the background questionnaire. The only additional information was their names and this was only used to distinguish between male and female participants. Table 1 shows the different classes with their gender distribution.

Table 1: Gender distribution of the PUPIL group by class

Class	Male	Female	Total
10-1	7	9	16
10-2	5	14	19
10-3	10	9	19
Total	22	32	**54**

Some mention should be made of the student’s progress with the school curricula, especially with regards to mathematics. School curricula in Germany are a matter of the *Länder,* i.e. the individual states. Eilenburg is situated in Saxony. Pupils of the 10^{th} grade receive lessons covering mathematical proof, exponential and logarithmic functions, trig metric functions, trigonometry, calculation, and presentation of objects, introduction to statistics and functions (Freistaat Sachsen 2001). By the time of the test the students had covered mathematical proof and linear, exponential and logarithmic functions.

2.2.2 Students on the forest science course

The students of the forestry course at the Technische Universität München, abbreviated with FOR, consisted of two groups; one being mainly first semester students and the second group consisting of higher semester students. The first semester test group did also include some students from third or fourth semesters. Nevertheless, in this study we will consider them as one group. The first and lower semester students numbered 21 whereas the higher semester group consisted of 22 students. Table 2 shows the entire group’s demographics.

Table 2: Demographics of the FOR group

Age	No. of students	Gender	No. of students	Prior Field of Study (%)	
20 – 22	16	Male	29	None	81,4
23 – 25	15	Female	14	Engineering	4,7
26 – 28	2			Business/Management	0,0
34	1			Social Sciences	9,3
Mean age	22			Natural Sciences	2,3
				Humanities	0,0
				Other	2,3

Highest Prior Degree (%)		Region of Origin (%)		First Language (%)	
Diploma	0,0	Germany	100	German	100
High School	100	Other	0,0	Other	0,0
Technical College	0,0				
Trade School	0,0				
Other	0,0				

The forestry students cover a range of topics during their 9-semester course. It involves a basic course including a general foundation in science and economics, natural scientific foundation of forestry, technical aspects of forestry production and economic and sociological foundations of forestry. The foundation course lasts 4 semesters and includes subjects such as mathematics, statistics, physics, general business management, general economics and theory of science. The fifth to ninth semesters are covered by the extension courses. These involve topics such as balance of material in forest ecosystem, climatology, silviculture, forest yield science, engineering biology and business management in forest and timber industries, forest politics, and management consultancy.

2.2.3 Students on the Sustainable Resource Management course

The Sustainable Resource Management course at the Technische Universität München is a three semester international master programme open to students from all over the world, although a selection process based on suitability is used to grant access to the programme. The course is taught in English and covers a wide range of topics related to resource management. The first semester covers basic courses such as environmental resource economics, information management, human resource management, remote sensing and project management (see also Biber and Kasperidus 2004 for more information). During the second semester the students select two elective courses as their focus. The third semester is spent on the master thesis. At the time of the assessment the students had covered the systems thinking lecture and parts of the system analysis course given by two of the authors. The participants numbered 34 and their demographics are shown in Table 3.

Table 3: Demographics of SRM group

Age (No.)	No. of students	Gender	No. of students	Previous field of study (%)	
23 – 25	10	Female	17	Engineering	35,3
26 – 28	12	Male	17	Business / Management	14,7
29 – 31	6			Social Science	14,7
32 and above	3			Natural Science	29,4
Mean age	25			Humanities	2,9
				Other	2,9

Highest previous degree (%)		Region of origin (%)		English as First language (%)	
Bachelor of Arts	17,6	Europe	41,2	Yes	11,8
Bachelor of Science	32,4	North America	0,0	No	88,2
MA, MS, Diploma	38,2	Africa	8,8		
Ph.D.	0,0	Asia and Middle East*	44,1		
High School	0,0	Latin America	5,9		
BE, JD, BBA, MD, CPA	5,9	Australia / New Zealand	0,0		
Other	5,9				

* Note that the majority of students come from Asia, not the Middle East.

2.3 Assessment procedure at university level

For the university groups the following procedure was used to conduct the tests individually for the two FOR groups and the SRM group. First of all we tried to find the best time slot for the task to achieve a high participation rate. The whole process required a total time slot of about one and a half hours for each group. After a brief introduction about the goals of the study, the participants were given 45 minutes to complete the three tasks.
The tests for the forestry students were both on one day and the likelihood of the test paper being handed to the other semesters was unlikely. The forestry students and SRM students have very little contact to each other. The students were asked to sit far enough apart to discourage any copying. The test subjects were allowed to leave the room as soon as they were finished with the questions.

3.0 Results

3.1 General results

The total number of participants, and therefore test papers to be evaluated was 131. Of these 34 belonged to the SRM students, 43 were from the forest science course and 54 were provided by the school students. In general the results show that subjects in all three groups were able to achieve a full score in all three tasks. Figure 1 shows the distribution of the total score for all groups and highlights the variation in overall performance. Only the SRM group included subjects with zero points. The forestry students outperform the other two groups by a tendency for achieving better total scores as well as having a higher minimum score. The pupils have the second largest variation in performance, although not as large as the SRM group who produced the lowest median score. The forestry students produced results with a median value around 10 points. The other two groups have the majority of their total scores below the 10-point mark. The high school students are above 5 points whereas the participants from the SRM course are on the 5-point mark.

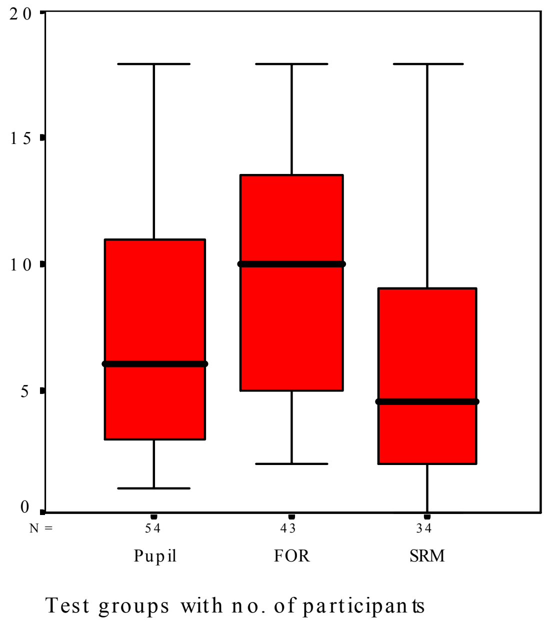


Figure 1: Total performance on all tasks by the test groups

3.2 Group comparison

The chapter will look at the specific differences between and within the groups. Table 4 shows the average performance of the individual groups according to the separate tasks and in total.

Table 4: Comparison of average group performance by task and in total. Numbers are a percentage of the average score.

Group	N	Total %	DS Task %	BT Task %	MC Task %
Pupil	54	41	57	47	26
SRM	34	34	58	39	16
Forestry	43	54	57	63	42
Total	131	43	57	50	29

The forestry students achieved the highest mean result with a total of 54 % correct solutions to the questions. Their scores for both the BT task (63 %) and the MC task (42 %) were higher than the scores of the other two groups. In the case of the MC task the difference in performance between the groups was the most pronounced. The BT task produced the best individual results by any group. In the case of the DS task the group results are very similar to each other and the SRM group scored the highest marks. It is remarkable to see that subjects that did not solve the DS task correctly were able to solve the BT task.
Figures 2, 3 and 4 show the details for the every group in their performance on the tasks.

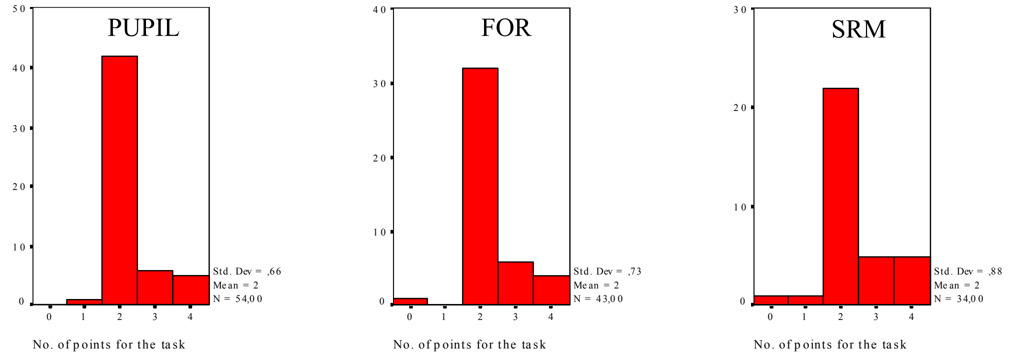


Figure 2: Performance on the Department Store (DS) task by the test groups.

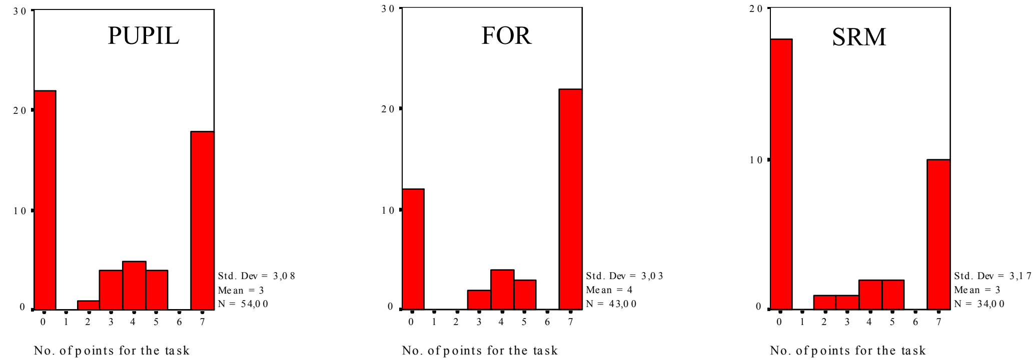


Figure 3: Performance on the Bathtub (BT) task by the test groups.

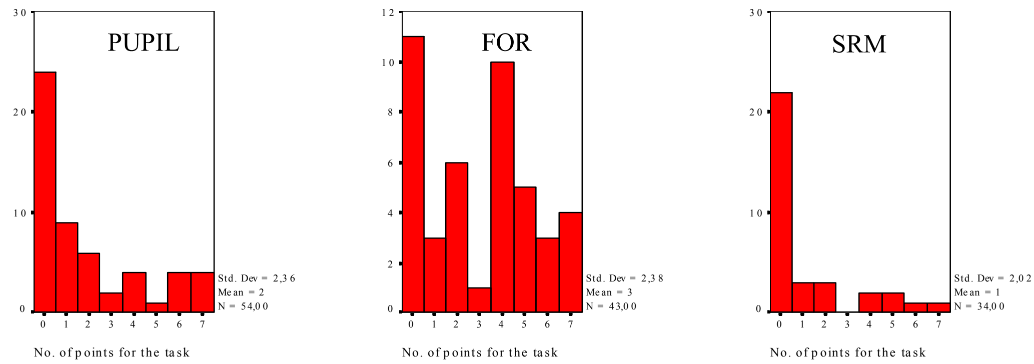


Figure 4: Performance on the Manufacturing Case (MC) task by the different groups

Table 5 covers the performance by all groups on all tasks. It shows the frequency and percentage depending on the number of points reached. The SRM group produced the best results on the DS task with 14.7% achieving the full four points. The pupils and forestry groups each had 9.3% answering the all questions correctly. The BT task changed this. The pupils had 33.3% fully correct answers, whereas the forestry group had 51.2% gaining 7 points. The SRM group on the other had 29.4% full score answers. The MC task produced the many sheets with zero points (44.4% for the pupils and 64.7% for the SRM). The forestry group had 9.3% or four persons with the full score. Nevertheless, the largest group was 25.6% with zero points.

Table 5: Group performance on all tasks showing frequency and percent

Groups		Pupils			SRM			Forestry	
Tasks	Points	Freq	%	Points	Freq	%	Points	Freq	%
DS task	1	1	1.9	0	1	2.9	0	1	2.3
	2	42	77.8	1	1	2.9	2	32	74.4
	3	6	11.1	2	22	64.7	3	6	14.0
	4	5	9.3	3	5	14.7	4	4	9.3
				4	5	14.7			
BT task	0	22	40.7	0	18	52.9	0	12	27.9
	2	1	1.9	2	1	2.9	3	2	4.7
	3	4	7.4	3	1	2.9	4	4	9.3
	4	5	9.3	4	2	5.9	5	3	7.0
	5	4	7.4	5	2	5.9	7	22	51.2
	7	18	33.3	7	10	29.4			
MC task	0	24	44.4	0	22	64.7	0	11	25.6
	1	9	16.7	1	3	8.8	1	3	7.0
	2	6	11.1	2	3	8.8	2	6	14.0
	3	2	3.7	4	2	5.9	3	1	2.3
	4	4	7.4	5	2	5.9	4	10	23.3
	5	1	1.9	6	1	2.9	5	5	11.6
	6	4	7.4	7	1	2.9	6	3	7.0
	7	4	7.4				7	4	9.3

3.3 Impact of subject demographics

Both the forestry and SRM groups were provided with questionnaires to collect demographic data in order to find possible connections between a participant's background and his or her performance on the systems thinking tasks. Using an analysis of variance the subject demographics and task results were evaluated. Neither the forestry nor the SRM group showed any significant impact on the different performance rates at $p \leq 0,05$.

Gender was the only demographic factor, which was included in the school student's assessment sheet. Table 6 shows the results of a t-Test on the mean performance for the individual tasks and in total by the two groups. The BT and the MC tasks as well as the total score produced a highly significant difference in performance. Male pupils coped better with the systems thinking tasks.

Table 6: Comparison of mean performance depending on gender by task and in total with significance levels indicated by p

Gender	N	DS task	BT task	MC task	Total
Female	32	2,25	2,50	1,00	5,75
Male	22	2,32	4,55	3,09	9,95
		p = 0,711	p = 0,015***	p = 0,001***	p = 0,002**

** = high significance; *** = very high significance

3.4 Comparison between the test groups

The following paragraphs will look at further differences between the groups with regards to the total and task performance. Table 7 shows the evaluation of the performance per task and in total for all groups combined. The BT task and MC tasks and the total score produced significant differences. In particular the MC task showed a very high significance at p = 0,002 using a test of variance.

Table 7: Comparison of test groups depending on task and total performance (significant items at p ≤ 0,05 in bold)

	DS task			**BT task**			**MC task**			**Total**		
Variable	df	F	p	df	F	p	df	F	p	df	F	p
All test groups	2	0,126	0,882	2	3,113	**0,048**	2	6,409	**0,002**	2	5,047	**0,008**

df = degrees of freedom; F = ratio of variance between / within treatments; p = level of significance

Further evaluation and comparison of the performance between the groups showed significant difference in the case of the forestry and SRM groups for the BT and MC tasks as well as for the total score, see Table 8. Significance between the test group results could also be determined in the case of the pupils and forestry students for the MC task and the total score. In all cases the significance was high thereby validating the hypothesis that there would be a difference in performance between the groups.

Table 8: Results of a t-test for comparison of test groups depending on task and total performance (significant items at p ≤ 0,05 in bold)

	DS task			**BT task**			**MC task**			**Total**		
Groups	Mean	df	p	Mean	df	p	Mean	df	p	Mean	df	p
Pupil / FOR	2,28 / 2,28	95	0,993	3,33 / 4,44	95	0,079	1,85 / 3,00	95	**0,021**	7,46 / 9,74	95	**0,028**
Pupil / SRM	2,28 / 2,35	86	0,649	3,33 / 2,74	86	0,382	1,85 / 1,18	86	0,171	7,46 / 6,26	86	0,274
FOR / SRM	2,28 / 2,35	75	0,690	4,44 / 2,74	75	**0,019**	3,00 / 1,18	75	**0,001**	9,74 / 6,26	75	**0,003**

df = degrees of freedom; p = level of significance

Figures 5 to 8 show the values and range of variation of the group's mean performance on the individual tasks and the total score. Figure 5 describes the variation for the DS task. No overlap occurs confirming that no significance could be determined between the groups for this task. The high school students have the lowest variation in mean performance whereas the SRM students have the widest spread in the points they reached. Figure 6 describes the effects of the BT task. A level of significance could be determined (see Tables 7 and 8).

The graph shows a marked change in variation of performance from the previous graph. The forestry group produces a higher level of performance than both the pupils and the SRM group. This trend continues in the MC task and for the total score, Figures 7 and 8 respectively. The significant differences are backed up in Tables 7 and 8. In particular the lack of overlap between the forestry group and the SRM give weight to the significant variation in performance.

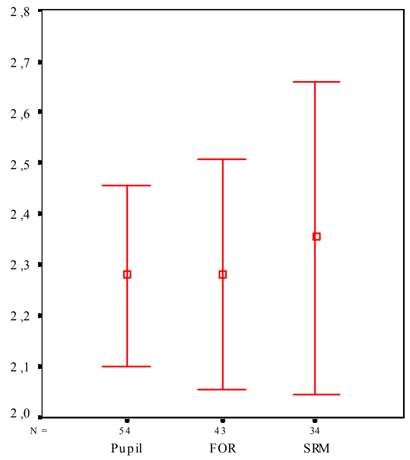


Figure 5: Variation of means for the DS task

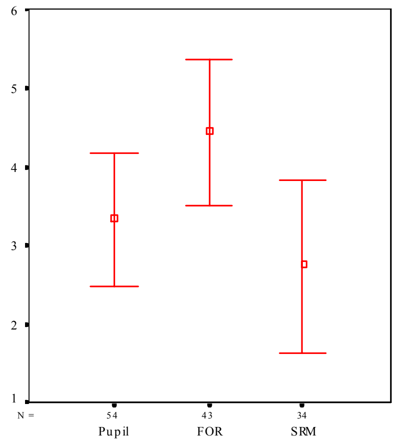


Figure 6: Variation of means for the BT task

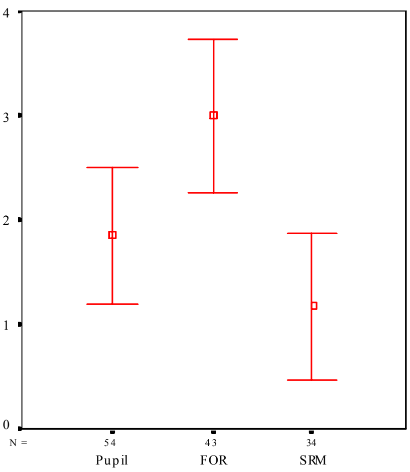


Figure 7: Variation of means for the MC task

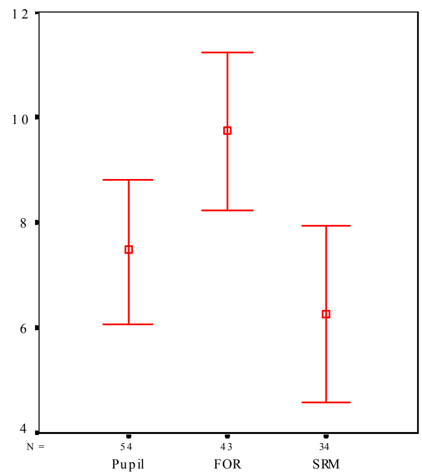


Figure 8: Variation of means for the total score

The analysis of the results indicated that gender played a strong role in determining the participant's performance. All groups were tested on the significance of gender with the result that the BT and MC tasks showed a high significance, see Table 9.

Table 9: Gender effect on task and total performance by all test groups (significant items at p ≤ 0,05 in bold)

	DS task			BT task			MC task			Total		
Variable	df	F	p	df	F	p	df	F	p	df	F	p
All test groups	4	1,122	0,349	5	2,453	**0,037**	7	2,380	**0,026**	18	1,373	0,159

df = degrees of freedom; F = ratio of variance between / within treatments; p = level of significance

3.5 Typical errors

The mistakes made by the participants are essential for understanding and discussing the assessment results. The following section will show typical errors and briefly discuss the mistakes made.

The mistakes for the DS task are cases where subjects believed minute 8 to be the time when most people are in the store and minute 16 when the fewest people are in the store. Minute 8 is the time when the difference in the flows is at its' highest and when the least people are leaving. The subjects who answer the questions in this way show a lack of differentiation between accumulation and the net rate. For the BT task the common mistakes were spreadsheet thinking (instead of drawing continuous lines for the changes in water level, the change is shown as steps) and pattern matching (the subject transferred the inflow or outflow pattern directly to the changes in water quantity). The MC task also included spreadsheet thinking and pattern matching (in this case transferring the order line directly to the inventory graph) as well as not showing a production overshoot, not letting the production and inventory start in equilibrium with the orders and not matching the content of the production graph to the inventory graph. The subjects also produced non-typical errors.
Figure 9 is a pupils' response to the BT task. Although the participant started the stock trajectory at 100 l it then dropped instead of rising with an increase in the net flow. The time sections (4, 8, 12 and 16 minutes) are ignored.

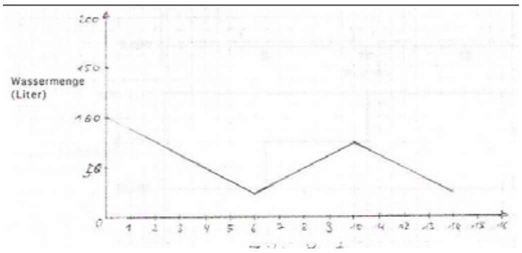


Figure 9: Pupil response to the BT task

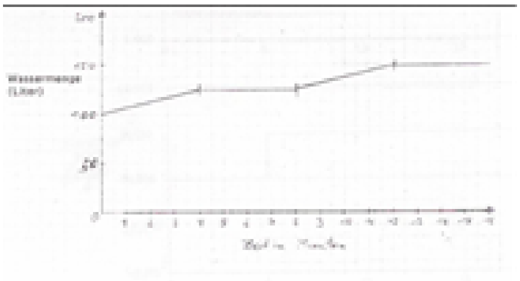
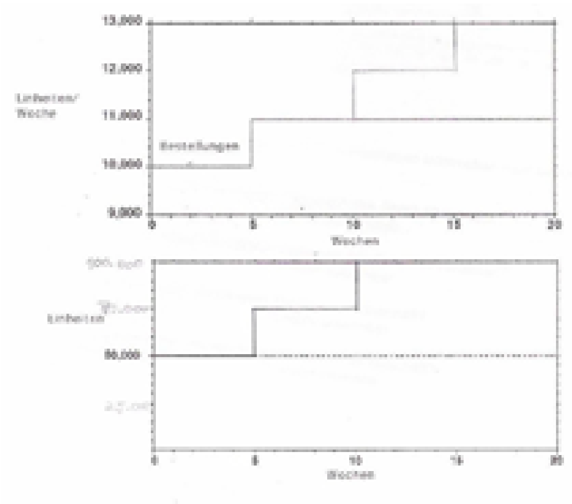


Figure 10: Pupil response to the BT task

In Figure 10, the subject has correctly understood that a distinctive change in the stock occurs every 4 minutes. The explanation in the margin provided an insight into the reasoning behind the answer: the subject stated that the initial 100 l are increased by 25 l, although water is also draining out of the tub. But this does not affect the water quantity because it is 0 l. Then another 25 l are added. The subject ignores the fact that every minute 25 l are added to and 50 l are drained out of the bathtub.



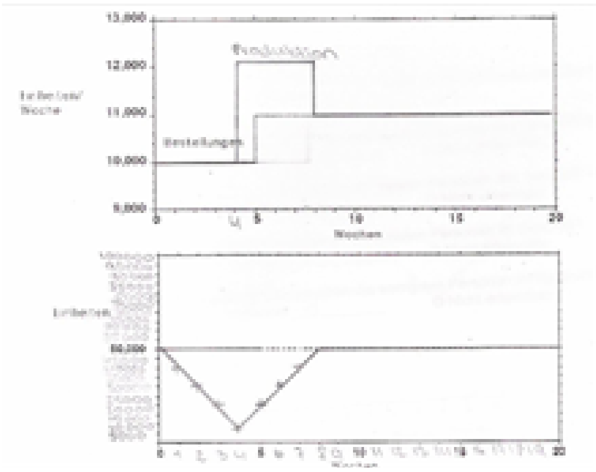
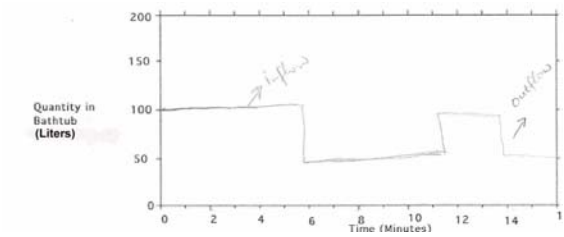


Figure 11: Pupil response to the MC task

Figure 12: Pupil response to the MC task

In Figure 11, the pupil answered the MC task by allowing both the production line and the inventory to rise every 5 weeks by 1,000 units in the case of the production and by 25,000 in the case of the inventory. This was a case of pattern matching. The development of the inventory does not coincide with the change in production. Instead the subject took the order line as a basis for both the production and inventory lines. Figure 12 is another example of a subject's response to the MC task. Here the participant included an overshoot in production and the rise and drop in the inventory, but the lines do not coincide with each other. The drop in inventory is too early and the change in production output begins before the increase in orders. The answer ignored the time delay for adjusting the production process. Nevertheless some basic principles, such as the drop in inventory due to an increase in orders, were understood.



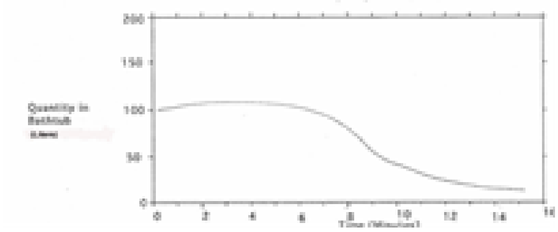


Figure 13: SRM student response to the BT task

Figure 14: SRM student response to the BT task

Figure 13, an answer from the SRM group for the BT task, where the participant combined the inflow and the outflow within one line, as indicated by the labelling. This is also a case of pattern matching and it lacks any reference to the relevant time sections. In Figure 14, the SRM student produced a solution to the BT task with no apparent understanding of the task requirements. The line drawn does not related to the changes in flows within the stock with only the starting point at 100 l showing any relation to the task setting.

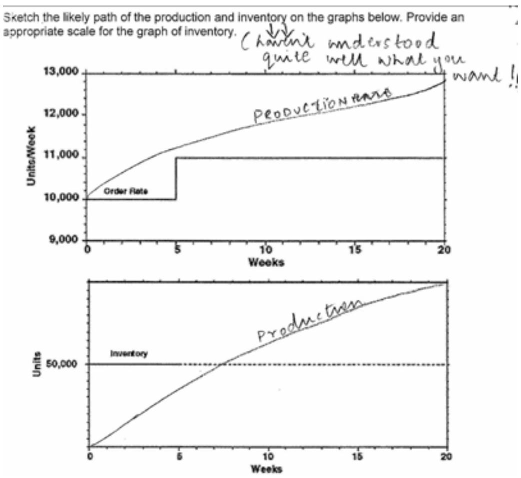


Figure 15: SRM student response to the MC task

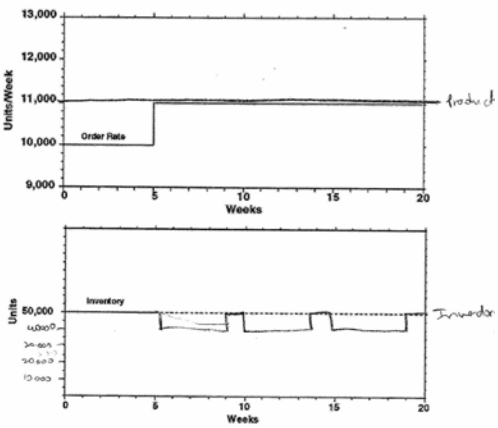


Figure 16: SRM student response to the MC task

The subject who produced the answer in Figure 15 provided her own explanation for the incorrect answer by stating confusion and uncertainty as to the requirements of the task despite having acknowledged important aspect on the sheet paper such as the lag in production adjustment and the rise in orders. But the answer does not coincide with her remarks on the paper. Both graphs included the production line, instead of production and inventory lines. The lines are of similar trajectory but do not coincide with the task information. Figure 16 is a further response to the MC task where production and inventory lines do not show any relevance to each other, nor are they connected to the information in the task setting. The subject ignored the fact that the production had to start in equilibrium and that a time lag as well as an overshoot would occur. Instead the production line was drawn to cover the 10% increase in orders beginning at point zero. The inventory on the other hand shows a repeated drop to 40,000 units lasting four weeks (reference to the time lag in adjusting the production) with a week recovery in between. The information contained in the task setting seems to have been partially and selectively included in both graphs but with no correlation to each other.

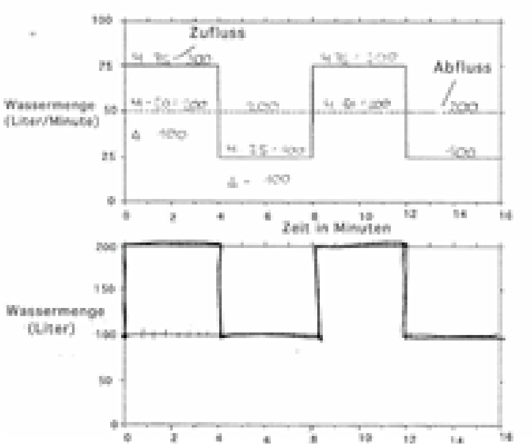


Figure 17: Forestry student response to the BT task

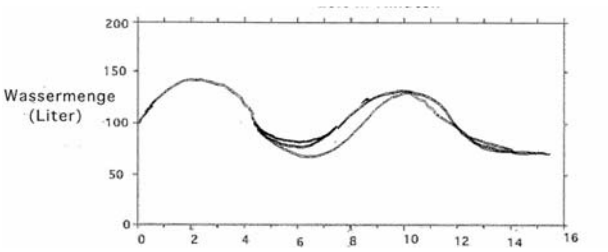


Figure 18: Forestry student response to the BT task

In Figure 17, the subject from the forestry group displayed an example of pattern matching despite the fact that the calculations are correct. The stock changes discontinuously in the same way as the net rate. The answer in Figure 18 on the other hand showed that the subject understood that the stock will change depending on the net flow, but failed to calculate correctly the exact figures and disregarded the effect of peaks and troughs when the net flow crosses zero.

The results and the group evaluation have implications on the systems thinking capacity of the participants. As this chapter has shown the results are overall poor and show a lack of understanding for system principles tested in the Systems Thinking Inventory. This is the overall result of the evaluation. Looking in more detail one can seen that the German university students performed better than the group of international students and the pupils, although analysis of the individual groups and their demographics showed no correlation between education and performance. One aspect of interest that appeared repeatedly was the performance difference between male and female participants. Here it was shown that male students and pupils outperformed their female counterparts. These and other implications of the results will be discussed in the next chapter.

4. Discussion

The research results show that the participants have a poor understanding of system principles, such as stocks and flows, time delays and feedback processes. The subjects produced fundamental errors with regards to the behaviour of systems and not just calculation mistakes. Besides, the required calculations were straightforward and should not have caused serious problems for people with their level of education. Instead the test groups displayed systematic errors in their handling of the tasks. In the case of the DS task subjects were able to understand the graph but failed to accumulate flows. The BT task caused many students to draw stock trajectories similar to the net rate, thereby failing to prove that they understand the difference between the net flow into a stock and how this reflects on the stock trajectory. The MC task showed fundamental systemic errors as the subjects ignored the conservation of material and failed to match the inventory to the production lines. The poor results from the three tasks are reflected in the total performance, a further indication of the overall low performance rate.

This is also the conclusion that other studies have reached, notably Kapmeier (2004) and Booth Sweeney and Sterman (2000). The poor performance is despite the fact, that two of the groups were at university level, although the forestry group was composed of a large number of first semester students. The Sustainable Resource Management students were even at master level and had received a lecture in systems thinking and system dynamics before being presented with the test. With this study we wanted to find out whether students in natural resource and environmental management courses could perform well on systems thinking tasks, such as the Bathtub dynamics series, and thereby verify the hypothesis that natural and environmental resource management would show higher capability in solving such tasks due to the implied focus on complexity and interrelatedness of their subject areas. This proved not to be the case. The test groups individually and as a whole performed poorly. Yet it is the difference within and between the groups and their varying performance results that provide the interesting aspects of the research.

A highly significant difference lay between the results of the different test groups in relationship to each other. This was determined for the BT task and MC tasks as well as for the total score. The MC task in particular caused a wide divergence in the rate of performance with the FOR group producing far better results than the other two. The difficulty of the task, possibly due to the appearance of a time delay, the necessity to conserve matter and match the production line to the inventory graph, were more easily dealt with by the forestry students whereas the SRM group produced their poorest results here. How could these differences come about? Subject demographics for the SRM students showed no significant impact, although the main differences between these two groups are their origin (and therefore different school system and curricula) and their previous studies and degrees (again a different university education or no previous university degree). If the argumentation follows the line that more education results in better performance, then the SRM group should have produced the best results of all three groups. In fact, their results were the poorest.

The students of the forestry course provided enough input to produce better results than the SRM subjects, although there was no significant difference between the higher semester students and the lower semester students. The forestry course in itself does not appear to increase the participant's capacity for solving the system thinking tasks of this research.

The performance differences show that the school students – although not significantly better – nevertheless produced better results again than the SRM group prompting the conclusion that the school curricula and teaching in Germany (in this case in the federal state of Saxony) covers enough ground to outperform the SRM group. Quite likely the performance of the school group in 6 to 8 years would produce similar results as the forestry students now, although this would be the matter of further observation. One could also say that age is a determining factor in improving systems thinking skills. As the students get older, and therefore gain more knowledge and understanding, their performance improves. Evaluation proved no significant connection between age and performance. What is more likely and a more interesting point is the matter of experience. The study by Capelo and Dias (2005) found that the managers with up to 14 years work experience performed better on the Bathtub dynamics tasks than did the student group in the same study. Experience can be seen as a combination of different factors, such as education, work, upbringing, discussions, listening, trial and error, successes and failures; basically life long learning. And although experience will usually grow with age, it need not be the determining factor. More effective would be to increase a student's systems thinking experiences through education. This is also supported by the recent research in this field by Pala and Vennix (2005).

Comparison with other research

A useful aspect of the Bathtub dynamics systems thinking tests is the comparability with the other studies. Table 10 shows two other studies in comparison with the average performance by the test groups from the current research. The other studies were conducted by Booth Sweeney and Sterman at the Massachusetts Institute of Technology (MIT) in 2000 and Kapmeier and Zahn with the students from the Stuttgart Institute of Management of Technology (SIMT) 2003 and the University of Stuttgart from 2000 – 2001 and 2002 – 2003.

The subjects from MIT and the University of Stuttgart produced the best average results whereas the SRM students performed poorly in comparison. Only the forestry students came close to competing with the SIMT group on the BT and MC tasks and with the MIT students on the MC task. The pupils were also able to produce results close to the SIMT group in the case of the MC task. Nevertheless all studies come to the conclusion that the performance is poor (Booth Sweeny and Sterman 2000; Kapmeier 2004) as does this study. It is necessary to look behind the bare figures to discuss the implications and reasons for the results.

Is education – current and previous - a key factor in promoting systems thinking skills? Booth Sweeney and Sterman (2000) found that there was only a weak link between education and performance, but their test groups outperformed the groups of this study, in particular the SRM group. The students from MIT were able to produce 44 % better results on the BT task than the SRM group and still 30 % better at answering the MC task. This is a considerable advantage. Yet the difference for the DS task was merely 7 %. In the case of the forestry students the gap was smaller (20 % better for the BT task and only 4 % for the MC task). Booth Sweeney and Sterman (2000) state that their test groups consisted of "highly educated subjects with extensive training in mathematics and science…" and "many had years of coursework and even undergraduate or graduate degrees in mathematics, engineering or the sciences”. Admittance to the MIT Sloan School of Management requires considerable expertise, experience, academic achievement as well as above average quantitative and analytical skills, according to the website (MIT Sloan School of Management 2005). A higher performance than a group of 16-year-old school students would be expected. And although this proved to be the case (8 % better for the DS task, 36 % for the BT task and 20 % for the MC task), do these results reflect the differences in education, training and experience? According to Booth Sweeney and Sterman (2000) the MIT results are disappointing for a group of such highly qualified persons, which would place the pupils in a comparatively good position.

Table 10: Comparison of average performance of the groups from this study with other groups. Numbers are a percentage of the average performance.

Group	DS Task %	BT Task %	MC Task %
Pupils	57	47	26
SRM	58	39	16
FOR	57	63	42
MIT (2000)	65	83	46
SIMT (2003)	66	65	31
Uni Stuttgart (2000 - 2001)	N/A*	68	69
Uni Stuttgart (2002 - 2003)	67	83	62

Note that the groups MIT, SIMT and Uni Stuttgart are those without Beer game experience
* Not available; data is missing

The other groups assessed by Kapmeier (2004), which this study is compared with generally show a higher performance than the pupils, SRM and forestry students. An exception is the MC task where the forestry students produced results 9 % better than the SIMT group. The University of Stuttgart group from both test years showed a considerably higher performance for all three tasks than any of the groups from the current study. Kapmeier (2004) related the success in the MC task to the student’s experiences with production issues during their university course, thereby supporting the idea that specific education will improve systems thinking skills. These groups were at master level in their education with business administration and engineering as their main backgrounds and the authors state that a higher level of ambition had been observed in the students over the years, especially those that included system dynamics in their elective course (Kapmeier 2004). Yet despite this the authors draw the same conclusion as Booth Sweeney and Sterman (2000) that the results, in view of the simplicity of the tasks, are far too poor.
There is the possibility that other factors contributed to the outcome of the results, such as motivation and incentives. Booth Sweeney and Sterman (2000) considered alternative explanations for the results of their study and came to the conclusion that there is a school of thought that insists on always providing incentives to all voluntary assessments, but at the same time research has also show performance to stay the same or even drop despite incentives. Booth Sweeney and Sterman (2000) continue by saying that the knowledge needed for successfully completing the tasks is simple enough not to be too demanding on the subject’s motivation. Nevertheless this study has found a number of answers lacking any sign of being truly motivated. More research would be needed to assess the amount of

motivation that flows into answering the questions. One option would be to assess the level of motivation by asking relevant questions either on the paper or during post-assessment interviews or by selecting the test subjects in such a way as to ensure sufficient motivation, although such a method would produce a bias in the results as well. The school students stand out, because of their very real assessment conditions and the fact that they knew nothing about the empirical research they were taking part in.

Booth Sweeney and Sterman (2000) also considered time to be a limiting factor in the performance results. The students in this study were given 45 minutes to complete the tasks and although the exact times were not noted down, the two university groups finished the papers well within the time limit. Most students started handing in their papers after about 20 – 25 minutes. Lack of time therefore could not have been a reason for the poor results, although this assessment only involved three tasks.

Language proved to have an impact on the test results, although it was not significant. Despite this there is some evidence that the subjects had problems in understanding the task settings. The MC and BT tasks are a point in case. Comments in the margins stating confusion and uncertainty are an indication of this assumption. Lack of understanding and uncertainty could also be seen on some answer sheets of the school group. Was language a barrier in the case of the SRM students and the description of the tasks a problem for the school students? Both would be a reflection of the subject's ability to handle information in an assessment text and transfer this through logical thought and mathematical calculations to the graphs. This is a requirement that occurs repeatedly during a school and university career. It would be an area, which needs more research. It can be said that the better the understanding, the higher the motivation will be to answer the questions, although this does not guarantee better results. As the studies have shown the mistakes are systematic and not a matter of comprehension. The common mistakes would continue, but it would prevent subjects from dropping out through a lack of understanding. More understanding of the individual answers would be necessary to find out which of the student's produced errors due to lack of understanding or lack of motivation.

This study found that gender repeatedly played a significant role in the group's performance rates. The pupils especially produced results, which turned out to have a highly significant gender difference. Not just the total score but also the BT and MC tasks had male participants performing better than female subjects. Male pupils seem to be more capable of solving the systems thinking problems. In the case of the SRM students only a slight significance could be determined in the case of the DS task. Gender did not have a significant impact on the group's performance. The forestry students as well showed a slight significance of gender having an effect on the participant's results. Male students performed on average better than female students and in the case of the BT task a significance of $p \leq 0{,}10$ could be determined. Booth Sweeney and Sterman (2000) discovered differences between men and women, in their study as well. Men produced marginally significant but consistently better results than women on all three tasks. Kainz and Ossimitz (2002) found a clear indication that male subjects solved the tasks consistently better than women. The tasks were designed to test stock and flow understanding as well as reading and interpreting graphs. The researchers found the results unexpected as the subject's mathematics grades showed no significant differences between

males and females (Kainz and Ossimitz 2002). The gender effect on performance appears to begin at least before students enter university, but is then reduced in its significance. This would indicate the need to address the problem as early as possible in a person's education career. More research into the gender-performance relationship would allow to pinpoint the time when the difference becomes evident as well as finding out more about the reasons behind the results and possible solutions to the situation. Kainz and Ossimitz (2002) press for more research with this issue, in particular with regards to research on systems thinking.
The understanding and solving of systems thinking tasks seems to be truly difficult, despite the simplicity and straightforwardness of the problems. The studies that used the Bathtub dynamics tests have shown this to be the case. Education and training are no guarantee for excellent performance (see also Pala and Vennix 2005). Yet, this is not the point at which to give up. Rather it provides an insight into the problem of increasing systems thinking skills in schools and universities. This can and should be followed by strategies and actions to change the situation from the point of view of teaching and learning these skills, in particular with regards to system dynamics. This can be achieved by focussing on the basics, such as stock and flow relationships, time delays, feedback processes and developing intuition, as well as removing deeply entrenched beliefs and theories about how the physical world functions (Booth Sweeney and Sterman 2000). By providing the necessary environment for learning these skills. This involves the essential technology, trained teachers and experimental laboratories (Gould-Kreutzer 1993). By adjusting the school curricula to accommodate system dynamics and systems thinking skills. By fostering a political climate that acknowledges the necessity for system dynamics education in schools. And by drawing in the key players: parents, schools administrators, politicians and last but not least the students.
Finally, much more research is needed to increase the understanding of system dynamic in education, in particular in Germany. And perhaps then we will able to support people in their ability to learn and, as Edwards (1999) puts it, "to break free from the failures of the past" and to understand "the reality of increasing interconnectedness".

5. Acknowledgements

Many thanks to Mr. Dieter Mannel, director of the Martin-Rinckart-Gymnasium in Eilenburg, for his enthusiastic support to carry out this study at his high school, to Mrs. Hilbert and Mrs. Mai, who applied the tasks in their mathematics classes and to the pupils, who were part of this study. We also thank all the students from the Technische Universität München, who participated in this survey. Finally, we thank the anonymous reviewers for their constructive and helpful advice.

6. References

Biber, P. and **Kasperidus, H.D.,** 2004: Integrated Modeling Approaches and System Dynamics in Education related to Sustainable Resource Management, Forestry, and Land Use Management. Proceedings of the 22th International Conference of the System Dynamics Society, July 25 -29, 2004, Oxford, England, 20 pages.

Booth Sweeney L. and Sterman J., 2001: Systems Thinking Coding Guide. Version 1.0, October 2001.

Booth Sweeney L. and Sterman J., 2000: Bathtub dynamics: initial results from a systems thinking inventory, System Dynamics Review 16, 249–286.

Booth Sweeney L. and Sterman J., 2002: Cloudy Skies: Assessing Public Understanding of Global Warming, System Dynamics Review 18(2).

Capelo C. and Dias J. F., 2005: Strategy Lab Experiences: Bathtub Dynamics. ADETTI/ISCTE, Portugal.

Diamond J., 2005: Collapse. How societies choose to fail or succeed, Viking, London.

Dörner D., 2004: Die Logik des Misslingens. Strategisches Denken in komplexen Situationen, Rowohlt Verlag, Reinbek.

Duffield M., 2001: Global Governance and the New Wars. The Merging of Development and Security, Zed Books, London.

Edwards M., 1999: Future Positive. International Co-operation in the 21st Century, Earthscan, London.

Fisher D. M., 2003: Student performance on the bathtub and cash flow problems. Proceedings of the 2003 International System Dynamics Conference, System Dynamics Society, Albany, New York.

Fisher D. M. and Guthrie S., 1996: Systems Thinking and System Dynamics in the CC-STADUS High School Project (How High School Students Become Systems Thinkers). Proceedings of the 1996 International System Dynamics Conference, System Dynamics Society, Albany, New York.

Forrester J. W., 1961: Industrial Dynamics, Cambridge, Massachusetts, Productivity Press.

Forrester J. W., 1992: System Dynamics and Learner-Centred-Learning in Kindergarten through 12th Grade Education, Paper D-4337, Cambridge, Mass.: MIT.

Forrester J. W., 1994: Learning through System Dynamics as Preparation for the 21st Century. Keynote Address for Systems Thinking and Dynamic Modelling Conference for K-12 Education June 27-29, Concord Academy, MA, USA

Freistaat Sachsen Sächsisches Staatsministerium für Kultus, 2001: Lehrplan Gymnasium Mathematik, Klassen- und Jahrgangstufen 5 – 12. Sächsische Druck- und Verlagshaus AG, Dresden.

Gould-Kreutzer J., 1993: Foreword: System dynamics in education, System Dynamics Review Vol. 9, no. 2; 101-112.

Heinbokel J. and Potash J., 2003: Bathtub dynamics at Vermont commons schools. Proceedings of the 2003 International System Dynamics Conference. System Dynamics Society, Albany, New York.

Kainz D. and 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. System Dynamics Society: Albany, New York.

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

Kapmeier F. and Zahn E., 2001: Bathtub dynamics: results of a systems thinking inventory at the Universität Stuttgart, Universität Stuttgart Lehrstuhl für Planung.

Kubanek G., 2003: Bathtub dynamics – Ottawa. Proceedings of the 2003 International System Dynamics Conference. System Dynamics Society, Albany, New York.

Lewis H. W., 1997: Why flip a coin? The Art and Science of Good Decisions, John Wiley & Sons, Inc. New York.

Lyneis J. and Lyneis D., 2003: Bathtub dynamics at WPI. Proceedings of the 2003 International System Dynamics Conference, System Dynamics Society, Albany, New York.

Ossimitz G., 2002. Stock–flow-thinking and reading stock–flow-related graphs: an empirical investigation in dynamic thinking abilities. Proceedings of the 20th International Conference of the System Dynamics Society, Palermo, Italy. System Dynamics Society: Albany, NY.

Pala, Ö. And Vennix, J. A. M., 2005: Effect of system dynamics education on systems thinking inventory task performance. System Dynamics Review 21, 2, 147 – 172

Quaden R. and Ticotsky A., 2003: Bathtub dynamics at Carlisle Public Schools. Proceedings of the 2003 International System Dynamics Conference. System Dynamics Society, Albany, New York.

Senge P., 1990: The Fifth Discipline, Doubleday, New York.

Sterman J. D., 2000: Business Dynamics: Systems Thinking and Modelling for a Complex World. Irvin McGraw-Hill, Boston.

Sterman J. D., 2002: All Models are Wrong: Reflections on becoming a system scientist, System Dynamics Review 18, 501 – 531.

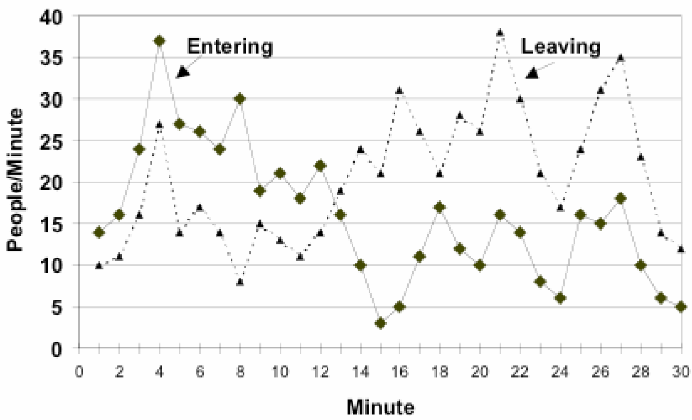
Zaraza, R. 2003: Bathtub Dynamics in Portland at SyMFEST. Proceedings of the 21st International Conference of the System Dynamics Society, New York, USA.

Appendix A: Bathtub Dynamics Tasks

Department Store task (DS)

This task shows a graphical representation of people leaving and entering a department store during a 30-minute period (see figure A1). The students are required to answer four questions concerning the flows (Question 1 and 2) and the accumulation of flows (Question 3 and 4). The correct answer to question 1 is minute 4 and for question 2 it is minute 21, which can be determined by pinpointing the peaks of the two lines on the graph. Question 3 refers to the point in time when the most people are in the store, i.e. when the leaving and entering lines cross each other after minute 13. Finally, question 4 asks the subject to state when the fewest people are in the store, which is after minute 30. Students could determine this by noting the size of the area covered by the entering and leaving graphs before and after minute 13. In addition, a box with "Can not be determined" was available to be ticked by the participant. The participants would still receive full points if their answer lies within a range of +/- 1 minute because of possible scale reading errors, while at the same time understanding the solution correctly.

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 provided information.

1. During which minute did the most people enter the store?
   Minute ____________ O Can't be determined
2. During which minute did the most people leave the store?
   Minute ____________ O Can't be determined
3. During which minute were the most people in the store?
   Minute ____________ O Can't be determined
4. During which minute were the fewest people in the store?
   Minute ____________ O Can't be determined

Figure A1: Description of the Department Store task. Source: System Thinking Inventory - Coding Guide (Booth Sweeney and Sterman 2001)

Bathtub task (BT)

The cover story is a bathtub containing water, which is being filled and drained at the same time. The BT task assesses the participant's ability to track the changing quantity of a stock, in this case water, depending on the inflow and outflow. The subject is required to draw changes in the water quantity as a trajectory on the graph (see figure A2). The task requires simple algebra in addition to an understanding of stock and flow relationships. The task is simple, containing no feedback processes and the flows are exogenous. The round numbers allow the net flow to be easily calculated and added to the stock.

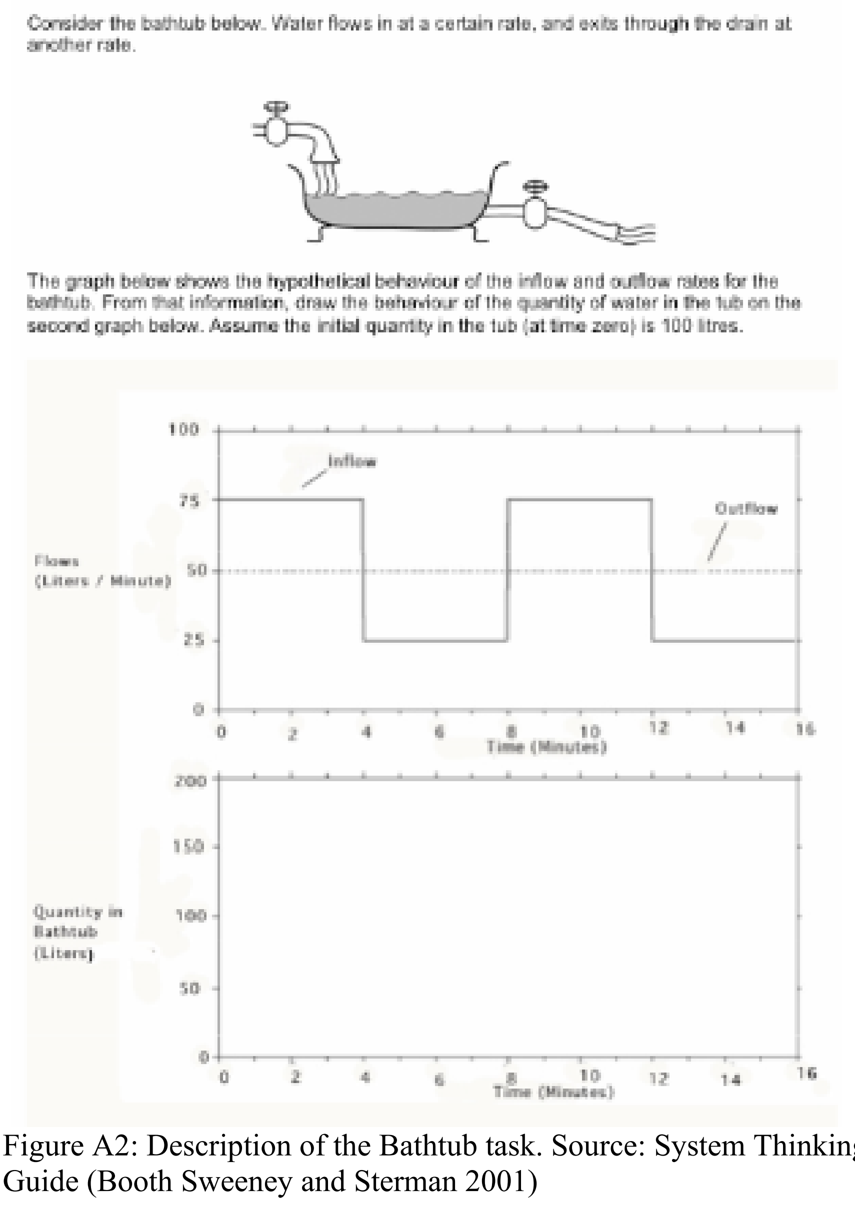


Figure A2: Description of the Bathtub task. Source: System Thinking Inventory - Coding Guide (Booth Sweeney and Sterman 2001)

Figure A3 shows a correct answer.

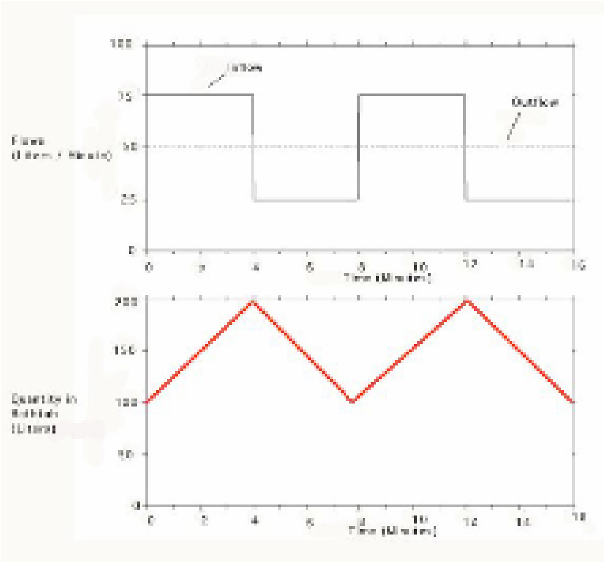


Figure A3: Correct answer for the Bathtub task. Source: System Thinking Inventory - Coding Guide (Booth Sweeney and Sterman 2001)

The following coding rules were applied:

- When the inflow exceeds the outflow, the stock is rising. This means that from time 0 to time 4, and from time 8 to time 12, the graph must show a line that is rising. If the subject shows the quantity in the stock rising during these two time periods but begins at the wrong point on the vertical axis, the response would still be coded as "correct." The response is marked correct (1) on this item if it is rising during both periods 0 to 4 and 8 to 12.
- When the outflow exceeds the inflow, the stock is falling. Between time 4 and 8, and between time 12 and 16, the graph should show a line that is declining.
- The stock should not show any discontinuous jumps (it is piecewise continuous). Discontinuous, refers to vertical jumps, such as steps.
- The peaks and troughs of the stock occur when the net flow crosses zero (i.e., t = 4, 8, 12). The peaks occur at time 2 and 14; the trough occurs at time 8.
- During each segment the net flow is constant so the stock must be rising (or falling) linearly. During the first segment (from time 0 to 4), the inflow is 75 units/time period and the outflow is 50 units/time period, so the net flow is 25 units/time period. The stock grows at a constant rate of 25 units/time period. In the next segment the net flow is – 25 units/time period, and so on. Subjects are marked correct if they show the correct slope in each of the segments, even if the starting point is incorrect.
- The slope of the stock during each segment is ±25 units/time period. During the first segment (from time 0 to 4), the inflow is 75 units/time period and the outflow is 50 units/time period, so the net flow is 25 units/time period. The stock grows at a

constant rate of 25 units / time period. In the next segment the net flow is –25 units/time period, and so on. Subjects are marked correct if they show the correct slope in each of the segments, even if the starting point is incorrect.

- The stock peaks at 200 units and falls to a minimum of 100 units. The quantity added to (or removed from) the stock during each segment is 25 units/time period * 4 time periods = 100 units.

Manufacturing Case task (MC)
The cover story is a manufacturing firm producing widgets at 10,000 units/week with a 50,000 units inventory, which is supposed to be maintained at all times. An increase of orders by 10% occurs in week 5 and remains at this higher level requiring the production to be stepped up. The changes in the production take four weeks to affect, the time delay. Therefore, the inventory will decrease while the production continues on the same level until the output can be adjusted. The adjustment results in a production overshoot followed by a return to stable manufacturing at a higher output. The developments of production and inventory have to be plotted on separate time graphs, see figure A4. Figure A5 shows a correct answer.

The task does not have a uniquely correct answer but a participant could score up to 7 points by covering the following criteria successfully:
The following coding rules were applied:

- Production must start in equilibrium with orders. (0 = no, 1 = yes)
- Production must be constant prior to time 5 and indicate a lag of four weeks in the response to the step increase in orders. (0 = no, 1 = yes)
- Production must overshoot orders to replenish the inventory lost during the initial period when orders exceed production. Production should return to (or fluctuate around) the equilibrium rate of 11,000 widgets/week (to keep inventory at or fluctuating around the desired level). (0 = no, 1 = yes)
- 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 = no, 1 = yes)
- Inventory must initially decline (because production < orders). (0 = no, 1 = yes)
- Inventory must recover after dropping initially. (0 = no, 1 = yes)
- Inventory must be consistent with the trajectory of production and orders (0 = no, 1 = yes).

Consider a manufacturing firm. The firm maintains an inventory of finished products. 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 maintains an inventory of 50 000 units to provide excellent customer service and they adjust production schedules to close any gap between the desired and actual level. Although the firm has enough capacity to handle variations in demand, it takes time to adjust the production schedule and make the product – a total lag of four weeks.

Now imagine that the order rate for the firm's products 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 the production and inventory on the graphs below. Provide an appropriate scale for the graph of inventory.

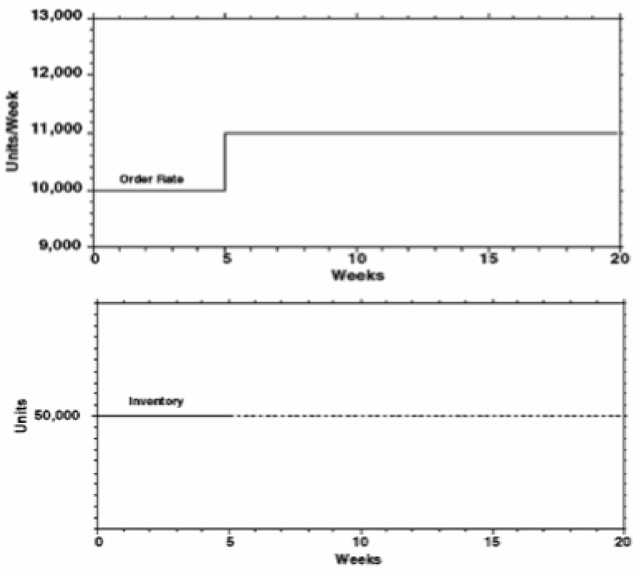


Figure A4: Description of the Manufacturing Case task. Source: System Thinking Inventory - Coding Guide (Booth Sweeney and Sterman 2001)

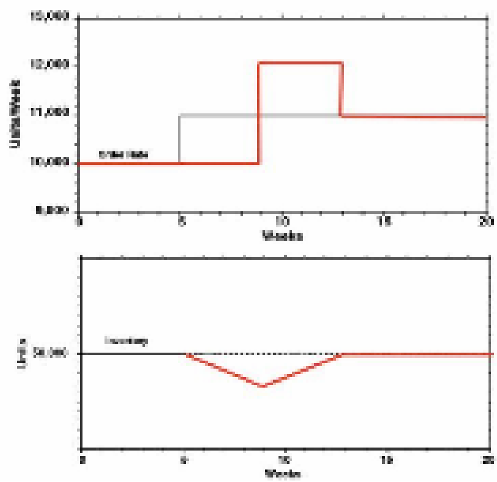


Figure A5: Correct answer to the Manufacturing Case task. Source: System Thinking Inventory - Coding Guide (Booth Sweeney and Sterman 2001)