Supporting decision-makers in managing stock-flow problems:

The effects of oral feedback on reasoning and decision-making

Dennis Jacobs[1*], Inge L. Bleijenbergh[1], Jac A. M. Vennix[1]

[1]Institute for Management Research, Radboud University Nijmegen

P.O. box 9108, 6500 HK Nijmegen, The Netherlands

*corresponding author: d.jacobs@fm.ru.nl, +393897881317

Abstract

This paper broadens the scope of traditional stock-flow experiments because for the first time the focus is both on product (task solutions) and process (reasoning process). By using a think aloud protocol, insight is gained in the reasoning processes of 8 participants in the female professors systems thinking inventory task. This knowledge is used to identify patterns and deficiencies in decision-makers' ability to manage simple stock-flow problems and to propose measures to improve this. Furthermore, the effects of elaborate oral feedback on the quality of participants' reasoning and ability to solve the task are assessed. The experiment shows that the most important reason for failure in managing stock-flow problems is decision-makers' lack of familiarity with the concept of average delay time for determining the outflow of a stock. Furthermore, the oral feedback is found to be effective for creating insight in simple stock-flow problems, although use of supporting tools like a spreadsheet is essential for gaining full insight and for successful management of simple stock-flow problems.

Keywords: female professors task, oral feedback, reasoning processes, stock-flow problems, think aloud protocol

1. Introduction

The ability to understand and manage stocks and flows – resources that accumulate or deplete over time and the in- and outflows responsible for it – is extremely important for comprehending and influencing processes in society, business and personal life (Cronin, Gonzalez, & Sterman, 2009). Over the years however, and in particular since the introduction of stock/flow tasks in 2000, research and experiments have proven that even highly educated people find it difficult to understand and manage even the most simple stock-flow problems (Booth Sweeney & Sterman, 2000; Pala & Vennix, 2005; Cronin, et al., 2009; Strohhecker, 2009, Sterman, 2010).

The bathtub experiment explored highly educated students' understanding of stocks and flows, of how they behave and of the relationships that exist between them. Even though the behavior shown by the stock and flows in the task was thought to be simple and familiar, more than half of the students were unable to answer the questions related to them successfully (Booth Sweeney & Sterman, 2000). Participants failed to recognize the behavior of the stock and the way the in- and outflow influenced it. For instance, most participants erroneously stated that the stock was highest at the moment the inflow was highest and the stock was indicated to be at its lowest level at the moment the outflow was highest. This type of experiment has been repeated many times over the years with different participants and cases, e.g. the 'department store task' (Sterman, 2002), and consistently produced similar results.

The term 'stock-flow failure' was introduced to denote some of these difficulties in identifying the behavior of and relationship between stocks and flows (Cronin, Gonzalez, & Sterman, 2009). The authors labeled one type of mistake often made as the 'correlation heuristic': "it appears that people often use intuitively appealing but erroneous heuristics such as assuming that the output of a system is positively correlated with its inputs. That is, people assume that the output (the stock) should 'look like' the input (the flow or net flow)" (p.117). Correlational reasoning has a negative influence on the ability to understand and solve a stock/flow task and it may therefore be assumed that it can lead to erroneous judgments and decisions in case of real-world stock-flow problems whose structures in general are far more complicated.

Several experiments indicate the need to better educate and assist (future) managers in order to prevent them making these mistakes. One of them is the 'female professors' task (Bleijenbergh, Vennix, Jacobs & Van Engen, 2011), in which participants were asked to balance an initially unbalanced personnel situation. More in particular, starting from a situation with 11% of female full professors and an objective to reach 50% balance, participants were asked to indicate what percentage of female professors they would hire each year and by what time the desired balance would be achieved. Furthermore, they were asked to write down an explanation for their choices. Only two

thirds of participants was able to recommend a hiring percentage that would support achieving the goal stated in the task, and only one third was able to predict more or less when the goal would be achieved. Despite this need, there have been only few experiments on assisting decision-makers while they are working on stock-flow problems. Sterman (2009) took a first step by providing participants with a 'correct' or 'incorrect' comment as feedback on answers. But results showed that this had only a small positive effect and some participants still gave incorrect answers after they were given feedback for the sixth time. This is why we perceive the need to take a new approach in studying stock/flow task performance and to study other and more elaborate forms of feedback.

Therefore, this study attempts to deepen our understanding on the participants' performance in stock-flow experiments, by employing a think aloud approach to gain insight in participants' reasoning processes (as has been suggested in the past by Pala et al., (2005), Cronin et al., (2009), and Sterman (2010)). Doing so, we expect to gain deeper knowledge about how decision-makers reason when solving stock/flow tasks and managing stock-flow problems, what difficulties they experience, how they cope with these problems, and which cognitive biases and heuristics are responsible for the stock-flow failure.

Bleijenbergh et al., (2011) already introduced an intermediate step in the female professors task by asking participants to write down a short justification for their answers. However, this attempt was limited by the fact that only 32% complied with this request, with in general short justifications, that were mostly written after the reasoning process towards the solution had taken place.

We now extend this task by not only introducing the think aloud requirement for observing participants' reasoning processes while performing the task, but also by introducing oral feedback aimed at enabling participants to think in a systemic way, as described by Hopper (2007), about stock-flow problems. Since this form of feedback provides participants with additional information and hints about the structure of the stock-flow problem, their decision-making process should improve and their problem-solving abilities should increase. Therefore, in this paper the following research questions will be addressed:

1) *How do participants in the experiment reason when solving a simple stock/flow task?*
2) *What is the effect of oral feedback on decision-makers' reasoning when solving a stock/flow task?*
3) *How effective is oral feedback to improve decision-makers' ability to solve a stock/flow task correctly?*

The structure of the paper is as follows. We will first discuss concepts on decision making and oral feedback in section 2. The quasi experiment that was used to test these hypotheses will be discussed in section 3. In section 4, the results will be presented. In section 5, the research questions

are answered and practical recommendation will be given. The last section is dedicated to discussion of this study.

2. Concepts

Effective and functional solutions

Earlier research on stock/flow tasks shows that insight in stock flow problems is often low. A variety of tasks has been used to explore cognitive biases people use in solving these tasks (Booth Sweeny and Sterman, 2000; Sterman, 2002). The most simple and most frequently used tasks are the bathtub, the cash flow and the department store task. More complex tasks are the Manufacturing Case, the CO2 zero emissions task (Booth Sweeny and Sterman, 2000; Sterman and Booth Sweeney, 2002) and the female professors task (Bleijenbergh et al., 2011). The latter task showed that few participants are able to select an effective percentage and estimate correctly the time needed to change an unbalanced personnel situation into a balanced one.

In this paper we give feedback to the participants regarding the quality of the solution they had to present by labeling their solution effective / ineffective and functional / dysfunctional (table 1). Solutions are regarded effective when they involve measures that are able to reach the specified goal of the task used in this study. A functional solution would involve measures to achieve the specified goal at the exact time chosen. Presenting a dysfunctional solution decision would mean that the goal would be reached sooner or later than expected, possibly leading to a decrease in motivation and support for the policy that is central in the stock-flow problem used in this study.

	Effective	**Ineffective**
Functional	Goal realized at desired time	Goal cannot be reached
Dysfunctional	Goal realized too early or too late	Goal cannot be reached

Table 1: Solution typology

Oral feedback

This study aims at examining the effect and effectiveness of oral feedback on the quality of decision-makers' reasoning and their ability to solve a stock/flow task. In real-world decision-making, giving and receiving oral feedback on stock-flow problems is common.

The advantage of oral feedback is that the sender of the feedback can fine-tune it to the specific context at that moment. Furthermore, it can be given directly and does not have to be written down, or modeled first. This contributes to a flexible process in which no time is lost.

In this study we put oral feedback into operation by defining it as asking questions related to the elements, structure and behavior of the stock-flow problem and the calculation method described later on (see also Appendix B).

Effectiveness of oral feedback

We consider the oral feedback given in this study to be effective when it enables participants to achieve improvements in their level of insight into the structure and behavior of the stock-flow problem. Simon defines insight as 'grasp of the structure of the problem' (Simon, 1989, p. 14). Applying Simon's definition of insight to this study, this means that when participants gain insight in the structure of the stock-flow problem, they are better able to take effective and functional decisions because they have identified more elements of the stock-flow problem and understand better how these elements behave and interact with each other.

Think aloud protocol

In order to assess the reasoning processes and the effect of the oral feedback on it, it was necessary to have access to decision-makers' reasoning during the performance of the stock/flow task. For this, we used a think aloud protocol. This means that a participant was "required to verbalize thoughts that he himself generated in the course of performing a task" (Ericsson & Simon, 1984, p. 78). Participants were asked repeatedly to verbalize all thoughts they had, every calculation they made, every option they considered and comments on everything they wrote down.

The use of a think aloud protocol allowed for primary data collection about the reasoning of participants, which was important because the oral feedback provided was aimed at supporting participants' insight into the structure of the stock-flow problem that they are faced with. The best way to assess the effect of the feedback, was by asking participants to immediately verbalize how they incorporated it in their reasoning towards solving the task. Furthermore, keeping in mind that "task-directed cognitive processes determine what information is heeded and verbalized" (Ericsson & Simon, 1984, p. 16), participants were required to focus only on solving the task.

3. Methodology

Stock/flow task

The ´female professors´ stock/flow task (Bleijenbergh et al., 2011) served as the basis for the task to be performed in the experiment (Appendix A). To serve the purpose of thinking aloud and assessing the quality of reasoning, it was slightly expanded. The most important change was the introduction of a fixed date (i.e. 2026) to achieve the target of an equal gender balance of professors.

Participants were provided a pencil and paper, a standard calculator and an Excel worksheet they could use to solve the task.

The stock-flow structure based on the reasoning process participants had to follow to solve the problem can modelled as in figure 1[1]. It consisted of two stocks representing the number of female and male professors, with each one inflow, and one outflow. The inflows represented the newly appointed professors who start working in a specific year. The outflows represented the number of professors who leave the university in a specific year. The inflows were determined by the appointment percentage, which had to be determined by the participants and the job openings. Job openings equal the number of professors leaving the university. The outflows were determined by the average delay time, the 10 year working period and the number of (fe) male professors currently employed (for model equations, refer to Appendix C). In this case, the time-period is years and the start time is January 1st, 2011.

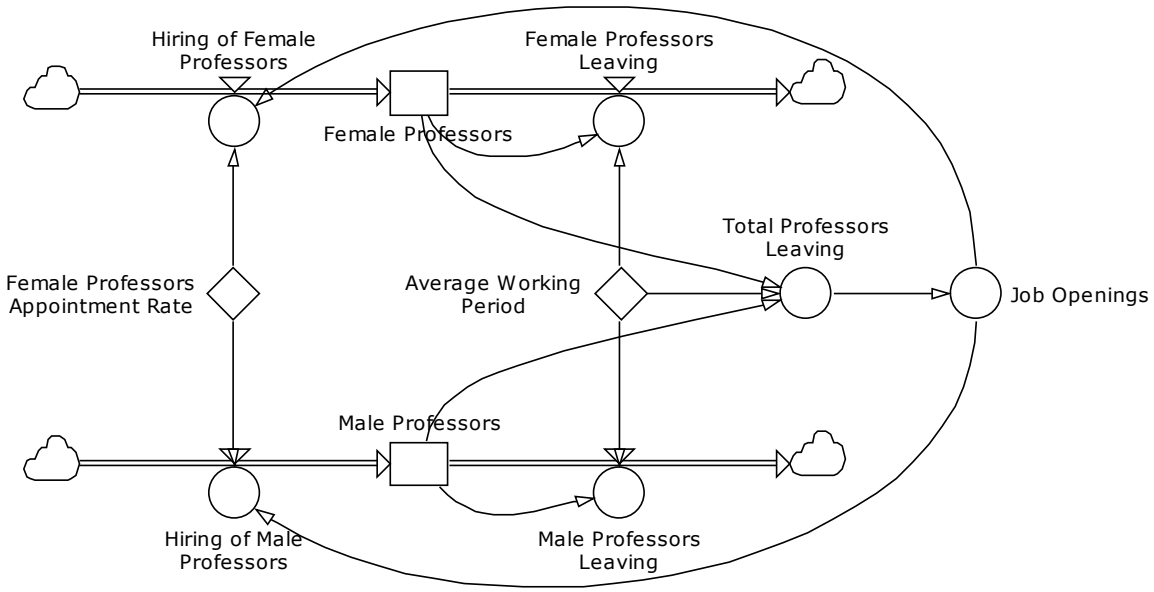

Figure 1: Structure of stock-flow problem

Because we expected participants, to reduce complexity, only to use discrete percentages, we determined the correct percentage for having equal numbers of male and female professors on January 1st, 2026 to be 60%.

Structure of the experiment

[1] Although this stock-flow model violates the rate-to-rate rule, it serves as a representation of the discrete reasoning process participants are expected to follow when solving the task. When simulating this model, a timestep of 1 year is used to represent participants' reasoning timestep, and thus generates the behaviour participants would be able to calculate.

The experiment consisted of two parts. In part one the objective was to gain insight in the reasoning process of participants before any feedback was provided. This part represented the behavior of participants solving stock-flow problems by themselves. The stimulus in part one was the stock/flow task. Participants were asked to think aloud about the case and what elements of it were useful for answering the question. Furthermore, they were expected to reason about the way to solve the task and finally, to determine the correct appointment percentage. Participants had fifteen minutes for this part. We considered a decision to be effective when a participant decided to hire at least 50 percent female professors each year, because, in order to reach a 50/50 representation of female professors an appointment rate of at least 50 percent is necessary.

In part two the objective was to assess the effect and effectiveness of the oral feedback. The stimulus was the oral feedback by means of questions that focused on specific elements of the structure of the stock-flow problem (Appendix B). Depending on how participants reasoned in the previous part and on which elements they already had incorporated in their calculations, we asked questions that would prime participants to think in a systemic way towards the solution of the task. Each question was intended to kick-start this process and a new question was only asked when participants failed to use the previous question to determine the correct appointment rate. Participants were expected to answer the questions and to use that answer, and the feedback that was provided on it, in their reasoning process. Participants had thirty minutes for this part. In case participants designed policies that were better able to achieve the target and that were also closer to the policy deadline of 2026, we assumed that insight was created and that the feedback was effective. When a participant was able to solve the task by choosing the appointment rate that would enable the desired 50/50 representation of female professors, the decision was considered to be functional.

Data collection

The qualitative data were collected by directly observing participants who were performing the female professors stock/flow task. Participants were free to say and do anything they wanted. The collected data were in the form of audio recordings who we transcribed into think aloud reports to allow for an easier analysis process. Because of this intensive method of data collection the number of research entities was kept small (N=8).

Three trial-experiments were conducted to assess the structure of the experiment, the effectiveness of the think aloud approach in this specific setting and the way the oral feedback was formulated. After that eight (under-)graduate students from the Business Administration program at the Nijmegen School of Management were selected by snowball sampling for participation in the final format of the experiment. These participants are well-educated, have some basic knowledge of stocks and flows and are expected to become decision-makers in the near future, so we considered them suitable for observing decision-makers' problem solving abilities.

Think aloud reports

Each experiment was recorded and transcribed afterwards. The transcription was shaped in terms of think aloud reports. Think aloud reports are reports in which: 'the cognitive processes [of the participant], described as the successive states of heeded information, are verbalized directly' (Ericsson & Simon, 1984, p. 16). In the report, verbalized thoughts are presented in such a way that related thoughts that occurred in a short time period are grouped together. The questions asked by the researcher are also in the report, in order to analyze these types of feedback.

4. Data analysis

Focus of analysis

In this section we analyze the results and the think aloud reports. While analyzing the think aloud reports in order to answer research question one, we focused on what information participants deemed important, how they used that information and in which sequence. Furthermore, we paid attention to whether or not general patterns could be identified.

In order to answer research question two, we determined the effect of the oral feedback on participants' reasoning. For this, we used the systems thinking questions provided in Appendix B, in combination with the think aloud protocols. While analyzing, we focused on what the first thing was participants thought and did after being provided with feedback. The data were also used to search for patterns of behavior after the first reaction and for common reactions among the participants.

In order to answer research question three, we determined the effectiveness of the feedback on solving a stock/flow task. We thus compared the outcome of the task in part two with the outcome of the task in part one. For this, we again used the think aloud reports in order to gain more knowledge about which questions created what kinds of insight. The analysis focused on what feedback enabled participants to improve their reasoning process, and which parts of the structure of the stock-flow problem were incorporated correctly in the reasoning and calculation towards the correct appointment rate. Furthermore, ineffective feedback was identified. Finally, whether or not a type of feedback was effective or ineffective was determined for all participants.

Selected appointment rates

As shown in table two, only five participants selected an appointment rate in both parts of the experiment. Their solution in part two was both effective and functional. In the end, they thus choose the effective and functional appointment rate after the oral feedback had provided them with enough insight to understand how to solve the task. However, in part one no participant had chosen the correct appointment rate. Three participants did not know how to determine the correct appointment rate in part two and in these cases, the researcher decided that it was best to stop the experiment because the

time limit was passed and the participants could not focus anymore. Only one participant made an ineffective decision, in part one.

	Participant							
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Appointment rate in part 1 (%)	50	75	65	50	50	80	7.8	85
Appointment rate in part 2 (%)	-	60	60	-	60.2	60	60	-

Table 2: Selected appointment rates in each part of the experiment

These results, especially the decisions made in part one, are somewhat better than the general results on the 'female professors' task conducted by Bleijenbergh et al., (2011). In this study only one participant did not recognize the meaning of the appointment rate, and did not show understanding of the threshold of 50% to achieve an increasing female professor population. In the female professors task 34% did not show understanding of the threshold. This may be explained by the fact that the task was changed to participants only having to decide on the appointment rate instead of on the expected year of goal achievement as well. Another possibility is that the slight textual changes that were made to the case have caused this result.

We now use the research questions formulated in section 1 as a guide for analyzing the gathered data.

How do decision-makers reason when solving a stock/flow task?

Use of information

The case provided all the information necessary but, not all information was identified and used in the reasoning process by all participants. However, it is possible that participants have used information without verbalizing it. In Appendix D all pieces of information from the case that participants have used according to the think aloud reports are shown.

The case contains elements which can be considered essential or unessential for solving the task. As shown in Appendix D, only participants 5, 7 and 8 incorporated all essential elements in their reasoning process. These participants should have been able to successfully solve the task, if they knew the right way of using this information. It is also an indication that full knowledge of what the essential elements in a task are does not mean that the structure and behavior of the stock-flow problem is understood (e.g. participant 8).

Finding the solution

It seemed like there were as many ways to try to solve this task as there are participants. Participants started with different elements and applied different calculations. Furthermore, the sequence in which the information was used varied. Six participants calculated the existing gap. The other two did not do determine it explicitly. Two participants did not use any supporting tool, and only one used the Excel worksheet.

Six participants talked explicitly about mathematics being necessary to solve the task. They either indicated that they had to calculate or come up with a general equation that would include all variables, or they stated that their mathematical skills were insufficient to solve the task.

Main problem

The main problem in the first part of the experiment was that seven participants were unable to determine the correct total outflow of professors:

Participant 2: "The number of female professors to flow in depends on in which period how many people leave. If in one period a lot of men leave because they were all hired in a certain period and have worked for those ten years, and thus leave…"

Participant 3: "Those ten years? of employment, that means... you don't know how long people are already working…"

Participant 4: "You just can't tell when people leave, so if suddenly in some year 117 people quit…"

Furthermore, some participants incorrectly reasoned that after 10 years all professors currently employed would have left because each year a cohort of 30 new professors would replace the last cohort of the previous year:

Participant 1: "When you want to achieve a 50/50 distribution in 2026, you must enforce that ratio in every year prior to 2026." "When a professor works for 10 years, in ten years from now nobody currently employed is still working at the university."

Participant 4: "The goal must be reached in 15 year. On average someone holds his position for ten years. So than in ten years everybody should be replaced."

Participant 5: After ten years everybody would have been replaced."

Because participants could not determine the total professors flowing out each year, six participants did not even speak about the number of female professors flowing out. The reason for these problems is that participants did not know how the concept of average delay time had to be used to determine the outflow. This confirms the conclusion of the previous experiment in Bleijenbergh et

al., (2011, p…..): “This clearly points towards a mental model which uses a pipeline delay or ‘first in first out’ (FIFO) approach”.

Participant feelings

Generally system dynamicists think that the SF tasks are quite simple. However, that is not how it is experienced by the participants. During the entire task, all participants stated to perceive the task as difficult and told that they felt becoming mentally exhausted. This lead to cursing, self-criticism and expressions of despair: “I really can’t solve it.” (participant 1), “What a misery!” (participant 2), “I feel pressured.”, “Very annoying.”, “I feel ashamed.”, “I get crazy.” (participant 3), “I don’t even know where to start.”, “Am I supposed to solve this?”; “I start to feel so stupid that I hardly know what to say anymore.” (participant 4), “May I stop?” (participant 5), “I am stuck.” (participant 6), “It is difficult, I am not a whiz kid.”, “I suck at this, horrible!” (participant 7), “It feels like I am in a math tutoring session.” (participant 8).

What is the effect of oral feedback on decision-makers’ reasoning when solving a stock/flow task?

Participants appreciated the guiding role of the oral feedback and indicated that it motivated them to solve the task. Because of this, participants did not ask continuously for more questions, but tried to proceed on their own as much as they could. An overview of questions used and the elements of the stock-flow problem they are related to, is presented in Appendix D.

Not all questions were asked, because some were deemed unsuitable for the level of insight participants had. This does not mean that they cannot serve as feedback for others.

We asked some questions more frequently for three reasons. First of all, questions, (i.e. question 40: ‘how many female professors are working at the university at 1-1-20xx?’) were answered incorrectly previously and needed to be answered correctly before proceeding in a systematic way. Secondly, questions, (i.e. question 13: ‘how many female professors are leaving the university in 20xx?’) were related to the part of the stock-flow problem that was most difficult to understand. The aim of asking how many female professors would leave the university was to prime participants to incorporate the outflow of female professors in their reasoning and to understand why they could not determine the correct number. The third reason is that question 41 (‘describe what happens with the professors in 20xx’) was considered to be a ‘final’ question. This means that a correct answer to this question was expected to create sufficient insight to solve the task correctly because all elements necessary of the stock-flow problem had to be incorporated. However, since this question was answered incorrectly many times, it can be concluded that we often assumed participants to have more insight in the problem than they actually had.

Sequence of questions

The sequence of questions depended on the progress participants had made in the previous parts of the experiment.

In five cases, we did not ask a question, but provided the participant with information about the answer to a specific question. This was done because those participants were expected to be able to solve the task, although they did not know how to proceed at that specific moment. Participant 1 was provided with a clue about how to determine the yearly outflow. Participants 3 and 8 were told how the outflow could be determined, or what the outflow was. Participants 7 and 8 were told that they had to use unrounded numbers in their calculations.

Because all participants reasoned differently in part one, we could not use a common sequence of questions. Still, a common pattern can be identified. This pattern corresponds to the way in which this task has to be solved. First, the current state of the stock-flow problem has to be known, then the outflow needs to be determined, after that, the inflow can be determined, and the new state of the stock can be verified. Because determining the outflow was the main problem, most questions focused on that part, after which the attention shifted to the future state of the stock. This was an iterative process, depending on the level of insight of participants.

Insight creating questions

In general, two questions and one hint can be said to have the best effect on all participants solving the task. Question 13 (‘how many female professors are leaving the university in 20xx?’) focused the attention of participants on the fact that this number varies and is dependent on the number of female professors employed. Question 41 (‘describe what happens with the professors in 20xx’) focused the attention of participants on incorporating the complete structure and interdependencies of the stock-flow problem in their reasoning. The hint to use the Excel worksheet primes participants to model this structure and to calculate the future values.

Besides these general observations, every participant had its own specific combination of questions that created insight in the structure and behavior of the stock-flow problem.

Participant 1: Relating the average working period of a professor with the total number of professors created insight in the annual average outflow. Asking how many female professors would be employed after one year created insight in the proportion of female professors in the total outflow, although the participant used rounded numbers.

Participant 2: Simultaneously asking how many professors leave each year, combined with the proportion of female professors in the outflow, created insight in how to determine the correct number of female professors leaving each year. The hint to use Excel enabled the structured calculation of future values. Simultaneously asking what would happen in 2011 and how many females would leave created sufficient insight to solve the task.

Participant 3: By asking what would happen in 2011, the insight was created that female professors also leave their jobs, and that each year this is the same proportion as in the population. The hint to use Excel created sufficient insight to calculate in a structured way and to solve the task.

Participant 4: By asking what the average working period was, insight was created in the total yearly outflow.

Participant 5: By asking how many female professors would leave in 2011, insight was created in how to determine the correct number of female professors leaving each year. The hint to use Excel enabled the structured calculation of future values. By asking how many professors would leave each year and reminding that the 33 female professors have an average employment time of 10 years, insight was created in how to determine the correct number of female professors leaving each year. Asking how many female professors would leave in 2013 enabled the participant to solve the task.

Participant 6: By asking how the number of female professors leaving in 2011 was calculated, insight was created in how to determine the number of female professors leaving each year.

Participant 7: By asking how many female professors would leave in 2011, and telling the participant not to use rounded numbers insight was created in how to determine the number of leaving female professors. Asking what would happen in 2011 created insight in the in- and outflow. Asking what would happen in 2012 and 2013 created insight how to determine the stock-flow behavior. The hint to use Excel created sufficient insight to calculate in a structured way and to solve the task.

Participant 8: Giving the total yearly average outflow created insight in the total yearly inflow. Asking how many female professors would leave in 2011 created insight in how to determine the number of female professors leaving each year. Asking what would happen in 2011 created insight in the way to determine the behavior of the stock-flow problem for one year.

Effect on the decision

As shown in Appendix D, the oral feedback enabled five participants to determine the correct appointment rate. Although the oral feedback clearly had a positive effect, its effect was insufficient for creating the necessary insight for all participants. The fact that participants who could not use the oral feedback to their advantage eventually gave up or had to be stopped trying because they ran out of time may indicate that the design and process of giving oral feedback needs to be improved.

In part two, five participants used all support tools available to them. Four participants only started using the Excel worksheet after they had been told explicitly to do so. Two participants indicated not to use the Excel sheet because they did not possess the required skills and did not know what information to put in and calculate.

Participant feeling

The oral feedback had mixed effects on participants' feelings. We assume that because specific questions were asked several times participants' self-confidence decreased and felt pressured: "I just don't get it!" (participant 3), "I can only repeat the answer I already gave" (participant 5).

On the positive side, participants indicated that they felt motivated and inspired by the questions, because it guided their thinking: "I actually start to like this now" "I find it horrible, but I am getting somewhere now. I am heading for triumph!" (participant 3), "all right, wait a second! Now I have a whole new stream of thoughts" (participant 7), "I had not thought about it in this way, brilliant!" (participant 8).

What is the effectiveness of oral feedback for improving decision-makers' ability to solve a stock/flow task?

To answer this question we focus on the feedback and the results in order to assess the effectiveness of the oral feedback. As shown in table 2, the effectiveness of the oral feedback process in terms of creating sufficient insight for solving the task in this experiment can be expressed as enabling five out of eight participants to reason to the correct percentage.

It is more difficult to determine the effectiveness of the individual questions. A total of 86 questions were asked, of which 54 (62.79%) were asked to the five participants who were able to solve the task. Although 41.86% of all questions were answered incorrectly, this does not mean that they were not effective. Because, even incorrectly answered questions provided participants with new motivation for solving the task, for example because they gained insight in how to determine the inflow of female professors given a certain, but incorrect appointment rate. Furthermore, we were provided with insight in the difficulties participants encountered, so that the sequence of questions could be fine-tuned to the participant. What makes determining individual effectiveness even more difficult is that participants gained insight from questions, but could not use that because they lacked the skills to apply it, for example by using the Excel worksheet.

Besides the oral feedback, the hint to use Excel for visualizing modeling and calculating the stock-flow problem that was given to 5 participants proved to be an essential part for enabling them to solve the task because it enables them to make calculations for consecutive years or because they could make interdependencies explicit.

5. Conclusions

In this section we answer the research questions and offer practical recommendations to support decision-makers in managing stock-flow problem in a better way.

Our first question was: *How do decision-makers reason when solving a stock/flow task?* In order to answer this question we analyzed the reasoning of eight Business Administration students during the performance of a stock/flow tasks. We found that they first identify the important information that they assume is relevant to the stock-flow problem. Secondly, they use and combine this information in various ways in order to calculate other variables that they think are necessary to solve the task. They show partial understanding of the structure of stock-flow problems, because they try to incorporate the outflow and the inflow into their reasoning. In general, they do understand the meaning of the appointment rate that is used to manage the balance between two types of a resource. Therefore, they are able to reason towards an effective decision. However, they are unable to determine the outflow when it is presented by means of concepts like 'resources in stock' and 'average time spent in stock'. When decision-makers cannot determine the outflow, they abandon this approach, and finally resort to guessing the rate. Because of this, they take a dysfunctional decision. Most participants did not use an Excel worksheet to determine the correct appointment rate. They told us that they do not possess the skills, or find using a pencil, paper and a calculator more convenient.

Our second question was: *What is the effect of oral feedback on decision-makers reasoning when solving a stock/flow task?* To answer this question we analyzed the effect of the open questions during the performance of a stock/flow task. The effect of oral feedback on participants reasoning was considerable and positive. The think aloud protocols showed they gain new insights into stock-flow problems, which offered them new opportunities to solve the task, and which motivates them to continue their efforts. When questions related to the outflow are asked, participants come to understand how to determine the outflow of stock-flow problems. When this difficult part is understood, most decision-makers are able to determine the behavior of stock-flow problems for the first time period. Only when participants were subsequently asked to repeat those calculations for the next years, they gained complete insight in stock-flow problems. However, they were still not able to solve the task correctly. To do so, they needed to be primed to use an Excel worksheet for their calculations, which helps them, to solve the task, albeit by trial and error. Although participants cannot answer all questions correctly, the questions still seem to be useful to them, because they guide the reasoning process towards the design of an effective and functional policy.

Our third question was: *how effective is oral feedback to improve decision-makers' ability to solve a stock/flow task correctly?* In order to answer this question we evaluated the results of the eight participants that performed a stock/flow task during which they received oral feedback. The effectiveness of oral feedback by means of questions seems to be considerable, because it enables them to gain sufficient insight in the structure and behavior of the stock-flow problem to determine what happens in one time period. Oral feedback by means of questions allows for fine-tuning the feedback to the level of participants' insight. This allows for a sequence of questions that fits both the

way in which the task needs to be solved and the response of the individual decision-makers. However, as we mentioned before, without priming participants to use an Excel worksheet for their calculations, they remain unable to extrapolate this reasoning to future time periods in order to solve the task. When decision-makers possess the skills to work with Excel, and understand the structure of the problem, it is only a matter of trial and error before they take a decision that is both effective and functional.

Practical recommendations

In order to assist decision-makers in the best way possible, it is important to achieve insight in their reasoning. On the basis of our results, we would give the following recommendations:

Description of stock-flow problems

The equation commonly used in system dynamics models for determining the output of a stock (i.e.: outflow = resources in stock / average time spent in stock) is often not known and understood by decision-makers. Therefore, descriptions of stock-flow problems, and especially of how the elements of the problem interact with each other, need to be made as explicit as possible. So, instead of leaving the calculation of the outflow to decision-makers, information on the outflow or instructions on how it can be calculated must be provided to them during or prior to the decision-making through training and education.

Use of oral feedback

It is highly recommended to educate and support decision-makers in managing stock-flow problems by means of asking questions related to the structure and behavior of the problem or task at hand. By doing so, decision-makers are provided with clues about the structure and behavior of the problem and about ways to solve the problem, but they would still need to reason on their own towards sufficient insight and the solution. The main advantage of oral feedback for the educator or consultant is that it can be fine-tuned to the level of insight of the decision-maker, allowing for a flexible and personal approach.

Break down the time period

Breaking down the complete time period of the task into basic time periods, for instance in years, simplifies solving a stock/flow task. Because the calculations are the same each year, decision-makers only need to figure out what happens during one period of time and extrapolate that. This reduces the complexity and makes clear which variables influence the stock that needs to be managed.

Use of supporting tools

Our results show that solving simple stock flow-problems is supported by the use of basic supporting tools like an Excel worksheet. We would recommend stimulating decision-makers to actually use them. Because the elements in the worksheet require input by means of an equation or reference to other elements, interdependencies can be identified easily and insight in the structure and behavior of the problem is gained quicker than in case the tool was not used.

6. Discussion

In this section we will discuss the results of this study and provide recommendations for future research. A first limitation to the results of the study is the snowball sampling method and the small sample size. This could have caused unknown variables to influence and limit the findings of this research. In order to gain more knowledge about the behavior of decision-makers managing stock-flow problems and about the effect of feedback aimed to prevent common reasoning errors future studies should aim for a larger sample size combined with non-random participant selection.

A second limitation is that, despite the use of a think aloud protocol, we might not have gotten all thoughts of the participants. Long silences on the audio recordings might be evidence of this. However, we have gained enough insight from the recordings to answer the research questions and these findings promise to be a good starting point for more focused and structured future research.

A third limitation is the pressure surrounding the performance of the task in an experimental setting. All participants indicated that they felt put under pressure and uncomfortable during the experiment, because they could not solve the task. Participants knew that the answer could be found, but because they were unable to find out how, they got irritated and were not completely concentrated. Although this response may be natural, it may have negatively affected the results. Therefore, in future studies more attention needs to be paid to making the participants feel at ease. Another solution may be to change the role of the researcher, or to remove the researcher from the process, by showing feedback and asking questions by means of computer software.

Future experiments should be more focused on finding reasons for why the oral feedback does not create the same degree of insight in the stock-flow problem for all participants. In this way, the effectiveness of the oral feedback can improve and consequently the required level of insight for managing stock-flow problems can be attained more easily. Both the individual questions and the oral feedback process in general should be re-examined for this purpose.

All participants described the task as difficult, and felt mentally exhausted afterwards. This effect of the experiment can be found in the think aloud reports, in which curses, exclamations of disbelief and despair, and self-criticism can be read frequently. However, when, after the experiment, the solution of the task was explained, participants indicated that it was in fact very logical, and stated that they should have been able to solve it. After this explanation, some participants felt relieved that they now knew how to solve the task, while others felt even a bit stupid.

References

Bleijenbergh, I., Vennix, J., Jacobs, E. & Van Engen, M. (2011). Reducing the gender gap: biases in understanding delay in personnel policies, submitted to System Dynamics Review.

Booth Sweeney, L., & Sterman, J. D. (2000). Bathtub dynamics: Initial results of a Systems Thinking Inventory. *System Dynamics Review , 16 (4)*, 249-286.

Cronin, M. A., Gonzalez, C., & Sterman, J. D. (2009). Why don't well-educated adults understand accumulation? A challenge to researchers, educators, and citizens. *Organizational behavior and Human decision Processes* (108), 116-130.

Ericsson, K. A., & Simon, H. A. (1984). *Protocol Analysis: Verbal Reports as Data.* Cambridge, Massachusetts, United States of America: MIT Press.

Gharajedaghi, J. (2006). *Systems thinking: Managing Chaos and Complexity.* Burlington: Butterworth-Heinemann.

Hopper, M.A. (2007), Proposing measures for assessing systems thinking interventions. UNLV Theses/Dissertations/Professional Papers/Capstones. Paper 282.

Pala, Ö., & Vennix, J. A. (2005). Effect of system dynamics education on systems thinking inventory task performance. *System Dynamics Review , 21 (2)*, 147-172.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference.* Boston: Houghton Mifflin Company.

Simon, H. A. (1989). *Models of thought, Volume 2.* East Greenbush: Yale University Press.

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

Sterman, J.D. (2010). Does formal system dynamics training improve people's understanding of accumulation? *System Dynamics Review, 26 (4),* 316-334.

Strohhecker, J. (2009). Does a Better Understanding of Accumulation Indeed Predict a Higher Performance in Stock and Flow Management? *International Conference of the System Dynamics Society*, (pp. 1-21).

Appendix A: The experiment

Dear participant,

In this experiment you are asked to assume the role of a consultant to the board of a Dutch university. You are asked to provide them with an advice on how to achieve their desired goal. After that, the research team will provide you with feedback on your advice. If your advice is incorrect, you will be provided with feedback aimed at assisting you in giving the correct advice.

While the research team is of course highly interested in what your advice will be, they are also interested in your reasoning process which leads to that advice. We therefore provide you with pencil and paper, a calculator and an Excel worksheet, and ask you to speak out everything that you think, calculate, write down, etc. while you are reading their case and when you are formulating and thinking about the advice that you will give them.

We will record your 'thinking aloud' on tape for two main purposes:

- we would like to be able to listen to the reasoning behind your advice again, and;
- we would like to understand in the best way possible how you came to your advice.

These recordings will be used for research purposes only, and you are presented the possibility to listen to it afterwards. Anonymity is guaranteed in the processing of the data and the use of the data within the research team.

The data gathered in this experiment will be used to contribute to better ways of assisting decision-makers in real-world situations.

Before you proceed to the case, we would like you to answer some general questions:

Gender	Male / Female
Date of birth	
Study	
Year of Study	1 / 2 / 3 / 4
University	

You may now continue to the case on the next page.

Appointing female professors

The Board of a Dutch university has decided to drastically increase the number of female professors. The target is to have 50% female professors at the university in January 2026. At the moment, the percentage of female professors is 11%. For 15 years already, the percentage of female students is around 50%, while 45% of the current PhD students are female. It is expected that this number will only increase in the future. This situation has led to the target being set at 50% for female professors. Considering the financial situation at institutions for higher education the assumption is that the number of professors employed at the university will not increase. The change therefore has to occur within the current number of professor positions.

At the university the number of positions for professors is 300. As stated, it is expected that this number cannot increase in the long run. There are currently 33 female professors and 267 male professors. Once appointed, a professor will hold his or her position on average for 10 years. When a professor retires or accepts a position at another university, a new professor, either male or female, must be appointed. As mentioned, the target is to get 50% female professors in total at the university in 2026.

Within the Board there is discussion about the strategy that has to be pursued to achieve this goal and they are convinced they have to implement a quota. This means that each year a fixed predetermined percentage of the open positions must be filled in by female professors.

The Board asks your advice about what fixed percentage this quota needs to be in order to achieve the 50% female professor in January 2026. This goal should not be achieved earlier or later, because for example budgets and policies will be designed with this goal as a very important underlying assumption. The new policy will be implemented in January 2011.

You may use pencil and paper, a calculator and an Excel worksheet in your reasoning process. Please remember to think aloud during the entire experiment.

You have fifteen minutes for this part of the experiment.

Yearly percentage of female appointments: .....……...%

When you have chosen your percentage, you may continue to the next page.

Now, the research team will assess the validity of your policy. If you have chosen the correct percentage, and reasoned in a correct way, the experiment ends. If not, the experiment continues and you will be provided with various forms of feedback on your policy to assist you in designing and reasoning to the right percentage.

Your percentage and reasoning were:

Percentage: Correct ☐ Reasoning: Correct ☐	Percentage: Correct ☐ Incorrect ☐ Reasoning: Incorrect ☐
The experiment has ended. Thank you for participating!	Please continue the experiment below.

The feedback consists of questions that are intended to help you gain more insight in the process of designing the correct policy. Please reason to an answer to these questions. You will be told whether or not your answers are correct. Please use these questions, your answers and the feedback to them in reasoning why your original policy was incorrect and/or try to reason to the correct percentage.

You may use pencil and paper, a calculator and an Excel worksheet in your reasoning process. Please remember to think aloud during the entire experiment.

You have thirty minutes for this part of the experiment.

Yearly percentage of female appointments: ....……...%

When you have chosen your percentage, you may continue to the next page.

Now, the research team will for the last time assess the validity of your policy. After that, the experiment ends.

Your percentage and reasoning were:

Percentage: Correct ☐ Reasoning: Correct ☐	Percentage: Correct ☐ Incorrect ☐ Reasoning: Incorrect ☐
The experiment has ended. Thank you for participating!	The experiment has ended. Thank you for participating! You may ask for the correct reasoning and percentage.

Appendix B: Oral feedback questions

Case

1. What is the goal of the university board?

Stock

2. How many professors are working at the university at 1-1-2011?
3. How many male professors are working at the university at 1-1-2011?
4. How many female professors are working at the university 1-1-2011?
5. What is the composition of the professors' workforce in % at 1-1-2011?
6. Which factors influence the workforce?

Rate of outflow

7. What is the average working period of professors at the university?
8. Does the average working period of professors' change?

Outflow

9. How many professors leave the university each year?
10. Is the same number of professors leaving the university each year?
11. How many professors are leaving the university in 20xx?
12. How many male professors are leaving the university in 20xx?
13. How many female professors are leaving the university in 20xx?
14. Is the same number of male professors leaving the university each year?
15. Is the same number of female professors leaving the university each year?

Case

16. What happens to the position of a professor who has left the organization?
17. How many vacancies may exist at the university?
18. How many professors need to be employed at any moment?
19. How long does it take before a new professor is appointed?
20. How many professors can be appointed each year?

Rate of inflow

21. What does the appointment rate you need to determine, mean?
22. What does a specific value of this appointment rate mean for the appointment of new professors?
23. Does this appointment rate change over time?

24. Does the effect of this appointment rate change over time?

Inflow

25. How many professors are appointed each year?
26. Is the same number of professors appointed each year?
27. How many male professors are appointed in 20xx?
28. How many female professors are appointed in 20xx?
29. How many professors are appointed in 20xx?
30. Is the same number of male professors appointed each year?
31. Is the same number of female professors appointed each year?

Stock

32. How many professors are working at the university at 1-1-20xx?
33. Which factors influence the composition of the workforce?
34. When does the composition of the workforce change?
35. How many male professors have left the university in 20xx?
36. How many female professors have left the university in 20xx?
37. How many male professors have been appointed in 20xx?
38. How many female professors have been appointed in 20xx?
39. How many male professors are working at the university at 1-1-20xx?
40. How many female professors are working at the university at 1-1-20xx?
41. What happens with the professors in 20xx?

Appendix C: Model equations

Average Working Period
Unit: Years
Value: 10

Female Professors
Unit: Professors
Initial value: 33
Female Professors (t) = Female Professors (t-dt) + (Hiring of Female Professors – Leaving Female Professors) *dt

Female Professors Appointment Rate
? (fraction) (to be determined by participants)

Female Professors Leaving
Unit: Professors / year
Female Professors / Average Working Period

Hiring of Female Professors
Unit: Professors / year
Job Openings* Female Professors Appointment Rate

Hiring of Male Professors
Unit: Professors / year
Job Openings * (1- Female Professors Appointment Rate)

Job Openings
Unit: Professors / year
Total Professors Leaving

Male Professors
Unit: Professors
Initial value: 267
Male Professors (t) = Male Professors (t-dt) + (Hiring of Male Professors – Leaving Male Professors) * dt

Male Professors Leaving
Unit: Professors / year
Male Professors / Average Working Period

Total Professors Leaving
Unit: Professors / year
(Male Professors / Average Working Period) + (Female Professors / Average Working Period)

Appendix D: Experiment Data

Information	Participant							
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Essential elements								
Average working period	x	x	x	x	x		x	x
Current number of female professors		x	x	x	x	x	x	x
Desired number of female professors		x	x	x	x	x	x	x
Desired percentage of female professors	x	x	x	x	x	x	x	x
End of policy	x		x	x	x	x	x	x
Number of professors		x	x	x	x	x	x	x
Start of policy					x		x	x
Unessential elements								
Current number of male professors		x		x	x	x	x	x
Current percentage of female professors		x	x	x		x		
Current percentage of male professors		x			x			
Desired number of male professors		x		x	x			
Length of policy		x	x[2]	x	x[1]		x	x[1]
Number of professors remains the same		x		x		x		
Appointment rate in part 1 (%)	50	75	65	50	50	80	7.8	85

Information provided in the case verbalized by participants (x= verbalized)

Element	Question	Times used	Element	Question	Times used
Goal	1	3	**Outflow**	17	1
Sub-total		*3 (3.49%)*	*Sub-total*		*36 (41,86%)*
Stock	2	2	**Case**	22	1
	4	2	*Sub-total*		*1 (1.16%)*
	5	1	**Rate of inflow**	25	1
	6	3	*Sub-total*		*1 (1.16%)*
Sub-total		*8 (9.3%)*	**Inflow**	29	1
Rate of inflow	7			30	1
Sub-total		*6 (6.98%)*		32	2
Outflow	9	10	*Sub-total*		*4 (4.65%)*
	10	5	**Stock in the future**	40	10
	11	2		41	17
	12	1	*Sub-total*		*27 (31.4%)*
	13	17	***Total***		*86 (100%)*

Used questions for oral feedback

[2] Length of policy was incorrectly identified as 16 years.