

MAIN ARTICLE

Metacognition in action as a possible explanation for stock-flow failureGürsu Aşık^{a*}  and Zerrin Doğança Küçük^{b*} *Abstract*

This study aims at examining the role of metacognition, which refers to one's ability to control and regulate their own thinking process through various activities in assessing dynamics of stocks and flows. The first research question focuses on the metacognitive activities used by individuals who solved such tasks correctly and who did not. The second question focuses on how successfully participants organized their thinking processes to arrive at a correct answer when prompted and permitted to retry. Forty undergraduate students were involved in the study, and concurrent think-aloud protocol was used to examine their thinking when performing two stock-flow tasks. The findings revealed that participants tend to have difficulties in reading, planning, monitoring, and checking activities. The effectiveness of metacognitive activities employed by the participants is likely to decrease from reading to checking, respectively. The study contributes to our understanding of metacognitive deficiencies in stock-flow failure and provides further research suggestions.

Copyright © 2021 The Authors. *System Dynamics Review* published by John Wiley & Sons Ltd on behalf of System Dynamics Society.

Syst. Dyn. Rev. **37**, 253–282 (2021)

Introduction

We are living in an ever-changing world surrounded by a pool of problems that are nested, mostly complex, and highly dynamic. As such, it is difficult to identify variables and their nested relations to make decisions in such dynamic systems. The challenge of interpreting the complexity of dynamic systems lies behind their stock-flow (SF) structures. Understanding stocks (accumulations) and their flows (rates of change) is a fundamental process at many levels of human life. For instance, balancing dietary habits and exercise regimes (Abdel-Hamid *et al.*, 2014), adjusting water levels to account for the differences in flow rates when water flows into and from a bathtub (Sweeney and Sterman, 2000), and deciding on CO₂ emission levels to control atmospheric CO₂ deposit (Moxnes and Saysel, 2009) are all related to identifying stocks and managing their flows in dynamic systems. The concepts of stocks and flows are foundational to systems thinking, and the

^a Faculty of Educational Sciences, Bahçeşehir University, Istanbul, Turkey

^b Department of Education, Maynooth University, Maynooth, Ireland

* Correspondence to: Zerrin Doğança Küçük, Department of Education, Maynooth University, Maynooth Co. Kildare, W23 F2H6, Ireland. E-mail: zerrin.dogancakucuk@mu.ie; Gürsu Aşık, Faculty of Educational Sciences, Bahçeşehir University, Beşiktaş, İstanbul, 34349, Turkey. E-mail: gursu.asik@es.bau.edu.tr

Accepted by Markus Schwabinger, Received 6 November 2020; Revised 18 April 2021; 14 July 2021 and 20 August 2021; Accepted 26 August 2021

System Dynamics Review

System Dynamics Review vol 37, No 4 (October/December 2021): 253–282

Published online in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sdr.1692

development of systems thinking is essential for solving problems that require effective decision-making strategies.

Considering the nature of problems in dynamic systems, the systems thinking literature introduces various failures and oversimplified strategies while solving perceived problems of a system and possible reasons for failures. In this article, there is an attempt to systematically observe problem-solving processes and compare strategies of individuals who could solve tasks assigned to them in an experimental setting with the ones who could not. This study aims to contribute to the systems thinking literature by identifying differentiated metacognitive strategies and to explain possible reasons of failures in solving system tasks.

Cronin *et al.* (2009) define “stock-flow (SF) failure” as the misunderstanding of the relationships between stocks and flows in a system. They identify “correlation heuristics,” that is the tendency of associating behaviors of stocks and inflows, as one form of SF failure. For instance, in their study of people’s decisions for managing the problem of climate change, Dutt and Gonzalez (2012) strongly oppose the general assumption that people’s decisions stemmed from their insufficient knowledge about climatology and climate processes. Rather than this general assumption, they argue that people tended to oversimplify the relationship between CO₂ emissions and CO₂ deposit as if there was a linear relationship between the inflow and the stock, respectively.

Gonzalez and Wong (2012, p. 4) use the term “linear illusion” to describe a primitive problem-solving strategy that people are often taught in mathematics classes in the early grades, where a problem is solved through a linear, simplistic process. The limits of this approach, they argue, is that it has led to people simply preferring linear reasoning at the expense of attending to the complexity of certain situations. As an illustration of this, Sterman and Sweeney (2002) pointed to common misconceptions around the relation between mean global temperature change and atmospheric CO₂ concentration change. Despite popular assumptions, these are not related in a linear fashion, even though similar trends exist between them. In parallel with the example about temperature and CO₂ stock in the atmosphere, there is another problem about understanding the behavior of a system in linear terms although stocks have inertia. In other words, stocks often accumulate over a relatively long time (delay) and their values do not change suddenly (Barlas, 2002) as the case for the atmospheric CO₂ stock demonstrates.

The interconnectedness of system structures (e.g. understanding various nested variables to solve traffic jams) and the existence of both visible and invisible system parts and processes (e.g. identifying groundwater as the invisible component of the water cycle as explained in Assaraf and Orion’s (2005) study) are other characteristic features of complex systems (Hmelo-Silver *et al.*, 2007) that might lead to SF failures. Hmelo-Silver *et al.* (2007) conducted a study on possible differences between novices and

experts in their understanding of complex systems and conclude that most of the differences are associated with the identification of invisible and indirect processes that take place within a complex system. In other words, novices generally tend to focus on direct causal relationships within a system and end up with an incomplete understanding of the system itself.

Previous studies conducted with undergraduate and graduate students reveal that less than half of the participants responded correctly to stock-flow problems (Cronin and Gonzalez, 2007; Gonzalez and Wong, 2012; Ossimitz, 2002; Sweeney and Serman, 2000). Lakeh and Ghaffarzadegan (2015) asked the well-known “department store task” (Serman, 2002) to 400 individuals with different background through an online platform; as expected, the percentage of the individuals who could solve the task was lower than the previous studies with undergraduate and graduate students. In their intervention, they changed the way the task was presented to the sample by devising a prior task for the sample to engage with. This was done with the view to supporting the sample to think analytically for the coming SF task. As a result of the intervention, Lakeh and Ghaffarzadegan (2015) observed an increase in the frequency of correct responses to the SF task. Similarly, Fischer and Gonzalez (2016) used some priming activities to focus on a system (at a global level) in its entirety, rather than through an attention to local details alone. To reflect this, they then modified the department store task by changing its subquestions to relate to the behavior of the system as a whole. Their intervention resulted in a decrease in SF failure for the experimental group (for whom the priming activities and the department-store task were both presented in a global rather than local format).

In their study, Cronin *et al.* (2009) revealed that poor understanding of stocks and flows was due to the problems people have in their decision-making processes rather than their inability to interpret graphs and lack of contextual knowledge about the tasks, motivation, and cognitive load. In the literature, there are a number of studies that make use of a variety of cognitive strategy tasks for understanding the nature of SF failures. For instance, Cronin and their colleagues (2009) conducted several experiments that involved changing the representation of data, the amount of data points on the tasks, and the context of the tasks. Ossimitz (2002) designed and applied six different systems tasks with different contexts and difficulty levels. Gonzalez and Wong (2012) developed authentic interventions by presenting tasks with different structures in varying sequences. In relation to the department-store task, they concluded that introducing a comparison task with two problems exhibiting the same system behavior (behavioral similarity) was more effective in lessening SF failures than introducing a comparison of problems using graphs with similar appearance (surface similarity). The majority of these studies mainly focused on contextual structures of the problem and cognitive strategies people use when solving SF problems. However, to understand the differences in individuals’ abilities in solving SF problems, it is important to examine their problem-solving behaviors and

thinking strategies during the problem-solving process. In particular, rather than just focusing on cognitive and contextual factors, individuals' use of metacognitive thinking processes was identified as a crucial variable for gaining insight into SF failure (Doyle, 1997; Gonzalez and Wong, 2012). Indeed, Cronin and their colleagues (2009) conclude that SF failure is related to the use of inappropriate heuristics, which is closely related to metacognitive functioning.

Flavell (1976), the leading voice in the area of metacognition, defines metacognition as "one's knowledge concerning one's own cognitive processes and products or anything related to them" (1976, p. 262). The definition is conceptualized around the active monitoring and regulation of one's cognitive processes (Garofalo and Lester Jr, 1985; Nelson, 1996). The monitoring function refers to the knowledge that one has of one's cognition, whereas the regulatory function refers to the use of this knowledge to orchestrate ongoing cognitive processes (Flavell, 1979). Individuals' metacognitive skills are shaped by the procedural knowledge needed to regulate and monitor cognition (Van Der Stel *et al.*, 2010). These metacognitive skills emerge throughout the orientation, planning, monitoring, evaluation, and elaboration phases of an individual's cognitive functioning (Veenman and Van Cleef, 2019).

Mandinach and Cline (1994) refer to systems thinking as a problem-solving strategy that accounts for changing components of a dynamic system with the help of models and simulations. While the focus of systems thinking is on problem solving, metacognition is utilized to monitor solution processes and to regulate problem-solving episodes related to exploring, understanding, and analyzing a task, making and implementing a solution plan, and verifying the answer (Schoenfeld, 1992). Schaffernicht and Groesser's (2016) System Dynamics Competence Framework also concentrates on particular skills related to the aforementioned metacognitive processes to learn and teach system dynamics. In light of the relationship between metacognition and systems thinking, it can be concluded that systems thinking requires effective use of metacognitive skills.

In order to make effective decisions in dynamic systems, individuals need to have a clear understanding of the relationship between stocks and flows (Sweeney and Sterman, 2000; Sterman and Sweeney, 2002; Cronin and Gonzalez, 2007; Cronin *et al.*, 2009). In addition to evaluating their decision-making, individuals should engage inner dialogs with themselves, during which they experience metacognition (Costa, 1984). Individuals should create a continuous flow of information about the relevant task by monitoring and controlling the mechanisms through this inner speech (Nelson and Narens, 1990). At this point, the reason for making poor decisions in system dynamics may be due to the lack of information, as well as the fact that metacognition cannot easily process information and bring those activities to the status of automatic skills in challenging new situations. Thus, it is

worthwhile focusing on individuals' use of metacognitive skills while solving stock-flow problems.

This study aims to examine the role of metacognitive skills in solving stock-flow tasks. Metacognition in this context refers to the effective use of strategies for understanding the problem and designing, monitoring, and executing an effective plan and also evaluating the possible solution during stock-flow tasks. Individuals who perform these steps effectively are expected to be successful in interpreting the stock-flow tasks and constructing a solution. The study attempts to address two questions in this manner. The first focus of the study is based on the proposition that individuals who solved stock-flow tasks correctly would differ in their use of metacognitive activities compared to those who engaged with the tasks incorrectly. It is expected that successful problem solvers exhibit a higher level of metacognitive skillfulness relative to unsuccessful problem solvers. In accordance with this purpose, we aim to reveal the differences between successful and unsuccessful individuals' metacognitive activities while working on SF tasks. Secondly, the study examines the prospective problem-solving processes of the individuals who were not able to give the right answer in their first attempts. After their first attempts to solve the problem, the participants, who offered an incorrect solution, were made aware of this by the facilitator. Immediately afterwards, extra time was given to the participants to rethink and resolve the problem and then participants' metacognitive processes while working on the given problem were examined. This intervention is similar to the approach taken by Lakeh and Ghaffarzadegan (2015) to understand the cognitive processes of participants after an initial change in response. We aim to determine which metacognitive activities have important roles in the solution-correction process and to show how metacognition leads participants to identify and correct mistakes after an external regulatory support (such as the facilitators' prompt) is provided. In summary, the research expects to gain detailed information on individuals' use of metacognitive skills while solving SF problems.

Method

Study group

The data was collected at Boğaziçi University, through its Undergraduate Program in Mathematics Education. It is the program that requires the highest score in the OSS examination (University Entrance Examination) among education-focused departments in Turkey. Eighty-two junior and senior students who were enrolled in this program were invited to participate. Participation was voluntary and no incentive was given. Forty

participants took part in the study with a 2.86 ($SD = 0.36$) grade-point average over 4.0. Gender distribution was formed as 22 women (55%) and 18 men (45%). The participants' ages ranged from 22 to 28 years old.

Data collection

Think-aloud protocols

In the study, a single-subject concurrent thinking-aloud technique was used to generate verbal protocol data (Fonteyn *et al.*, 1993). Concurrent protocol analysis is considered to be a valid source of data on thinking (Ericsson, 2006) to elicit verbal reports of thought sequences reflecting participants' short-term memory without altering their cognitive process (Gero and Tang, 2001). In concurrent think-aloud techniques, the subjects are required to verbalize their thoughts simultaneously while performing a specified task.

The verbalization of thoughts while working on a task is acknowledged as one of the most effective methods for gaining insight into individuals' metacognition (Schraw, 2010). In addition to this, think-aloud protocols are recognized as one of the most promising techniques for collecting and analyzing data on dynamic decision-making (Doyle, 1997). Hence, the think-aloud technique was chosen to gather data about participants' metacognitive skills in this study. A systematic observation checklist (Veenman *et al.*, 2000) was utilized on the scoring of this data. The participants' behaviors and discourse while solving stock-flow tasks were assessed through systematic observations inspired by the systematical-observation checklist (Veenman *et al.*, 2000) and scored in relation to the occurrence of metacognitive activities.

For this study, the metacognitive skillfulness systematical-observation checklist (Veenman *et al.*, 2000) was modified based on the structure of the tasks. In its original version, 12 metacognitive activities under three phases were identified in the checklist. The activities in the checklist are characteristics of metacognitive skillfulness particularly for math exercises (Schoenfeld, 1985; Van der Stel *et al.*, 2010). In this study, based on the observations made in the pilot study, four activities for the first task and two activities for the second task could not be observed due to the nature of the tasks. The metacognitive activities on the systematical observation checklist are presented in Table 1.

On the systematical-observation checklist, Veenman *et al.* (2000) identified three phases and the corresponding metacognitive activities as represented on Table 1. Activities 1 through 5 represent participants' preparation and orientation to the problem before acting, activities 6 through 10 express systematical orderliness presented during task performance, and lastly, activities 11 and 12 emphasize the evaluation activity during and after problem

Table 1. The metacognitive activities in systematical-observation checklist for stock-flow tasks

Activities	Criteria	Department-store task	Bathtubtask
	Phase 1: Orientation		
1	Entirely reading the problem statement	✓	✓
2	Selection of relevant information needed to solve the problem	✓	✓
3	Paraphrasing of what was asked for	✓	✓
4	Making a drawing related to the problem	X	X
5	Estimating the possible outcome	X	✓
	Phase 2: Task performance		
6	Designing an action plan before actually calculating	✓	✓
7	Adhering to that plan	✓	✓
8	Calculation outcomes	X	✓
9	Avoiding negligent mistakes (such as switching number)	✓	✓
10	Orderly note taking of problem-solving steps	X	X
	Phase 3: Evaluation		
11	Monitoring the ongoing problem-solving process	✓	✓
12	Checking the answer	✓	✓

The cells with a sign “X” represent the metacognitive activity eliminated for the task.

solving (Veenman *et al.*, 2000). The scoring range for each metacognitive activity was 0 to 2. A score of 0 was given if the activity was absent. One point was given if the activity was initiated, but not completed. Two points were given if the activity was clearly presented. Further scoring examples can be seen in the findings section.

Before data collection, a pilot study with two university students took place. During the pilot study, we practiced this systematical observation procedure on given tasks with the randomly selected two university students. The metacognitive activities and predicted overt behaviors for each phase were defined during the pilot study based on the literature. By aiming to fairly evaluate each session, the detailed rubric of the systematic-observation checklist was formed. Considering the importance of capturing individuals’ actions along with verbal expressions, a video-analysis technique instead of verbatim transcription was used in the analysis. Video-analysis techniques extend the depth and richness of analysts’ work by enabling them to virtually revisit the studied scene repeatedly and in this way gain greater insight of events that took place (Markle *et al.*, 2011). All recorded videos were watched together by both analysts and judged collaboratively with the help of low-inference notes taken. When writing up the findings of the study, we intended to capture participants’ actions and

voices through thick descriptions. Thick description refers to analysts' task of describing and interpreting observed behavior within its particular context (Ponterotto, 2006).

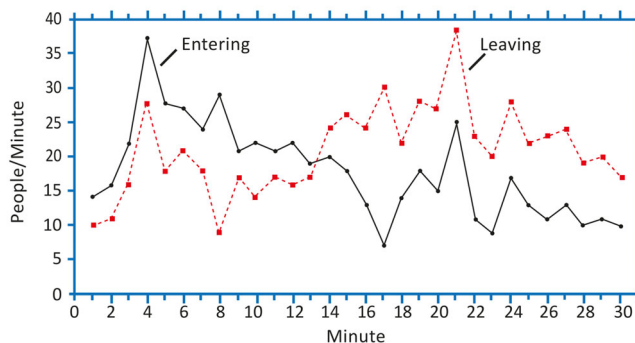
Procedure

The two common stock-flow tasks were administered in individual sessions throughout a three-week period. There was no particular time restriction for each session, though participants were expected to complete the tasks both within 30 minutes. Although we took notes about each participant's problem-solving process, the sessions were videotaped with participants' written permission for us to revisit the study in order to strengthen the validity of the analyses. During the problem-solving sessions, all interactions between the facilitators and the participants were kept to a minimum in order to avoid interfering with participants' flow of thoughts. In cases where participants paused for longer than 3 seconds, we would remind them to keep thinking aloud.

At the beginning of the data collection, the participants were informed by reading the instructions about the procedure. As a first problem, the department-store task was presented (Figure 1). Whenever a participant

Fig. 1. The department-store task (Sterman, 2002)
[Color figure can be viewed at wileyonlinelibrary.com]

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

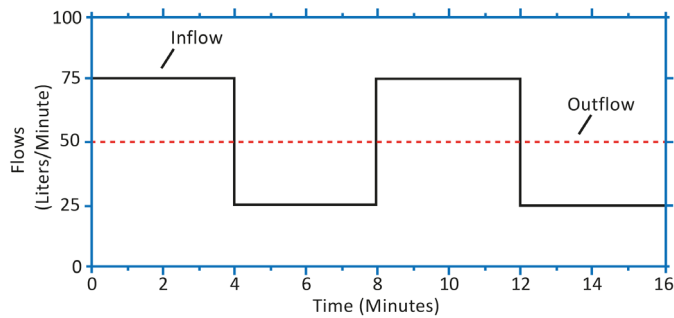


Please answer the following questions.

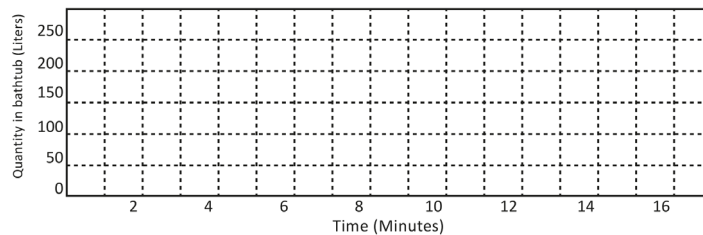
Check the box if the answer cannot be determined from the information provided.

- During which minute did the most people enter the store?
Minute _____ Can't be determined
- During which minute did the most people leave the store?
Minute _____ Can't be determined
- During which minute were the most people in the store?
Minute _____ Can't be determined
- During which minute were the fewest people in the store?
Minute _____ Can't be determined

Fig. 2. The bathtub task (Sweeney and Sterman, 2000) [Color figure can be viewed at wileyonlinelibrary.com]



Consider a bathtub. Water flows in at a certain rate, and exits through the drain at another rate. The graph below shows the hypothetical behavior of the inflow and outflow rates for the bathtub. From that information, draw the behavior of the quantity of water in the tub on the second graph below. Assume the initial quantity in the bathtub (at time zero) is 100 liters.



indicated that they completed the task, the facilitators then provided feedback on whether their solution was correct or not. The feedback was limited to “correct” or “incorrect” only, with no other information being provided. In the case of an incorrect response, the participants were allowed to think about the task again. Their ongoing metacognitive activities in the correction processes were reassessed. After the first problem was completed, the bathtub task (Figure 2) was presented as a second problem, and the same steps as the first were followed.

Stock-flow tasks

Two stock-flow tasks were presented to the participants. The questions were kept as their original versions. The first task, which is called the “department-store task” (Sterman, 2002), includes a graph showing the number of people entering and leaving a department store each minute over a 30-minute interval (Figure 1). The system involves a single stock (the number of people at the store) with one inflow (people entering) and one outflow (people leaving).

The second task, which is called the “bathtub task” (Sweeney and Sterman, 2000), aims to infer behavior of the stock from information on the flows. The task is among the simplest possible examples of stock-and-flow

Table 2. The descriptive statistics about the solutions for each task

	Department-store task			Bathtub task		
	1st attempt correct	2nd attempt correct	Response incorrect	1st attempt correct	2nd attempt correct	Response incorrect
Number of participants (<i>n</i>)	6	8	26	20	7	13
Percentage (%)	15%	20%	65%	50%	18%	32%

thinking. The outflow is constant, and the inflow follows a simple pattern. The task is shown on Figure 2.

Performance on the stock-flow tasks

The performance of this particular study group is summarized in Table 2. The given descriptive statistics are useful to identify successful and unsuccessful problem solvers on the tasks. The table also indicates at which attempt the participants reached the correct solution.

In the department-store task, six people (15%) gave a correct response to the four subquestions correctly at their first attempt, while eight (20%) were able to answer the question correctly at their second attempt. In addition to this, 26 participants (65%) did not offer correct answers to all subquestions in this task. The majority of the participants were able to give correct answers to the first two subquestions (subquestions *a* and *b*) which focus on flow by asking participants to determine the times where most people would enter and leave the store (88% and 78%, respectively). On the other hand, few were able to give correct answers to the stock questions (subquestions *c* and *d*) which asked participants to determine the times where the fewest and most people would stay in the store (28% and 25%, respectively).

In the bathtub task, there were more correct responses compared to the department-store task. Half of the participants ($n = 20$) solved the task at their first attempt, and seven more participants (18%) correctly determined the quantity of water in the bathtub at their second attempt. The remaining 13 participants (32%) were not able to find the correct answer on the given task.

Findings

In this section, a comprehensive analysis of the participants' metacognitive skills while solving the two stock-flow tasks is presented. First, findings on the use of metacognitive skills in each particular task are reported descriptively. Then, a detailed qualitative analysis is presented for each stage of problem solving classified by Veenman *et al.* (2006). To present a

Table 3. Relative ratios of each metacognitive skill on the assigned tasks

Scores	Department-store task			Bathtub task		
	0	1	2	0	1	2
Orientation						
Reading	2.5%	7.5%	90%	2.5%	20%	77.5%
Selection	12.5%	30%	57.5%	17.5%	25%	57.5%
Paraphrasing	22.5%	17.5%	60%	40%	10%	50%
Estimating	Not Observed			85%	2.5%	12.5%
Task performance						
Designing a plan	20%	52.5%	27.5%	22.5%	20%	57.5%
Adhering to plan	32.5%	40%	27.5%	22.5%	25%	52.5%
Calculation outcomes	Not observed			5%	5%	90%
Avoiding mistakes	2.5%	22.5%	75%	20%	25%	45%
Evaluation						
Monitoring	40%	40%	20%	32.5%	20%	47.5%
Checking	90%	5%	5%	60%	22.5%	17.5%

quantitative picture for each metacognitive skill, three levels were determined: 0 indicates no performance of the skill; 1 indicates partial performance of the skill; and 2 indicates complete performance of the skill. Table 3 presents different levels and relative ratios of each metacognitive skill at these levels. As mentioned in Table 1, some skills could be observed across both assigned tasks, and they are indicated with check marks, while other skills could not be observed due to the nature of the tasks given, and they are indicated with “X”s. Examples include “estimating” and “calculation outcomes,” which were not relevant to the department-store task.

In the following sections, the observed metacognitive activities are critically examined based on the individual responses of the participants. Additionally, some common and unique responses addressing the particular metacognitive activities during the problem-solving sessions are exhibited with variations identified among successful and unsuccessful problem solvers in understanding, interpreting, constructing strategies, and monitoring their own processes.

Successful and unsuccessful problem solvers are denoted as SPS and UnSPS throughout the text, respectively.

Analysis of the department-store task

Orientation

The department-store task can be considered as a difficult task with continuously changing flow patterns. For the orientation part, the first two subproblems focusing on flows by asking the participants to determine the time at which most people entered and left the store were taken into account

(Figure 1). Reading the text fully and carefully, the first metacognitive step in problem solving, was accomplished by 36 participants (90%). Only four participants (10%) skipped reading the text in detail and moved on to the subquestions. Among these four participants, one UnSPS did not spend any time in reading the question but started to translate the question into Turkish and attempted to explain the graph. In their second attempt, this participant did not read the first sentence again and asked whether the subquestions (a) and (c) were the same. This was not the only participant who was confused about the two subquestions. By contrast, one participant read the question fully and loudly and then said that:

Excerpt 1: Entirely reading the problem

“I did not understand the question completely. I would like to read it again.”

This was one of the SPSs who could solve the department-store task in their first trial. As is exemplified in the given quote above, this participant was willing to comprehend the problem at first and tried to figure out the problem by repeating the distinctive phrases of the text several times.

The activity of selecting relevant information was closely related to reading the graph precisely. Five participants were unable to select the relevant information to solve the department-store task. For instance, one of the participants explained that:

Excerpt 2: Deficiency in selection of the relevant information

“The highest number of people in the store is at the 4th minute, because it is the minute that the highest number of entering takes place.”

This UnSPS assumed that only entering affects the total number of people gathered in the store and selected the irrelevant information initially. Although the graph on the task is clear and exact, six participants (15%) included the phrases “nearly” or “approximately” in their answers for the first two subproblems. In fact, these participants could not achieve an exact answer by using those approximations. Only one of these six participants emphasized the exactness of the answer at their second attempt. The remaining 29 participants (73%) were successful in selecting the relevant information and responded to the two related subproblems correctly. Among these participants, only one of them emphasized the focus as “it asks to find out exactly at which minute it took place.”

Another UnSPS spent some time figuring out a piece of irrelevant information to solve the task. The excerpt below is from their second trial:

Excerpt 3: Deficiency in selection of the relevant information

“The [number of] people entering and leaving are given. But the number of people in the store is not given... (This participant still insisted that) We do not know exactly how many entered and left.”

Another UnSPS spent a few minutes clarifying whether the intersection points of the entering and leaving graphs were exact, deciding that there was no need to mention “approximately” in their response.

A distinctive phrase in the problem (“over a 30-minute period”) offered participants a potential clue to the solution, as it implied that the department-store task would take place over an accumulated period of time. Only three participants demonstrated awareness of this phrase while solving the task. Two participants repeated the phrase “over a 30-minute period” a few times during reading the task while the other participant paraphrased “the difference throughout the minutes.” These three participants were among the six SPSs who could solve this task at their first trials.

Task performance

Task performance, which is the phase where an individual takes actual steps to solve a problem by using their own knowledge and skills, starts with a planning activity. The subproblems (c) and (d) in the department-store task reflected a sense of planning. During problem-solving sessions, three different levels of planning were identified. Level 0 plans included unorganized, incoherent, and false chunks of information. In this task, eight participants (20%) misinterpreted the graphs and constructed their plans on just one flow (either leaving or entering). Level 1 plans were the most frequent; 21 participants (52.5%) grounded their answers in the differences between the two flows at their first trial. Level 2 plans included strategies like summing up the differences of the flows at each minute or comparing the areas covered by the flow curves (namely, accumulation of the people in the store over 30 minutes). Eleven participants (27.5%) were able to design their plans at Level 2. Among these 11 participants, nine participants succeeded in solving the problem. While a SPS spent a considerable amount of time adding the inflows and subtracting the outflows at each minute, the remaining eight were able to identify the breakpoint. A few selected responses for subproblems (c) and (d) addressing each level of planning are placed in Table 4.

With respect to planning, although more than half of the participants’ plans were at Level 1, only one participant adhered to their Level 1 plan and made calculations at each minute. During their calculations, this participant also questioned whether there should have been a simpler solution.

An UnSPS changed their plan while explaining it as follows.

Table 4. Levels of planning skills with related quotations for the department-store task

Levels of planning	Level 0: unorganized plans or one-flow oriented plans	Level 1: plans focused on two flows	Level 2: plans focused on accumulation over time
Selected responses	d) The moment when the highest number of people left the store.	c) The moment when the difference is the highest. d) The moment when the difference is the lowest.	c) I have to shade the area below the graph [shading]. Up to here it is increasing, then decreasing. This is the breakpoint [showing the point]. c) We should add up the differences as cumulative. c) After this point [the breakpoint], the number of people leaving is higher.

(c) and (d) stand for the third and fourth subproblems of the department-store task.

Excerpt 4: Changing the plan

“Maybe we need to consider the number of people entering and leaving at the store.

(The participant said so, however, did not stick the Level 1 plan and decided to convert the plan into a one-flow oriented Level 0 plan.)

...I have to think the other way around (for the subquestion d) Because the number of people who left the store is the highest (pointing to the 4th minute).”

There are some instances where changing the plan resulted in correct solutions for the task. Eight of the SPSs started with a Level 1 plan but then successfully identified the breakpoint on the graph. After that, they focused on the two areas separated by the breakpoint on the graph. Although time recording was not a focus of this study, the participants who successfully identified the breakpoint seemed to spend relatively less time in completing the task as soon as they identified the breakpoint.

The instant when the problem solvers realized the breakpoint was also crucial in terms of metacognitive questioning and self-talk. For instance, one SPS figured out the breakpoint while responding to the third subproblem. This participant mentioned the breakpoint as “the point where entries and departures change” and continuously emphasized the breakpoint and articulated self-assurance when they said “Yes, this is the point.”

Another SPS explained that their strategy was based on the breakpoint.

Excerpt 5: Adhering to the plan

“People start to enter and it is equal at this point (showing the breakpoint). Number of people entering is higher up to this point. And, after this point, number of people entering will decline.”

It should be noted that the explanation also implies a sense of accumulation rather than mentioning discrete time intervals. The interesting point was that some participants explained their strategies in a metacognitive manner, while some completed the task soon after they realized the breakpoint.

Another critical instant was when the participants realized that people were accumulating in the store. Among the participants, nine of them (22.5%) mentioned accumulation explicitly, while five out of the nine failed to solve the task due to time-consuming additions over minutes. Only three participants calculated the difference between the number of people leaving and entering at each minute and added them up one by one patiently. During their calculations, these participants realized the increasing trend of the stock of people at the store and the declining trend after the 13th minute (breakpoint). At that instant, they were able to complete the task successfully.

For instance, one SPS, who successfully solved the task at their first trial, felt dissatisfaction with the on-going plan and also displayed monitoring skills:

Excerpt 6: Changing the plan

“It (the graph) can be misleading. I am just focusing on the extremes (in the graph) (Then, noticing the accumulation) ...We will add the differences... cumulatively.”

It is important for problem solving to be aware of possible errors made and to progress in a controlled manner to avoid these. For instance, a few UnSPS insisted on mentioning intervals rather than exact minutes for subquestions (a) and (b) after developing their plans. These participants were scored 0 on Table 3. Some minor mistakes, such as responding “23” to the subquestion (b) due to misreading the graph, were scored as 1 in Table 3.

Evaluation

The skill of monitoring is expressed in how one validates one’s comprehension of a task and checking the solution steps. For the department-store task, eight participants clearly demonstrated high-level monitoring activities. Among them, five were SPSs. There was one SPS who solved the task very quickly, and unfortunately, we could not observe any monitoring steps in their session (Table 3). Although three participants were considered as

having high-level monitoring skills, they could not solve the task due to minor mistakes. The problem-solving strategies of 16 participants were evaluated as low-level monitoring while the remaining 16 people did not show appropriate monitoring activities at all.

An example of the high-level monitoring activity could be seen when a SPS commented on the task as follows.

Excerpt 7: High-level monitoring

“As long as the number of people entering to the store is much more than the number of leaving people, that means the entering line stays over the leaving line on the graph (showing the graph and the trends) and the number of people inside the store should increase.... On the contrary, the number of people should decrease. Therefore, the last minute on the graph (pointing the 30th minute) should indicate the time fewest number of people in the store.”

In this excerpt, the SPS gave an adequate and consistent explanation to their plan at the task performance stage. We can also identify high-level monitoring in the excerpt below which includes clear evidences of self-talks of a SPS:

Excerpt 8: High-level monitoring

“The highest number (of people) is at 13th minute (responding to subquestion c). Because after 13th minute, number of people leaving is higher and the number of people in the store is decreasing.... I consider whether there are any people before this minute (pointing first minute)... The first area indicates an increase while the second area indicates a decrease. The second area is bigger so it means there are already some people inside. Then, it is 30th minute for question d.”

As exemplified in the two excerpts above, two SPSs demonstrated a validation process by monitoring the correctness of their cognitive operation. As emphasized before, most of the participants started to search for a solution by trying to find the minute when the difference between the number of people entering and leaving the store on the chart was the highest. A few of them realized that the solution to the task was linked to accumulation. The ones who reevaluated their solution process and figured out where the real solution lies were also considered as having high-level monitoring skills.

Some of the participants who gave incorrect answers emphasized their doubts about the accuracy of results with statements such as “The answer is..., but I am not sure,” “probably I have a mistake,” and after spending some extra time on the task, “I know the answer is not correct, but this is my answer.” They were aware of the fact that they made mistakes during problem solving, but they could not take a step to come up with a new solution.

This implied that they have metacognitive monitoring skills even if it was low-level monitoring.

The third and final phase of metacognitive skillfulness is considered as evaluation, and it includes the monitoring and checking the correctness of the given result. Although we had the opportunity to make clearer observations about participants' monitoring strategies, it was difficult to observe participants making clear statements in checking the correctness of their solutions. One of the clear excerpts we identified for checking activity was as follows.

Excerpt 9: Checking the solution

“There is no need to evaluate deeply; the total number of people is the highest when the number of people leave the store is the least.”

This UnSPS was aware of the fact that they had a dissatisfying response but failed to identify the potential error in the solution and, in a sense, justified the lack of checking by the statement. Most of the participants did not include checking in their thinking process during cognitive activity, and this was considered as a deficiency in the thinking processes of most of the participants. The same is also valid for the second task.

Analysis of the bathtub task

The bathtub task is rather simple compared with the department-store task with one constant outflow and one inflow with a graph having a square wave pattern. Twenty participants (50%) were able to solve the task at their first trials, and seven more participants were able to solve the task at their second trials as given in Table 2. In contrast to the department-store task, we were able to observe some estimation processes while solving the bathtub task. Estimating possible outcomes is one of the metacognitive skills in the orientation subscale of the observation checklist (Veenman, 2006). Although participants did not make any estimates while solving the department-store task, six participants (15%) made estimations for the bathtub task. Among these six participants, only one participant took partial credit for one's estimation since it was estimation about a limited time interval in the task.

Excerpt 10: Partially estimating the possible outcome

“It (the level of water) will rise since the graph is constant (indicating the first four-minute time interval).”

One important finding is that, five participants, who took full credit for the skill of estimation, were able to solve the task at the end of the problem session. Most of the estimations were found to be related to the behavior of the

graph given in the task. Three of the estimations were listed as the excerpts below.

Excerpts 11–13: Estimating the possible outcome

“It will be a linear graph, because it is a first order function.”

“Since the flows are constant, there is no need to draw something parabolic.”

“Since the behaviors (of inflow and outflow rates) are symmetric, it (the amount of water) will stay the same (at the end).”

In the Excerpt 13, the SPS meant that the trend of the inflow rate is symmetric with respect to the trend of the outflow rate at each four-minute interval; hence the areas in between inflow and outflow are eliminated.

There was only one participant who made two estimations on the single task.

Excerpt 14: Estimating the possible outcome

“There should be an incline because the water entered is much more than the water left.” (This was the estimation about the interval referring to 0–4 minutes on the graph.)

“*Since the water inflow is more than water outflow, there will be a rise (in water level).*” (The second estimation relied on the shape of the whole graph and was about the amount of water at the end of the 20-minute time period.)

Paraphrasing is a key metacognitive skill which could be used to exhibit participants’ understandings of problems. Thirteen participants could not solve the bathtub task even at their second trials. Among these UnSPSs, two of them made clear explanations about the first graph given. These explanations actually revealed the underlying reasons about why they could not solve the task.

Excerpts 15-16: Deficiencies in paraphrasing

“75 liters of water flowed at the end of four minutes.”

“50 liters of water came inside (the bathtub) along 16 minutes.”

As indicated in the Excerpt 16, this UnSPS seemed to calculate the net flow as 50 L but did not concentrate on the amount of water being accumulated across all of those minutes. These statements implied misinterpretation of the graphs on the part of the participants.

By contrast, the SPSs mentioned the amount of water per minute. Both of the excerpts below can be considered as indicators of comprehending.

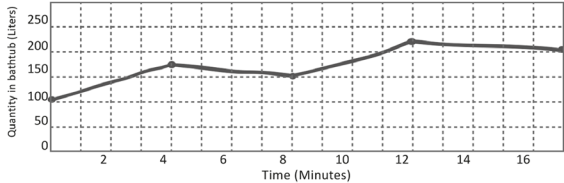
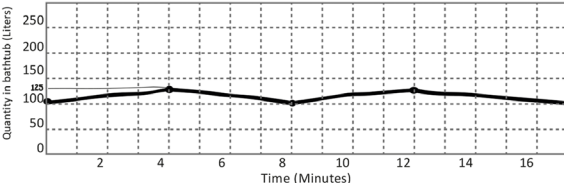
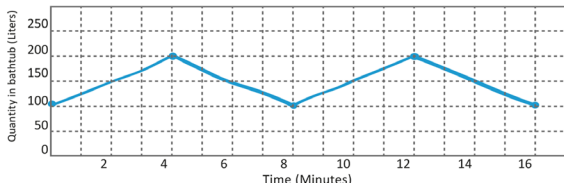
Excerpts 17-18: Paraphrasing

“We should talk for each minute.”

“25 liters of water is added at each minute. At the fourth minute, it makes 100 liters of water.”

The same levels of planning in the department-store task were also designated for the bathtub task. The corresponding oral responses and the drawings are placed in Figure 3 for each level of planning. It was easier for us to observe the planning phase of this task due to existing symmetric patterns given on the graph. Although one participant was able to solve the task with basic understanding of calculus, nine participants (22.5%) were not able to draw the graph at their first trial. Among these nine participants, eight of them did not include the outflow (amount of water exiting through the drain)

Fig. 3. Levels of planning skills with related excerpts and relevant graphs created for the bathtub task [Color figure can be viewed at wileyonlinelibrary.com]

Levels of Planning Skills	Selected Sample Drawings	Corresponding Excerpts
<p>Level 0: Unorganized plans or plans focused on one flow</p>		<p>“75 liters of water entered up to the fourth minute [never mentioned about the outflow]...”</p>
<p>Level 1: Plans focused on two flows</p>		<p>“We will start with 100... 75 [liters of water] entered, 50 [liters of water] went out.”</p>
<p>Level 2: Plans focused on accumulation over time</p>		<p>“We should talk in minutes... It is increasing like 25, 25... along four minutes.”</p> <p>“We should find the areas for the entries and the outputs.”</p>

on the given graph in their unorganized and erroneous Level 0 plans. In addition to this, these participants ignored the starting point of the target graph, even though this was explicitly specified as 100 L in the question. Their common problem was that they spent too much time trying to solve this task even more than 10 minutes for some of the participants. Another interesting finding was that eight participants who ignored the outflow could not solve the task, even at their second trials.

Eight participants (20%) were able to draw their graphs with the correct starting point taking into account the two flows. Their Level 1 plans failed in calculating the change in the amount of water minute by minute. For instance, the plans assigned to Level 1 in Figure 3 seemed to be acceptable, but the participants simply calculated the net flow in relation to every four-minute interval (imitating the first graph given) without taking each individual minute into account.

Level 2 plans (57.5% of the plans) were identified and privileged because of their organization by accumulation over minutes. Units on the given graph were helpful for the participants to base their plans on area computation. Some participants spent some time figuring out the change at each minute before converting their plans based on area computation, while others focused on area from the outset. Examples of the two variations of Level 2 plans were placed in Figure 3. Although the participants with relatively more organized plans (Levels 1 and 2) were expected to adhere to their plans, five participants (12.5%) drew the target graph carelessly. For example, they did not care to match the corresponding points on the axes and as a result the values on the graph became incorrect.

We identified “over 30-minute time” as a potential clue for the department-store task. “Per minute” was identified as the key phrase for this task while observing the skill of monitoring. For instance, an UnSPS had some struggles to figure out the task and asked a metacognitive question after getting feedback:

Excerpt 19: Low-level of monitoring

(During the first trial, after reading the task)

“75 liters (of water) flowed in by the end of 4th minute and 25 liters (of water) flowed out in between the minutes 4 to 8.” (Paraphrasing)

(In the first plan, there was only inflow but outflow was ignored.)

(After getting feedback, the participant read the task all over again.)

“I am confused: Is the inflow of 75 liters for every minute or is it for 4 minutes?”

This UnSPS showed low level of monitoring since they were able to identify the problem in one’s plan. However, they did not take any steps to correct the plan.

Conversely, a SPS identified the units in the graph and rephrased “*per minute*” a few times while explaining one’s plan. The SPS was confident about the plan by stating the following.

Excerpt 20: High-level of monitoring

“I verified inside myself and I think I solved the question correctly.”

Discussion

This study was conducted to clarify how individuals’ metacognitive processes work in solving stock-flow tasks. The study emerged from the statement “we know little about the mental procedures people use in solving these problems and why” by Gonzalez and Wong (2012, p. 4) and contributes to our knowledge by demonstrating how individuals utilize metacognitive strategies to monitor and facilitate problem solving in stock-flow tasks. Within the existing systems-thinking literature, various studies have been conducted to understand why people are unable to solve system tasks by focusing on different variables (Cronin *et al.*, 2009; Gonzalez and Wong, 2012; Lakeh and Ghaffarzadegan, 2015; Qi and Gonzalez, 2015; Fischer and Gonzalez, 2016). In this particular study, implementing the concurrent think-aloud protocol to examine the statements of how participants solved the tasks can be considered an authentic way of exploring such metacognitive activities. The think-aloud protocol sheds light on the difficulties participants experienced while solving stock-flow tasks (e.g. “a and c are too similar,” approximations even though the graphs are rigid.).

The system literature includes several studies on individuals’ failure to solve stock-flow tasks, including those conducted in more prestigious universities (Sweeney and Serman, 2000; Ossimitz, 2002; Cronin and Gonzalez, 2007; Gonzalez and Wong, 2012). Of the 40 participants, 15% were able to successfully solve the department-store task and 50% the bathtub task in this study. Although it was not the main purpose of the study, the results reveal that the percentages of the participants who solved the given tasks correctly are low, which is consistent with the literature. Considering the relative ratios of using metacognitive strategies on the tasks shown in Table 3, it was seen that the participants mostly read the given task and do the calculations correctly, but they performed most poorly in the monitoring and checking. This result suggests that the participants have difficulties in using strategies to control the accuracy and correctness of cognitive operations and results, while they easily use metacognitive strategies such as reading and calculation based on readers’ previous knowledge or background information.

In the orientation phase, while most of the participants read the given task entirely, they were less involved in selecting relevant information and paraphrasing the problem. These two strategies are critical steps in which an individual increases the level of comprehension of the information given in effortful problem-solving tasks. The deficiencies or inconsistencies in the use of these two metacognitive strategies would negatively affect an individual's ability of in-depth thinking by interrupting the skills such as self-questioning (Joseph *et al.*, 2016), finding key words for the solution (Van De Pol *et al.*, 2019; Lippmann *et al.*, 2021), noticing possible errors (Yeung and Summerfield, 2012) and monitoring the solution process more actively (Garofalo and Lester Jr, 1985; Fiedler *et al.*, 2019). Considering that stock-flow tasks are novel and nonroutine problems for most of the participants, this result suggests that the deficiencies while comprehending the problem are likely to lower the ability to solve stock-flow tasks. This finding needs to be examined in future research to unearth that increasing the task's reading comprehension with guided questioning or interventions would improve one's performance in solving stock-flow tasks.

In the study of Fischer and Gonzalez (2016), we see that the success rate increases when the structure of the SF task is changed from local to global manner, shifting focus from system structure to system elements. This is because, in order to think about the SF tasks, the simple building blocks and elements that make up the system must be understood and regulated. The ability to analyze the given task by dividing it into smaller elements is a metacognitive strategy that the individual should utilize (Aşık, 2015). To optimize problem-solving ability, individuals should have metacognitive regulation and apply metacognitive strategies to manipulate cognitive processes (Schoenfeld, 1992). It is important not to reduce the cognitive demand but to guide individuals to think meta-cognitively about complex tasks using relevant information. At this point, it may be worthwhile to examine the effect of providing a set of generic question stems to the participants and asking them to use these question stems to create their own questions (Joseph *et al.*, 2016) while solving stock-flow tasks. In this way, participants can work individually to create their own questions and can reflect more deeply on the stock-flow tasks.

One important finding related to the reading comprehension of the tasks is the critical role of the key words given within the text. It was observed that the participants who read and/or paraphrased all the instructions clearly during the tasks focused on the phrases "over a 30-minute period" for the first task and "per minute" for the second, and they tended to give more correct responses on each task. It should be noted that "over a 30-minute period" implies continuity, and it was expected that people should add or subtract differences of inflow and outflow on each minute. In the second task, "per minute" implied the rate of change in the water stock. Drafting the structure of the problem in one's mind requires a deeper orientation

than simply reading the problem situation (Schoenfeld, 1992; Whimbey *et al.*, 2013). In line with the literature (Thiede *et al.*, 2003; Lippmann *et al.*, 2021), this result of the study revealed that individuals who noticed key words were able to monitor their problem-solving process more accurately. Although phrases such as “over a 30-minute period” might be generally considered insignificant for any test taker, as adjunct parts of a problem to be read quickly or skipped, these statements are important details to bear in mind if one is to devise a plan for solving the problem during the orientation phase. Key words foster problem-solving performance by means of decreasing the ease of processing (Kintsch *et al.*, 1990; Lippmann *et al.*, 2021). Not paying enough attention to key words in the problem-solving process may result in people’s tendency to underestimate accumulations and make their decisions based on flows but not on accumulations over time (i.e. correlation heuristics; Cronin *et al.*, 2009).

The findings related to the task-performance phase revealed that, in line with our expectations, the participants designed an action plan and adhered to that plan at a lower rate in the department-store task which had higher cognitive demands, compared to the bathtub task. In the department-store task, more than 50% of the participants could not put forward a clear action plan and could not perform actions consistent with their plans. High cognitive-demand tasks require individuals to think abstractly, analyze information, and make connections (Stein and Smith, 1998; Van de Walle *et al.*, 2016). In line with the requirements, it was observed that participants in the study struggled to visualize the problem, to select priority information needed for solution, and to switch to a secondary plan when faced with a challenge. As Cronin and their colleagues (2009) emphasize that the use of inappropriate strategies can lead to stock-flow failures; similar overt behaviors such as designing incorrect action plans and adhering and insisting on those plans emerged as significant results of this study. The findings also revealed that those who do not start the problem-solving process with a clear action plan have a low rate of adherence with the plan they have developed. To this end, an important finding is that five out of nine participants (56%) who clearly emphasized the term “accumulation” in their planning process obtained the correct response. The literature highlights that extracted key words after a certain progression while working on a task is more effective in performance and thinking accuracy (Lippmann *et al.*, 2021; Waldeyer and Roelle, 2021). The findings of the study support the literature and calls for further investigations into the phenomena.

For both tasks, the weakest performance was observed in the metacognitive strategies of monitoring and checking the outcome compared to others. It was observed that many participants did not check their answers (more than 80%) and performed poorly in monitoring (more than 50%). The reason why many participants did not perform well in the monitoring activity may be related to their confidence judgments (Schraw, 2009; Lingel

et al., 2019). They may have an absolute confidence in their actual performance and thus overestimate their ability in the task (Schraw, 2009). This finding is intriguing but because we were not able to collect data on the participants' confidence accuracy about their actual performance, further research is needed to clarify the role of metacognitive judgments in solving stock-flow tasks. On the other hand, checking the results is one of the most important metacognitive activities, but most of the participants did not exhibit accurate performance in this issue. However, the answers to the questions in the given two tasks could be checked easily. In the department-store task, checking the answers the participants provided within a minute or more than a minute could easily reveal that the answer was not correct. In the bathtub task, reconstructing the graph backwards could be an easy way to check the correctness of the given answer. At this point, focusing on checking the answer as a metacognitive activity with retrospective explicit questioning soon after the answer is given might be a further research suggestion for stock-flow tasks.

The second research question was related to how metacognitive strategies worked on the participant's decision-making after receiving feedback prompting them to go over their wrong answers. The purpose of giving feedback was to support the monitoring and checking process of individuals' problem-solving activity. It was expected that participants would check and reconstruct their plans after getting the incorrect prompt. This expectation was fulfilled by the problem solvers who were successful at their second attempts. For both of the stock-flow tasks, the individuals who answered correctly after getting the incorrect prompt were able to identify the incorrect part of their previous solution. Therefore, these participants were regarded as having high monitoring skills. For this group, some participants tried to change their present plans after getting the prompt. "I had already given the correct response to this part" [showing the correct part of one's response]; "Should I solve it by calculating the differences that occur each minute as different number of people walk into and leave the store?"; "Should I solve it by paying attention to the area given below the graph?"; "My response might be incorrect due to the fact that I overly focused on the peak and trough points in the graph" were example utterances and self-questions that they formed. Those statements reveal that the successful participants were more aware of their own solution plans.

On the other hand, when unsuccessful attempts were evaluated, two cases came to the foreground. First, many of the UnSPSs gave unsaturated answers without understanding stock-flow relations. They validated their responses after providing it, rather than before. At the end of the evaluation, many of them confirmed their answers without any changes. In other words, these participants ignored the monitoring process and only made a solution check. Second, when these participants were given the incorrect prompt as feedback, it was observed that they went on working on the task without any

strategy changes. Other than that, some participants insisted on the correctness of their results by repeating the solution process irrespective of the prompt. Our expectation about the UnSPSs' second problem-solving trial was that they would repeat the metacognitive processes cyclically, starting with the step of whether they understood the problem correctly. Our expectation was partially supported through our observations. Most of the participants preferred to detect their faults only in more cognitive processes such as mathematical operations or drawing. This situation can be evaluated as an indicator that although the participants were told that their answer was wrong, they did not initiate the process of rethinking/examining what was really asked in the given question. In other words, the participants get stuck in lower-level cognitive processes, but they should foster higher-level cognitive functions to come up with a solution that makes sense. We believe that this is an interesting finding and should be considered for further research.

This study presents some implications for the fields of education and system dynamics. Although the two SF tasks are structurally simple, these non-routine or insight problems (as Cronin and Gonzalez, 2007, call them) offer challenges even for college students. These tasks could be useful for identifying and supporting metacognitive skills when teaching various subject matters. On the other hand, the effect of explicit and systematic metacognitive training on participants' achievement in SF tasks can also be considered as a separate topic to be studied. This study also has implications for the system dynamics field, by demonstrating the potential for future studies aiming at identifying important phrases for the wording of systems tasks and modifying feedback.

Conclusion

Metacognition has been recognized as one of the most relevant predictors of accomplishing complex learning and solving effortful problem tasks (Schoenfeld, 1992; Davidson and Sternberg, 1998; Dignath and Büttner, 2008; Van Der Stel *et al.*, 2010). Stock-flow tasks being cognitively demanding domain-specific questions constitute an important area where metacognitive activities can be investigated. The purpose of this research was to examine the metacognitive activities of the participants in solving stock-flow tasks and to make inferences on the possible reasons for stock-flow failure. As these metacognitive deficiencies are understood, training programs can be designed to help people overcome stock-flow failures.

Our results point out that people have difficulty in using their existing metacognitive skills effectively in domain-specific stock-flow tasks. The metacognition literature reveals that opportunities for engaging in metacognitive tasks are, at times, unavailable to students, or are at least deficient in certain ways (Winne, 1996; Veenman *et al.*, 2000). The results highlight

that the reason for such failure might be participants' inability of using metacognitive skills due to task unfamiliarity or following inappropriate metacognitive steps in particular situations (Veenman *et al.*, 2006), such as stock-flow tasks. Although there are many real-world examples where individuals can reflect on systems thinking, curriculum-based study environments where people can systematically study on and experience the stock-flow tasks are limited. This makes stock-flow tasks much more novel and effortful. Thus, individuals have difficulties in using metacognitive activities effectively (Van Der Stel and Veenman, 2014), and their systems thinking skills remain weak.

Our current findings suggest that, among the metacognitive activities investigated in this study, participants have difficulties in reading comprehension, planning, monitoring, and checking activities. The effectiveness of metacognitive activities employed by the participants is likely to decrease from reading to checking respectively. For the reading-comprehension phase, although participants read the problem entirely, they were less involved in selecting relevant information and paraphrasing the problem. The results show that the participants who made estimations about the solution and recognized the important phrases within the text were more consistent in problem solving by comprehending the text and identifying the given task. However, it should be noted that estimating the answer is one of the domain-specific metacognitive activities pertaining to mathematics (Van Der Stel and Veenman, 2014). For the planning phase, it was observed that participants struggled to select priority information needed for solution and to switch to a secondary plan. They could not come up with a clear plan, and thus difficulties arose in complying with their existing plans. These are also among the reasons why the participants are not able to monitor effectively. For monitoring and checking, those who failed in the given stock-flow tasks had difficulty in determining a proper task strategy and error detection, and they also put little effort into and spend less time checking the correctness of their answers even if they were told that their answer was wrong. In a similar vein, it was noted that those who received feedback on their wrong answers preferred reviewing their own responses to find the error rather than reading the overall problem again and constructing a new strategy. Some of these participants even insisted on their responses even though they received feedback. Investigating the possibility of an individual's confidence accuracy in their actual performance has the potential to hone our understanding of the failure in stock-flow tasks.

An interesting finding about checking the solution correctness was that although they were preservice mathematics teachers, they lacked the metacognitive activity of "checking the answer." One of the most important responsibilities of being a teacher is to facilitate metacognition by modeling their own thinking and prompting reflective thinking in students (Pintrich, 2002; Jacobse and Harskamp, 2012). The fact that even teacher

candidates who are responsible for teaching metacognition do not spare time for checking the solution correctness may be a new research focus that needs to be studied in the future. Secondly, emphasizing the evoking of key words for stock-flow thinking such as “over a 30-minute period,” “per minute,” and “throughout” by underlining them in the text or using guided questions to trigger their metacognitive questioning is a further research suggestion. Lastly, studies that will investigate the accuracy of learners’ perceptions of their own performance, which is defined as metacognitive calibration, and their use of metacognitive activities in solving stock-flow tasks could yield comparative findings and provide further suggestions for understanding stock-flow failure.

Acknowledgements

We would like to thank Dr. Sean Henry for the thoughtful feedback and critical insight.

Biographies

Gürsu Aşık is an assistant professor of Mathematics Education at Bahçeşehir University in Turkey. His academic research interests predominantly focus on students’ mathematical problem-solving skills and metacognition which refers to planning, monitoring and regulation of one’s understanding and performance. He is also interested in raising and maintaining student’s curiosity in learning and teaching.

Zerrin Doğança Küçük is an assistant professor of Science Education at Maynooth University in Ireland. Besides environmental education and understanding complex environmental problems, her research interests are about development of systems thinking skills of students and preservice teachers. She designed system courses for middle school students and preservice teachers.

References

- Abdel-Hamid T, Ankel F, Battle-Fisher M, Gibson B, Gonzalez-Parra G, Jalali M, Kaipainen K *et al.* 2014. Public and health professionals’ misconceptions about the dynamics of body weight gain/loss. *System Dynamics Review* **30**(1–2): 58–74.
- Aşık G. 2015. *A Design Study on Metacognitive Training in Problem Solving [Unpublished Doctoral Dissertation]*. Marmara University: Istanbul, Turkey.

- Assaraf BO, Orion N. 2005. Development of systems thinking skills in the context of earth system education. *Journal of Research in Science Teaching* **42**(5): 518–560.
- Barlas Y. 2002. System dynamics: systemic feedback modeling for policy analysis. In *Knowledge for Sustainable Development - an Insight into the Encyclopedia of Life Support Systems*. UNESCO-Eolss Publishers: Paris, France, Oxford; 1131–1175.
- Costa AL. 1984. Mediating the metacognitive. *Educational Leadership* **42**(3): 57–62.
- Cronin M, Gonzalez C. 2007. Understanding the building blocks of system dynamics. *System Dynamics Review* **23**(1): 1–17.
- Cronin M, Gonzalez C, Sterman J. 2009. Why don't well-educated adults understand accumulation? *Organizational Behavior and Human Decision Processes* **108**: 116–130. <https://doi.org/10.1016/j.obhdp.2008.03.003>.
- Davidson JE, Sternberg RJ. 1998. Smart problem solving: How metacognition helps. In *Metacognition in Educational Theory and Practice*, Hacker DJ, Dunlosky J, Graesser AC (eds). Lawrence Erlbaum Associates: Mahwah, NJ; 47–68.
- Dignath C, Büttner G. 2008. A meta-analysis on intervention studies at primary and secondary school level. *Metacognition and Learning*, **3**(3): 231–264. <https://link.springer.com/article/10.1007/s11409-008-9029-x>
- Doyle JK. 1997. The cognitive psychology of system thinking. *System Dynamic Review* **13**: 253–265.
- Dutt V, Gonzalez C. 2012. Decisions from experience reduce misconceptions about climate change. *Journal of Environmental Psychology* **32**: 19–29.
- Ericsson KA. 2006. Protocol analysis and expert thought: Concurrent verbalizations of thinking during experts' performance on representative tasks. In *The Cambridge Handbook of Expertise and Expert Performance*, Ericsson KA, Charness N, Feltovich P, Hoffman RR (eds). Cambridge University Press: Cambridge; 223–241.
- Fiedler K, Ackerman R, Scarampi C. 2019. Metacognition: monitoring and controlling one's own knowledge, reasoning and decisions. In *The Psychology of Human Thought: An Introduction*, Sternberg RJ, Funke J (eds). Heidelberg University Publishing: Heidelberg; 89–111.
- Fischer H, Gonzalez C. 2016. Making sense of dynamic systems: how our understanding of stocks and flows depends on a global perspective. *Cognitive Science* **40**: 496–512.
- Flavell JH. 1976. Metacognitive aspects of problem solving. In *The Nature of Intelligence*, Resnick LB (ed). Erlbaum: Hillsdale, NJ; 231–235.
- Flavell JH. 1979. Metacognition and cognitive monitoring: a new area of cognitive developmental inquiry. *American Psychologist* **34**: 906–911.
- Fonteyn ME, Kuipers B, Grobe SJ. 1993. A description of think aloud method and protocol analysis. *Qualitative Health Research* **3**(4): 430–441.
- Garofalo J, Lester FK Jr. 1985. Metacognition, cognitive monitoring, and mathematical performance. *Journal for Research in Mathematics Education* **16**(3): 163–176.
- Gero JS, Tang HH. 2001. The differences between retrospective and concurrent protocols in revealing the process-oriented aspects of the design process. *Design Studies* **22**(3): 283–295.
- Gonzalez C, Wong H. 2012. Understanding stocks and flows through analogy. *System Dynamic Review* **28**(1): 3–27.
- Hmelo-Silver CE, Marathe S, Liu L. 2007. Fish swim, rocks sit, and lungs breathe: expert-novice understanding of complex systems. *The Journal of the Learning Sciences* **16**(3): 307–331.

- Jacobse AE, Harskamp EG. 2012. Towards efficient measurement of metacognition in mathematical problem solving. *Metacognition and Learning* 7: 133–149.
- Joseph LM, Alber-Morgan S, Cullen J, Rouse C. 2016. The effects of self-questioning on reading comprehension: a literature review. *Reading & Writing Quarterly* 32(2): 152–173.
- Kintsch W, Welsch D, Schmalhofer F, Zimny S. 1990. Sentence memory: a theoretical analysis. *Journal of Memory and Language* 29: 133–159.
- Lakeh AB, Ghaffarzadegan N. 2015. Does analytical thinking improve understanding of accumulation? *System Dynamic Review* 31(1–2): 46–65.
- Lingel K, Lenhart J, Schneider W. 2019. Metacognition in mathematics: do different metacognitive monitoring measures make a difference? *ZDM* 51(4): 587–600.
- Lippmann M, Danielson RW, Schwartz NH, Körndle H, Narciss S. 2021. Effects of keyword tasks and biasing titles on metacognitive monitoring and recall. *Metacognition and Learning* 16(1): 233–253.
- Mandinach EB, Cline HF. 1994. *Classroom Dynamics: Implementing a Technology-Based Learning Environment*. Lawrence Erlbaum Associates: Hillsdale, NJ.
- Markle DT, West RE, Rich PJ. 2011. Beyond transcription: Technology, change, and refinement of method. *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research* 12(3): 21. Retrieved from: <http://nbn-resolving.de/urn:nbn:de:0114-fqs1103216>.
- Moxnes E, Saysel AK. 2009. Misperceptions of global climate change: information policies. *Climate Change* 93: 15–37.
- Nelson TO. 1996. Consciousness and metacognition. *American Psychologist* 51(2): 102.
- Nelson TO, Narens L. 1990. Metamemory: a theoretical framework and new findings. *The Psychology of Learning and Motivation* 26: 125–141.
- Ossimitz, G. (2002). Stock-flow-thinking and reading stock-flow-related graphs: an empirical investigation in dynamic thinking abilities. *Paper submitted for the 2002 System Dynamics Conference, Palermo*.
- Pintrich PR. 2002. The role of metacognitive knowledge in learning, teaching, and assessing. *Theory Into Practice* 41(4): 219–225.
- Ponterotto JG. 2006. Brief note on the origins, evolution, and meaning of the qualitative research concept “thick description”. *The Qualitative Report* 11(3): 538–549.
- Qi L, Gonzalez C. 2015. Mathematical knowledge is related to understanding stocks and flows: results from two nations. *Systems Dynamic Review* 31: 97–114.
- Schaffernicht MF, Groesser SN. 2016. A competence development framework for learning and teaching system dynamics. *System Dynamics Review* 32(1): 52–81.
- Schoenfeld AH. 1985. *Mathematical Problem Solving*. Academic Press: Orlando.
- Schoenfeld AH. 1992. Learning to think mathematically: problem solving, metacognition, and sense-making in mathematics. In *Handbook of Research on Mathematics Teaching*, Grouws DA (ed). MacMilan Publishing: New York; 224–270.
- Schraw G. 2009. A conceptual analysis of five measures of metacognitive monitoring. *Metacognition and Learning* 4(1): 33–45.
- Schraw G. 2010. Measuring self-regulation in computer-based learning environments. *Educational Psychologist* 45: 258–266.
- Stein MK, Smith MS. 1998. Mathematical tasks as a framework for reflection: from research to practice. *Mathematics Teaching in the Middle School* 3(4): 268–275.

-
- Sterman JD. 2002. All models are wrong: reflections on becoming a systems scientist. *System Dynamics Review* **18**: 501–531.
- Sterman JD, Sweeney LB. 2002. Cloudy skies: assessing public understanding of global warming. *System Dynamics Review* **18**(2): 207–240.
- Sweeney L, Sterman J. 2000. Bathtub dynamics: initial results of a systems thinking inventory. *System Dynamics Review* **16**: 249–286.
- Thiede KW, Anderson MCM, Theriault D. 2003. Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology* **95**(1): 66–73.
- Van de Pol J, De Bruin ABH, Van Loon AH, Van Gog T. 2019. Students' and teachers' monitoring and regulation of students' text comprehension: effects of metacomprehension cue availability. *Contemporary Educational Psychology* **56**: 236–249.
- Van de Walle JA, Karp KS, Bay-Williams JM. 2016. *Elementary and Middle School Mathematics: Teaching Developmentally, EBook, Global Edition*. Pearson Education Limited: UK; <https://ebookcentral.proquest.com/lib/nuim/detail.action?docID=5185860>.
- Van Der Stel M, Veenman MV. 2014. Metacognitive skills and intellectual ability of young adolescents: a longitudinal study from a developmental perspective. *European Journal of Psychology of Education* **29**(1): 117–137.
- Van Der Stel M, Veenman MVJ, Deelen K, Haenen J. 2010. The increasing role of metacognitive skills in math a cross-sectional study from a developmental perspective. *ZDM Mathematics Education* **42**: 219–229.
- Veenman M. 2006. The role of intellectual and metacognitive skills in math problem solving. In *Metacognition in Mathematics Education*, Desoete A, Veenman M (eds). Nova SP: New York; 35–50.
- Veenman MV, Van Cleef D. 2019. Measuring metacognitive skills for mathematics: students' self-reports versus on-line assessment methods. *ZDM* **51**(4): 691–701.
- Veenman MV, Van Hout-Wolters BH, Afflerbach P. 2006. Metacognition and learning: conceptual and methodological considerations. *Metacognition and Learning* **1**(1): 3–14.
- Veenman MVJ, Kerseboom L, Imthorn C. 2000. Test anxiety and metacognitive skillfulness: availability versus production deficiencies. *Anxiety, Stress and Coping* **13**(4): 391–412.
- Waldeyer J, Roelle J. 2021. The keyword effect: a conceptual replication, effects on bias, and an optimization. *Metacognition and Learning* **16**: 37–56.
- Whimbey A, Lochhead J, Narode R. 2013. *Problem Solving & Comprehension*. Routledge: New York.
- Winne PH. 1996. A metacognitive view of individual differences in self-regulated learning. *Learning and Individual Differences* **8**(4): 327–353.
- Yeung N, Summerfield C. 2012. Metacognition in human decision-making: confidence and error monitoring. *Philosophical Transactions of the Royal Society B: Biological Sciences* **367**(1594): 1310–1321.