



**HAL**  
open science

# Students' processing of modelling problems with missing data

Jonas Kanefke, Stanislaw Schukajlow

► **To cite this version:**

Jonas Kanefke, Stanislaw Schukajlow. Students' processing of modelling problems with missing data. Twelfth Congress of the European Society for Research in Mathematics Education (CERME12), Feb 2022, Bozen-Bolzano, Italy. hal-03759032

**HAL Id: hal-03759032**

**<https://hal.science/hal-03759032>**

Submitted on 23 Aug 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Students' processing of modelling problems with missing data

Jonas Kanefke and Stanislaw Schukajlow

University of Münster, Münster, Germany

[jkanefke@uni-muenster.de](mailto:jkanefke@uni-muenster.de), [schukajlow@uni-muenster.de](mailto:schukajlow@uni-muenster.de)

*The problems that are found in the real world, mathematics, and science are usually ill-defined problems. By contrast, the problems given in classrooms tend to be well-defined. It can be useful for students to solve modelling problems so they can learn how to deal with ill-defined problems. In a qualitative study of ninth to eleventh graders ( $N = 12$ ) in different secondary school tracks in Germany, we investigated how students process modelling problems with missing data. We found that students have difficulties noticing when data are missing from some problems. When students notice the missing data, they notice it at the very beginning of their processing, after understanding the problem, or while validating the mathematical results. After noticing, either the students were able to make a realistic assumption about the missing data using various strategies, or they made no assumption at all. The theoretical and practical implications of the study are discussed.*

*Keywords: Mathematical modelling, ill-defined problems, missing data.*

## Introduction


Having students solve problems is an integral part of mathematics teaching all over the world. The problems that are typically given in school have a clear structure, include all relevant data, do not include any superfluous data, and have exactly one solution. This type of problem is known as a well-defined problem. By contrast, the problems encountered in everyday life and at work are often ill-defined and have multiple solutions. Consequently, their solution processes differ from the processes needed to solve well-defined problems. In mathematics education, ill-defined problems are defined as problems situated in a specific context, where one or more aspects of the problem are not well specified, the problem description is not clear, or all the data needed to solve the problem is not provided in the description (Jonassen, 2000). One important type of ill-defined problem is a problem with a connection to reality (e.g., a modelling problem). Modelling problems are characterized by a demanding process of transferring information between the real world and mathematics. Modelling problems can be ill-defined in different ways, for example, when the initial state of the problem is unclear and the initial data are missing, so-called modelling problems with missing data or open (-ended) modelling problems. Modelling problems as ill-defined problems have rarely been the focus of research yet. In this study, we aimed to analyze how students process modelling problems with missing data and how they overcome the difficulties that occur while solving these problems. The theoretical foundations of this research are theories about the processes involved in solving modelling problems (Blum & Leiss, 2007) and problems with missing data (Krawitz et al., 2018).

## Theoretical background and research questions

### Modelling problems

Mathematical modelling is an important competency that is part of mathematical literacy and is included in many national curricula and in mathematics teacher education. Modelling problems have

been found to be difficult for students to solve due to the cognitive complexity of such problems. In order to describe which activities students need to engage in to solve modelling problems, several approaches and numerous theoretical models of the solution process have been developed in previous research (e.g., Blum & Leiss, 2007; Galbraith & Stillman, 2006). An idealized process for finding a solution to a modelling problem proposed by Blum and Leiss (2007) is the following: The first phase involves understanding the problem and constructing an individual situation model. Second, the students have to construct a real model by simplifying and structuring the situation model. Afterwards, the real model is transformed into a mathematical model. The mathematical model allows the student to apply mathematical procedures to compute a mathematical result. Then the mathematical result is interpreted with regard to reality in order to obtain a real result that should be validated with respect to the real situation. The validation can lead to the need to revise the solution and the constructed models by applying the modelling cycle again. There are various types of modelling problems with different characteristics (Maaß, 2010). An important characteristic of modelling problems that is particularly relevant to this study is that they often do not contain all the data needed to find an accurate solution (i.e., the initial state of the problem is unclear). Following Maaß (2010), such modelling problems will be titled modelling problems with missing data. An example of a modelling problem with missing data is the “Fire Brigade” modelling problem (see Figure 1). Missing data for this modelling problem include the position in which the fire engine is parked or the height of the fire engine where the ladder is attached.

<b>Fire Brigade</b>	
<p>There is great happiness at the fire brigade! The authorities have approved a request for a new turntable ladder vehicle. Now, the fire brigade can rescue people from great heights using a basket attached to the end of the ladder. The vehicle has an engine with over 250 hp and has space for 3 firefighters. The length of the turntable ladder is 30 meters. In addition, the firefighters were informed that some safety rules must be observed when using the new vehicle. The vehicle must maintain a safe distance of 7 meters when firefighting a burning house. In addition, care must be taken to ensure that the operation is not obstructed by power lines. Furthermore, in the event of a storm, the vehicle should not be operated as a matter of principle when wind speeds reach 100 km/h.</p> <p>From what maximum height will the fire department be able to rescue people from danger with this new vehicle?</p>	


**Figure 1: The “Fire Brigade” modelling problem, modified from Schukajlow and Krug (2014)**

### **Characteristics of modelling problems with missing data**

In ill-defined problems, the data, goals, and operators are not clearly specified (Jonassen, 2000). Modelling problems with missing data share these important characteristics with ill-defined problems, including the opportunity to develop different solutions to the same problem. From a normative point of view, the ability to solve modelling problems with missing data can be useful for students' actual and future lives because most of the problems that occur in their daily lives and at work are ill-defined. Schukajlow and Krug (2014) identified benefits of prompting students to find multiple solutions to modelling problems with missing data on students' interest, experiences of

competence, and autonomy. In another study, a feeling of autonomy that came from solving modelling problems with missing data was found to be one of the important sources of students' interest in these problems (Schulze Elfringhoff & Schukajlow, 2021). In previous research, many studies have used modelling problems with missing data without emphasizing that data were missing, which is an important feature of the problem. A typical class of modelling problems with missing data are Fermi problems, where a large amount of information is missing, and reasonable assumptions and a series of estimates are necessary for finding a solution (Ärlebäck & Bergsten, 2013), much more than in the problem presented in Figure 1. Therefore, the special type of problem used in this study is characterized as unique because it is missing relatively little data, but the data that are missing are crucial for the solution with respect to reality.

In order to solve modelling problems with missing data, such as the one in Figure 1, three steps seem to be crucial from a theoretical point of view (Krawitz et al., 2018): (a) Students must identify the problem as an ill-defined problem by noticing that the numerical data given in the problem is not sufficient to solve the problem adequately; (b) they must identify which quantities have to be estimated; and (c) after noticing which data are missing, they have to make assumptions about how to deal with the missing data. In order to make assumptions, students need to conceptualize the real-world situation, which requires realistic considerations and extramathematical knowledge. Further, estimation skills and strategies, such as the reference point strategy, are necessary. For the “Tree Track Adventure Park” modelling problem (Figure 2), students must notice during their processing that Mr. Meier needs enough rope so that the rope that will connect the two trees can be securely attached to each tree. Thus, students must estimate how many meters of additional rope will be needed to securely attach the rope to each tree by wrapping the rope around the trees. To make an assumption, students can either just estimate how much rope will be needed, or they can compute the amount of rope that will be needed after estimating how thick the trees are and how many times the rope will be wrapped around them.

<p><b>Tree Track Adventure Park</b></p> <p>Mr. Meier is the owner of a tree track adventure park. He opened this attraction for young and old guests 12 years ago. Now he has to replace the steel ropes to which the safety lines are clipped. These steel ropes are very expensive. To replace them, Mr. Meier needed more of the steel cable than originally planned. He has only 22 m of steel rope left for the last zip line between two trees. The two trees are 17 m apart. So that visitors can then slide from one tree to the other, the steel rope is attached to the starting tree at a height of 10 m and to the ending tree at a height of 8.5 m.</p>	
	
<p>Does Mr. Meier have to buy more steel rope or is the existing steel rope sufficient?</p>	

**Figure 2: The “Tree Track Adventure Park” modelling problem, modified from Schukajlow and Krug (2014)**

If the students neglect the real-world aspect of attaching the rope, they can still calculate how many meters of rope are needed between the trees and give this as an answer, thus concluding that Mr. Meier has enough rope, but this is in fact an unrealistic answer. The rope that is available is not enough because Mr. Meier needs more rope to attach the rope to the trees, so he needs to buy more rope.

No or inappropriate assumptions in a modelling problem can also result in an inadequate situation model and an inadequate mathematical model for the problem situation (Chang et al., 2020). Thus, it is important for students to master the steps needed to solve modelling problems with missing data. Previous studies have pointed out that students have trouble solving modelling problems with missing data because they seem to separate their knowledge of the real world from their mathematical knowledge and tend to ignore the context of the problem (Galbraith & Stillman, 2001). However, in-depth research on students' individual work and thinking with respect to the demands of such problems is rare.

### **Research questions**

In this study, we addressed students' processing of modelling problems with missing data. Thereby, we first analyzed the extent to which students noticed missing data while processing modelling problems with missing data. Second, we investigated whether students made assumptions about the relevant missing data in the modelling problems used in this study. Thus, we posed the following research questions:

1. How do students process modelling problems with missing data?
  - a. To what extent do students notice that data are missing?
  - b. Did students make assumptions about the missing data?

### **Method**

#### **Participants**

Participants were 12 students (seven girls and five boys) from different secondary school tracks in Germany. Participants came from high-, middle-, and low-track classes. They were between the ages of 14 and 16 and were in Grades 9 to 11. All students participated voluntarily in the study with their parents' permission. To select the participants, we followed the principle of maximum variation. Thus, we chose students who varied in gender, age, and mathematical performance. To assess students' mathematical performance, we considered students' math grades, the types of classes they were taking, and the types of schools they were attending. Students' grades ranged from very good to deficient. It was ensured that all participating students had already covered the unit on the Pythagorean theorem, which was required to solve the problems in the study, in previous mathematics lessons.

#### **Procedure**

In order to address the research questions, we conducted a three-step procedure that combined the methodological approaches of thinking aloud, stimulated recall, and interview. The procedure began with instructions for the think-aloud method, where students watched a video with a demonstration of the think-aloud method and practiced the method to solve a nonmathematical task. After the instructions, first, all students worked individually on the same given modelling problems with missing data using the think-aloud method. Students were videotaped as they processed the modelling problems. Second, after solving all the given problems, the stimulated recall was conducted. Video recordings of students working on all the problems administered in the study were shown to the students immediately after they finished working on the problems. Students were asked to pause the

video recordings whenever they wanted to add something that was going through their minds about the situation that was shown on the screen. If the interviewer thought that the situation shown in the video recording was relevant to the research question, and the students did not pause the video recording, the researcher could pause it too. By doing this, the insights from the first step could be intensified. In the third step, the students were administered a semi-structured interview with in-depth questions about their work on the three problems. In particular, questions were asked about how students handled the missing data from the problems.

### Modelling Problems

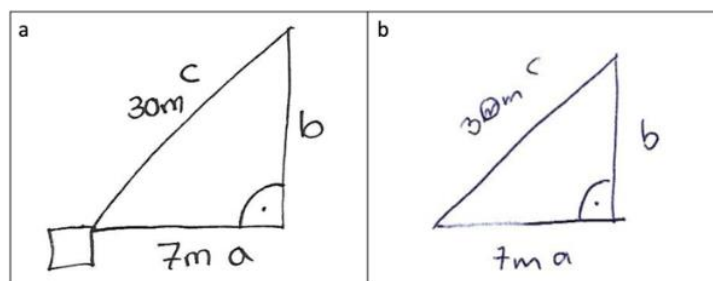
In this study, we used two modelling problems with missing data. Both problems could be solved by applying the Pythagorean theorem, which is part of national and international curricula. The two problems—“Fire Brigade” (Figure 1) and “Tree Track Adventure Park” (Figure 2)—were modified from prior studies (Schukajlow & Krug, 2014). For our study, the relevant data that was missing was always numerical. Thus, students needed to make assumptions about the missing quantities in order to solve the problems correctly. All problems were developed so that students were also able to perform calculations without noticing that important data were missing. For the “Fire Brigade” modelling problem, the height of the fire engine, and for the “Tree Track Adventure Park” problem, the additional rope needed to attach the rope to the trees were considered the relevant pieces of missing data.

### Data Analysis

The transcripts of students’ responses were analyzed using a qualitative content analysis (Mayring, 2014). The deductively developed coding scheme was used to code the sequences with regard to modelling activities from the modelling cycle by Blum and Leiss (2007). Another coding scheme was used to code how the students dealt with the characteristic demands of modelling problems with missing data (Krawitz et al., 2018). More specifically, we coded whether students had noticed the relevant missing data from the modelling problem and had made a realistic numerical assumption.

### Results

We analyzed whether the students noticed that relevant data was missing from the problems (Research Question 1a). In the solutions given by 7 of the 12 students for the "Fire Brigade" problem, the students noticed that the height of the fire engine was missing (Figure 3). Among other objects, Anton’s mathematical drawing included the fire engine with its height (Figure 3, left).



**Figure 3: Excerpts from students’ solutions to the “Fire Brigade” modelling problem**

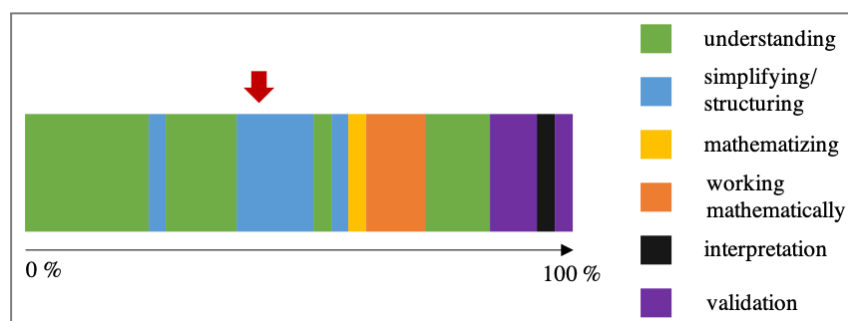
In the sequence presented below from Anton’s processing while thinking aloud, he commented:

Anton: Okay. (...) Here is a house wall, I don't know how big it is yet. (...) The safe distance must be seven meters. Seven. (...) And the height of the fire/the turntable ladder is thirty meters. (...) The turntable ladder is not completely on the ground.

By contrast, in the drawing made by Paula (Figure 3, right), data about the height of the fire engine was not considered. The sequence presented below illustrates that she was not thinking about the missing data while making her drawing.

Paula: The turntable ladder is thirty meters and uhm (...) the safe distance of seven meters. So that's seven meters. And then (...) the ladder would have to be angled so thirty (...) thirty meters. And that is then (...) the height.

For the “Tree Track Adventure Park,” students were less likely to notice the relevant missing data (only 2 out of 12 students noticed). For the “Fire Brigade” problem, we investigated how students’ noticing of the missing data interacted with their modelling process by analyzing when (i.e., between which sequences of modelling activities) the students noticed that the height of the fire engine was missing. Figure 4 illustrates Anton’s modelling process with its modelling activities (Blum & Leiss, 2007), and the arrow indicates when he noticed the missing data. It took Anton four minutes and 17 seconds to process the “Fire Brigade” modelling task.



**Figure 4: Anton’s modelling process**

Anton noticed the height of the engine between sequences in which he was simplifying/structuring. Students usually noticed the missing data in one of two different phases in the modelling process. First, the relevant missing data were noticed at the beginning of the modelling process after an initial understanding and after or during simplifying/structuring, but usually before working mathematically. An example of this case is Anton. Second, some students noticed the missing data after they had completed the modelling cycle activities (i.e., after mathematizing and working mathematically and usually between the validation sequences). These students noticed the missing data after they had already obtained an initial mathematical result. In the validation phase, students noticed that the answer they had calculated was not appropriate for the problem situation. In this case, some students began to correct their previous solution by considering the missing data, whereas other students stuck with their inadequate solution without making any changes. Further, noticing the missing data was not related to the activities of mathematizing or working mathematically. After students noticed that the height of the fire engine was relevant data, they should make a numerical assumption about the missing data and consider this assumption in the next steps of their modelling process. However, only five of the seven students who noticed the missing data also made a numerical assumption. The other two of the seven students did not make any assumptions. An illustration of this can be seen in a sequence of Julius’ solution while thinking aloud.

Julius:           Although, (...) the vehicle height, but I think it doesn't matter.

Julius does not see the height as being particularly important for solving the problem. In order to answer Research Question 1b, we analyzed the extent to which the students made appropriate reality-based numerical assumptions. Comparing the students' numerical assumptions with the real data, we found that all the students' assumptions were within an acceptable range around the real data. Different strategies for making a numerical assumption were found. For example, how Berta made her numerical assumption is described in the following sequence from the stimulated recall.

Interviewer:   How did you do that?

Berta:           I (...) when you look outside, you can see the trees, and then I thought to myself, or in the climbing forest, there are sometimes trees like that, and then I thought to myself that (...) yes, if you just put an arm around it, I don't know, then that's (...) usually a meter or two, if you, if those are thicker trees.

## **Summary and Discussion**

We investigated students' processing of modelling problems with missing data (i.e., real-world problems where the initial state of the problem is ill-defined). In this study, we analyzed whether students noticed relevant missing data and whether they made realistic assumptions about the missing data. An important finding of the study was that students seem to have trouble noticing the relevant missing data in the modelling problems. One possible explanation for this finding could be that students fail to think about the real-world situation of the modelling problem and just take the quantities given in the problem description and then work mathematically with these quantities. This idea would contribute to previous research on mathematical modelling by pointing out that students do not appropriately consider the real-world context of a problem when they process problems with missing data. The modelling problems used in this study were designed to analyze how students deal with missing data and to make it possible for students to perform calculations without noticing the missing data. However, ignoring the missing data in these tasks leads to unrealistic answers. As the necessity of noticing missing data can easily be overlooked, an intensive consideration of the real-world context was essential for solving these problems. As the number of students who noticed the missing data varied considerably between the modelling problems, this finding suggests that whether students notice missing data depends on the specific characteristics of the task. These characteristics might be the context of the problem or the use and nature of the missing data. With respect to the modelling cycle (Blum & Leiss, 2007), we found that some students noticed the missing data directly after reading the problem description, indicating that these students try to develop a more in-depth understanding of the given situation before continuing to process by mathematizing and working mathematically. In the other cases in which students noticed the missing data, the in-depth understanding of the real-world context took part at the end of the processing, after mathematization and working mathematically. Another important finding of our study is that students who noticed the missing data did not always make a numerical assumption, but when students did make an assumption, the assumptions were realistic. This indicates that measurement skills or estimation strategies are not the key obstacle to making assumptions. One possible explanation for the results regarding difficulties in including assumptions in processing might be that problems with missing data and the ability to make assumptions might not be a part of students' mathematics classes. Another possible explanation may be students' conceptions and beliefs about mathematics.



A practical contribution of our study is that students need practice processing modelling problems with missing data in school. Therefore, teaching methods should address students' ability to notice missing data and make assumptions. As the results of our study show, close attention should be paid to the phases of simplifying, structuring, and validation that were particularly relevant for the noticing of missing data. Teaching methods need to stimulate these phases intensively.

## References

- Ärlebäck, J. B., & Bergsten, C. (2013). On the Use of Realistic Fermi Problems in Introducing Mathematical Modelling in Upper Secondary Mathematics. In R. Lesh, P. L. Galbraith, C. R. Haines, & A. Hurford (Eds.), *Modeling Students' Mathematical Modeling Competencies: ICTMA 13* (pp. 597–609). Springer Netherlands. [https://doi.org/10.1007/978-94-007-6271-8\\_52](https://doi.org/10.1007/978-94-007-6271-8_52)
- Blum, W., & Leiss, D. (2007). How do students and teachers deal with mathematical modelling problems? The example sugarloaf and the DISUM project. In C. Haines, P. L. Galbraith, W. Blum, & S. Khan (Eds.), *Mathematical modelling: Education, engineering and economics* (pp. 222–231). Horwood.
- Chang, Y.-P., Krawitz, J., Schukajlow, S., & Yang, K.-L. (2020). Comparing German and Taiwanese secondary school students' knowledge in solving mathematical modelling tasks requiring their assumptions. *ZDM*, 52(1), 59–72. <https://doi.org/10.1007/s11858-019-01090-4>
- Galbraith, P., & Stillman, G. (2001). Assumptions and context: Pursuing their role in modelling activity. In J. F. Matos, W. Blum, K. Houston, & S. P. Carreira (Eds.), *Modelling and mathematics education* (pp. 300–310). Horwood Publishing.
- Galbraith, P., & Stillman, G. (2006). A framework for identifying student blockages during transitions in the modelling process. *ZDM*, 38(2), 143–162. <https://doi.org/10.1007/BF02655886>
- Jonassen, D. H. (2000). Toward a design theory of problem solving. *Educational Technology Research and Development*, 48(4), 63–85. <https://doi.org/10.1007/BF02300500>
- Krawitz, J., Schukajlow, S., & Van Dooren, W. (2018). Unrealistic responses to realistic problems with missing information: What are important barriers? *Educational Psychology*, 38(10), 1221–1238. <https://doi.org/10.1080/01443410.2018.1502413>
- Maaß, K. (2010). Classification scheme for modelling tasks. *Journal Für Mathematik-Didaktik*, 31(2), 285–311. <https://doi.org/10.1007/s13138-010-0010-2>
- Mayring, P. (2014). *Qualitative content analysis: Theoretical foundation, basic procedures and software solution*.
- Schukajlow, S., & Krug, A. (2014). Do multiple solutions matter? Prompting multiple solutions, interests, competence, and autonomy. *Journal for Research in Mathematics Education*, 45(4), 497–533. <https://doi.org/10.5951/jresmetheduc.45.4.0497>
- Schulze Elfringhoff, M., & Schukajlow, S. (2021). O que torna um problema de modelação interessante? Fontes de interesse situacional em problemas de modelação. *Quadrante*, 30(1), 8–30. <https://doi.org/10.48489/quadrante.23861>
- Verschaffel, L., Greer, B., & de Corte, E. (2000). *Making sense of word problems*. Swets and Zeitlinger.