Research in Science Education (2021) 51:1459–1479
https://doi.org/10.1007/s11165-020-09929-7

Pre-service Biology Teachers' Responses to First-Hand Anomalous Data During Modelling Processes

Sabine Meister[1] · Moritz Krell[2] · Maximilian Göhner[2] · Annette Upmeier zu Belzen[1]

Published online: 24 March 2020

Abstract

In this research project we investigate the role of responses to anomalous data during modelling processes. Modelling is seen as a comprehensive practice that encompasses various aspects of scientific thinking; hence, it is an important style of scientific thinking, especially if analysed from a process-based perspective. Therefore, it provides the opportunity to understand the role of anomalous data on scientific thinking from a broader perspective. We analysed how pre-service biology teachers ($N = 11$) reacted to self-generated anomalous data during modelling processes induced by investigating a water black box. The videotaped and transcribed modelling processes were analysed using qualitative content analysis. If anomalous data were recognised, a majority of explanations were based on methodical issues. This finding supports results from previous studies investigating responses to first-hand anomalous data. Furthermore, we found four response patterns to anomalous data during modelling processes: *no recognition*, *no explanation*, *methodical explanation*, and *model-related explanation*. Besides, our study indicates by trend a systematic relation between response patterns to anomalous data and modelling strategies. Consequently, the improvement of responses to anomalous data could be a promising way to foster modelling competencies. We are convinced that an integrated approach to anomalous data and modelling could lead to deeper insights into the role of data in scientific thinking processes.

Keywords Anomalous data · Modelling · Pre-service science teachers · Scientific thinking

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11165-020-09929-7) contains supplementary material, which is available to authorized users.

✉ Sabine Meister
sabine.meister@hu-berlin.de

Extended author information available on the last page of the article

Introduction

One goal of science education is that students should become scientific thinkers in order to participate in today's science-based society. Scientific thinking[1] is an umbrella term including practices like asking questions, formulating hypotheses, conducting investigations, and evaluating data (Fischer et al. 2014; Hetmanek et al. 2018; Rönnebeck et al. 2016). These practices of scientific thinking are included in national standards for science education of several countries (e.g. Australia: VCAA 2016; Germany: KMK 2005; USA: NGSS Lead States 2013). However, scientific thinking cannot be described as one specific construct. Moreover, practices of scientific thinking can be used differently depending on the style of scientific thinking that is in focus (Kind and Osborne 2017). Following the theoretical framework of Lehrer and Schauble (2015), scientific thinking is described as a modelling process that encompasses various other practices. Data evaluation plays an essential role in the modelling process, since data represent pieces of information used for building and developing models (Jeong et al. 2007; Samarapungavan 2018). Data that are not in line with prior expectations, so-called anomalous data (Chinn and Brewer 2001), have the potential to induce the change of an initial model about a phenomenon in focus (Dounas-Frazer et al. 2018; Khan 2008; Author 2017). Hence, the evaluation of anomalous data is a core activity in modelling (Giere et al. 2006). However, most research in science education on responses to anomalous data is conducted with a focus on experimentation as scientific processes (Crujeiras-Pérez and Jiménez-Aleixandre 2019; Lin 2007) as well as conceptual development (Chinn and Brewer 1998; Hemmerich et al. 2016). Research on the role of anomalous data during modelling processes on the other hand is limited. Furthermore, studies on modelling competencies show that pre-service science teachers have difficulties conducting scientific modelling processes (Krell et al. 2017). The present article combines research on anomalous data and modelling in order to investigate the role of responses to anomalous data during modelling processes. Therefore, the aim is to get deeper insights into these scientific processes and, if and how they might relate to each other. In consequence, the combination of these two research perspectives offers new implications for science and science teacher education to support both, handling anomalous data and modelling.

Theoretical Background

The analysis and evaluation of empirical data is a central part of scientific thinking, since it provides the information needed for gaining knowledge about a phenomenon (Jeong et al. 2007; Samarapungavan 2018). Therefore, data builds the empirical foundation for most scientific thinking styles, like experimental exploration or modelling (Kind and Osborne 2017). However, how the data is used might differ between the styles. In the case of modelling, data induce the development of models in science, especially if the data contradict the initial assumptions within these models (Chinn and Brewer 2001; Dounas-Frazer et al. 2018; Khan 2008; Krell et al. 2017). Since handling data plays such an important role in science, a lot of research had been done to investigate skills and knowledge that are needed for data evaluation

[1] Equivalent to Lehrer and Schauble (2015), we use the term *scientific thinking* in this contribution instead of *scientific reasoning* that is used in other contributions (Krell et al. 2018) for a similar construct. Following Lehrer and Schauble (2015), reasoning is the argumentative part of scientific thinking and, therefore, scientific thinking of higher-order.

in the psychological sciences (e.g. Chinn and Brewer 2001; Jeong et al. 2007; Samarapungavan 2018) and in science education (e.g. Duncan et al. 2018; Gott et al. 2008). In addition to investigations on the general practice of handling data, some studies focus specifically on so-called anomalous data. Anomalous data are defined as evidence that is not in line with, and would not be predicted by, a person's mental model (Chinn and Brewer 1998). Therefore, anomalous data can be a driving force for evaluating and developing existing theoretical models. However, previous studies on responses to anomalous data show that students and undergraduates tend to reject or reinterpret the data in order to hold on to their initial models (Chinn and Brewer 1998; Lin 2007).

Responses to Anomalous Data

The research on responses to anomalous data done by Chinn and Brewer (1998) provide an empirically tested taxonomy defining eight different types of responses: *ignoring*, *rejection*, *exclusion*, *uncertainty about validity*, *abeyance*, *reinterpretation*, *peripheral theory change*, and *theory change*. The grouping of reactions to the eight response types is based on how the reaction fits the criteria: perceiving the data as valid, explaining the anomalous data, and changing the initial theory. The majority of undergraduates' reactions showed in a paper-pencil-instrument was classified into the response *rejection* that is characterised by questioning the validity of the data and explaining the anomaly with methodical issues (Chinn and Brewer 1998). A consecutive study supported the taxonomy by investigating eight graders' responses show anomalous data (Mason 2001).

However, these studies provided anomalous data in written text passages as part of paper-pencil assessments—so-called second-hand data. Evaluating externally presented second-hand data instead of self-generated first-hand data affects the way the data are perceived and evaluated (Hug and McNeill 2008). Second-hand data are often seen as authoritative by students and provide an opportunity to evaluate complex data patterns; hence, second-hand data promote content-related reasoning. In contrast, first-hand data are more likely to be evaluated critically by students promoting the discussion about limitations and uncertainties in data (Hug and McNeill 2008). Furthermore, in the previous studies that focus on how anomalous data induces conceptual development (Chinn and Brewer 1998; Mason 2001), participants were primed with a theory that was then challenged by the data. They, therefore, did not assess participants' own mental models. Other studies on anomalous data focus on first-hand anomalous data generated by the participants in laboratory settings (Crujeiras-Pérez and Jiménez-Aleixandre 2019; Lin 2007; Pickering and Monts 1982; Toplis 2007). In contrast, anomalous data in laboratory settings do not exclusively address conceptual development; they also induce critical thinking about the experimentation process and data in general (Lin 2007; Pickering and Monts 1982). In a study with undergraduates, responses to anomalous data during an experimentation course with chemical contexts were investigated (Lin 2007). The findings supported the application of the taxonomy of responses to anomalous data proposed by Chinn and Brewer (1998) for laboratory settings under the condition of adding *uncertainty of interpretation* as an additional category (Lin 2007). However, in line with the previous findings, most responses were classified into the category *rejection*, indicating that, independent of whether the data are externally given (second-hand) or self-generated (first-hand), the validity of anomalous data is often in question.

Findings of a recent study (Crujeiras-Pérez and Jiménez-Aleixandre 2019) show another phenomenon regarding first-hand anomalous data, with many students struggling to recognise

self-generated anomalous data at all. This finding corresponds to other studies describing similar difficulties of students to recognise anomalous data during experimentation (Chinn and Malhotra 2002; Toplis 2007).

However, so far, research on anomalous data focused on either how it can induce conceptual development or how anomalous data are handled during experimentation processes that represent just one of various scientific thinking styles. Furthermore, responses are mostly investigated post hoc from a product-based perspective by analysing learners' written explanations (Chinn and Brewer 1998) or laboratory reports (Lin 2007). Therefore, a process-based analysis of responses to anomalous data during another style of scientific thinking like modelling (Kind and Osborne 2017) has the potential to provide more holistic insights into this field of research, especially if modelling is seen as a broader theoretical framework that encompasses different styles of scientific thinking (Lehrer and Schauble 2015).

Modelling in Science and Science Education

Modelling is a main practice used by scientists to make sense of the world (Giere et al. 2006) and is, therefore, essential for scientific thinking (Lehrer and Schauble 2015; Passmore et al. 2017). From a theoretical point of view, models are 'epistemic tools' for sense-making (Knuuttila 2011). Meaning that models are developed and applied by scientists to make sense of phenomena; models in science are hypothetical entities and modelling is the process in that those entities are developed and evaluated, contrasted with evidence, to explain and predict natural phenomena. The strategy of scientific modelling can be generally described as follows: 'The modeler's strategy is to gain understanding of a complex real-world system via an understanding of simpler, hypothetical system that resembles it in relevant respects' (Godfrey-Smith 2006, p. 726). Hence, modelling incorporates investigations into the experiential world, the development of explanations and predictions, and the evaluation of the latter by means of collecting further evidence. From this perspective, 'model-based and evidence-based reasoning are intertwined […]; evidence is used to construct models, and models are used to inform the search for evidence' (Dounas-Frazer et al. 2018, p. 6). Modelling can be conceptualised as a non-linear, cyclical process, which includes activities of exploration, model development, and model evaluation (Göhner and Krell 2018; Krell et al. 2017; Clement 2009; Giere et al. 2006).

Several models of the modelling process have been proposed in the educational literature, which can be used as heuristic tools to structure, analyse, and support the creative and iterative process of modelling in science (e.g., Clement 2009; Giere et al. 2006; Gilbert and Justi 2016; Khan 2011; Krell et al. 2017). For the present study, a theoretical framework of the modelling process was used, which has been developed in previous studies (Göhner and Krell 2018; Krell et al. 2017) and integrates frameworks of analogical reasoning (Clement 2009) and scientific reasoning through models (Giere et al. 2006). In this theoretical framework of the modelling process (Fig. 1), the development of a model is motivated by the perception of a phenomenon. Subsequently, the phenomenon is modelled through activation of analogies and experiences that are then used for model development. The evaluation of the developed model includes comparing to what extent it is in line with the observations. The application of the model includes the deduction and evaluation of predictions, which are based on the model, using empirical evidence. This leads to the evaluation of assumptions and to further insights about the underlying phenomenon. Hence, the modelling process is led by the aim of sense-making using models as epistemic tools (Gilbert and Justi 2016; Passmore et al. 2014). This theoretical

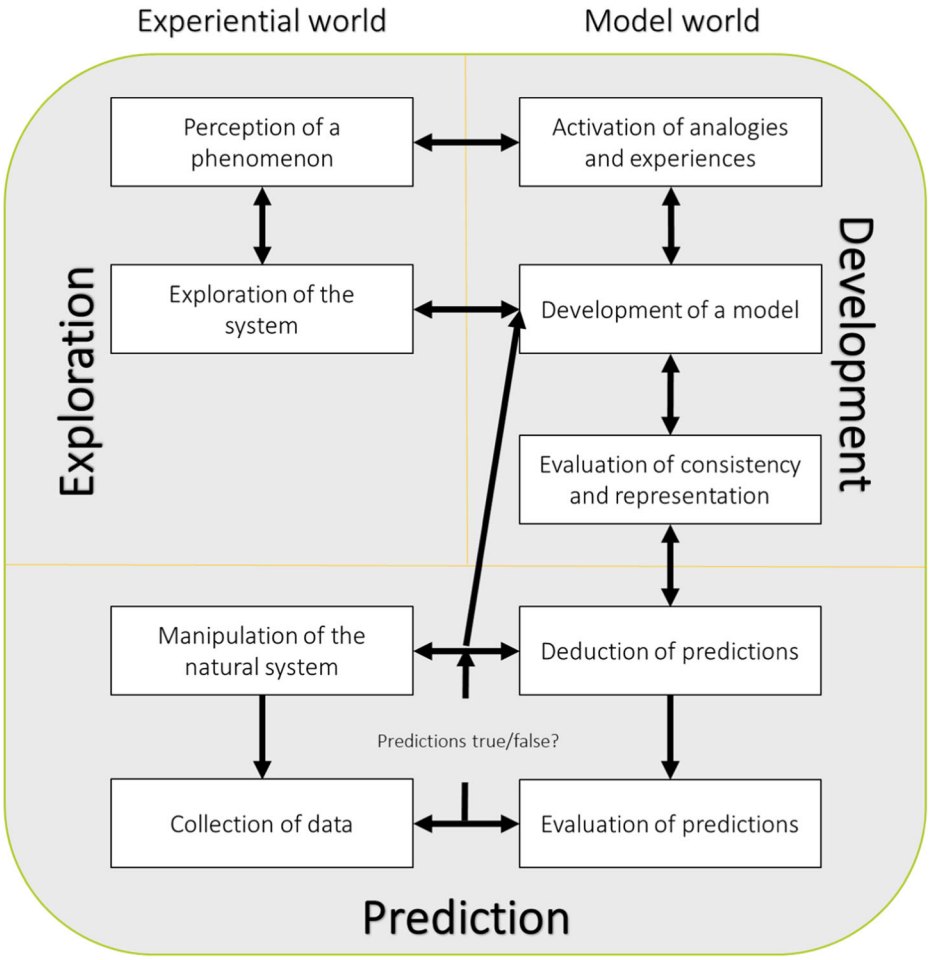

Fig. 1 Idealised illustration of the modelling process (Krell et al. 2017; cf. Clement 2009; Giere et al. 2006)

description of the modelling process is—of course—highly idealised, but was helpful as an heuristic framework for data analysis in the present study.

An appreciation of modelling and the role of models in science as epistemic tools (meta-modelling knowledge) is essential for scientific thinking (Giere et al. 2006; Knuuttila 2011; Schwarz and White 2005). Therefore, this understanding should be fostered in science classes, in order to enable students to understand the nature of science and its procedures, its capabilities, and its limitations (Passmore et al. 2017; Schwarz and White 2005). In science education standard documents and policy papers in various countries (e.g. KMK 2005; NGSS Lead States 2013), it is demanded that students at school should develop modelling skills and an understanding of the role of models in science. Consequently, such modelling competencies also have to be part of science teachers' professional competencies (Günther et al. 2019; Justi and van Driel 2005). However, most studies suggest that teachers' meta-modelling knowledge is rather limited (Justi and Gilbert 2003; Krell and Krüger 2016). Furthermore, models are mostly used as explanatory representations and not as epistemic tools in science classes

(Campbell et al. 2015; Khan 2011). Hence, understanding and using models as epistemic tools seem to be challenging for science teachers.

Strategies of Modelling in Science Education

One way to foster the use of models as epistemic tools in science education is to engage learners in performing modelling processes. However, different strategies of modelling can be theoretically distinguished, each addressing different epistemic aims (Svoboda and Passmore 2013). In science education, five modelling pedagogies have been proposed (Oh and Oh 2011): exploratory modelling (investigating a pre-existing model), expressive modelling (developing a model for expressing ideas about a phenomenon), experimental modelling (deducing predictions from a model and test them empirically), evaluative modelling (comparing and evaluating alternative models of/ for the same original), and cyclic modelling (being engaged in the cyclic process of model development, evaluation, and modification). Interestingly, expressive and exploratory modelling are the most frequently used pedagogies in science education (Campbell et al. 2015) and being the ones with a higher focus on using models as explanatory representations, rather than using models as epistemic tools. This finding on how modelling is used in science education can be the result of the previously mentioned lack in teachers' meta-modelling knowledge. However, empirical and process-based studies on modelling strategies of science teachers for scientific problem solving are rather limited (Nicolaou and Constantinou 2014). Recently, author and colleagues (Göhner and Krell 2018; Krell and Hergert 2020; Krell et al. 2017) analysed pre-service biology teachers' modelling activities during investigating a black box problem (see supplementary online material, source 1, for the single modelling activities). Based on the occurrence and transitions between these activities, the authors identified five different modelling strategies:

> *(1) Exploration only*: Pre-service teachers do not develop a model at all; that is, they only explore the underlying system (although the task explicitly demanded to develop a model).
> *(2a) Focus on exploration*: Pre-service teachers engage in activities related to exploration and model development, but specifically focus on exploration.
> *(2b) Exploration and model development*: Pre-service teachers engage in activities related to exploration and model development in a rather balanced way, that is, without a specific focus on exploration.
> *(3a) Single prediction*: Pre-service teachers engage in activities related to exploration and model development and use their model to predict the system's behaviour once.
> *(3b) Repeated prediction*: Pre-service teachers engage in activities related to exploration and model development and use their model to predict the system's behaviour repeatedly and cyclical.

Based on a current sample of $N = 31$, it was found that most pre-service biology teachers ($n = 11$) did *focus on exploration* when modelling the black box, followed by *single prediction* ($n = 9$), *repeated prediction* ($n = 5$), *exploration and model development* ($n = 5$), and *exploration only* ($n = 1$; Göhner and Krell 2019). That is, 14 of 31 respondents used their models to evaluate assumptions about the black box. Hence, this process-based study suggests that pre-service teachers are able to develop and use models as epistemic tools for sense-making in a problem-based task like investigating a black box.

Aims and Research Questions

In the present study, we acknowledge the importance of data evaluation, especially the evaluation of anomalous data, in the process of modelling as one of different styles of scientific thinking. We aim for more holistic insights into the process of evaluating first-hand anomalous data by expanding the previously done research that focusses on experimentation as a scientific thinking style (Crujeiras-Pérez and Jiménez-Aleixandre 2019; Lin 2007). Furthermore, we look on responses to anomalous data during the modelling process, in contrast to analysing only the product of the inquiry process. Hence, in order to investigate the role of anomalous data during these modelling processes, we combine research on responses to anomalous data and on modelling processes. The key objective of this study is to identify the role of anomalous data during modelling processes. More specifically, the following research questions are discussed:

1. Which responses to anomalous data during modelling processes do pre-service science teachers show?
2. To what extent can pre-service science teachers' responses to anomalous data be grouped into response patterns?
3. Which relation can be found between pre-service science teachers' response patterns to anomalous data and their modelling strategies?

Method

Sampling

Empirically, we draw on the study on pre-service teachers' modelling strategies (Göhner and Krell 2018; Krell and Hergert 2020; Krell et al. 2017) and selected 11 pre-service biology teachers from this study's sample (currently $N = 31$). The selection was based on two sampling criteria: course of studies (undergraduate and graduate pre-service biology teachers) and score in a scientific reasoning competencies multiple-choice test (low and high; Table 1). Data about scientific reasoning competencies were gathered from an ongoing longitudinal study at the university (Hartmann et al. 2015). Low and high scientific reasoning competencies were defined as $x < M - 1SD$ or $x > M + 1SD$, respectively, based on the sample of the longitudinal study. Pre-service biology teachers with each combination of the sampling criteria (i.e. four sampling groups) were selected in order to increase the probability of variation in their reactions to anomalous data during the modelling processes. We aimed for three participants

Table 1 Study sampling ($N = 11$) based on the used sampling criteria. Acronyms in parentheses

		Scientific reasoning competencies	
		Low	High
Course of studies	Undergraduates	$n = 3$ (Angelina, Celine, Jonathan)	$n = 3$ (Jenny, Lauren, Iris)
	Graduates	$n = 2$ (Jim, Frida)	$n = 3$ (Cynthia, Floyd, Carlo)

for each of the sampling cells. Since only two graduate students showed low scientific reasoning competencies, we got a total sample of 11 participants for the present study.

Procedure

In order to engage pre-service biology teachers in a scientific modelling process, we used a water black box (cf. Lederman and Abd-El-Khalick 2002). This black box has a funnel on its top to fill in water and a pipe at the bottom where the water flows out and contains an arrangement of inner siphons that causes a specific output pattern depending on the amount of water that is filled in (Krell et al. 2017). The black box approach is established in science education research to study the processes of scientific thinking (e.g. aspects of nature of science and argumentation) and modelling (Krell and Hergert 2020; Lederman and Abd-El-Khalick 2002; Passmore and Svoboda 2012). In such approaches, the black box represents a natural phenomenon under study, and the exploration of the black box represents the process of scientific discovery. Hence, it was shown that black box activities are suitable to elicit modelling processes, with models used as epistemic tools to discover the black box (Author 2017; Passmore and Svoboda 2012). The black box approach was intentionally chosen instead of an authentic and content-rich scientific problem, in order to reduce the influence of prior knowledge on the pre-service biology teachers' modelling processes. However, it is recognised that this might be one limitation of the study since some authors emphasise the important role of content knowledge for scientific modelling (Ruppert et al. 2017).

The pre-service biology teachers got the task to graphically develop (i.e. to draw) a model of the inner system of the black box and to think aloud during their modelling process (Ericsson and Simon 1998). The individual performances are video-taped and fully transcribed. No time constraint was given. The drawings of the participants are defined as the models in this study; an assessment of mental models was not possible with the applied methodical setting of videography (Werner et al. 2019).

Because of the specific output pattern and the dependency on the amount of water that is filled in, pre-service biology teachers have to repeatedly generate data to recognise the output pattern that gives clues about the inner system of the box. Therefore, participants are likely to be confronted with self-generated anomalous data several times during the modelling process (illustration and more detailed description published in Krell et al. 2017).

Data Analyses

In order to answer the first research question, responses to anomalous data during the modelling processes of pre-service biology teachers were analysed within the methodical framework of qualitative content analysis (Schreier 2012). For this purpose, a category system has been developed deductively based on studies about responses to anomalous data obtained in laboratory settings (Crujeiras-Pérez and Jiménez-Aleixandre 2019; Lin 2007) and about pre-service biology teachers' modelling activities (Krell et al. 2017). Data of two pre-service biology teachers were used to inductively refine the category system, in order to ensure its empirical validity (e.g. by adding new categories or by differentiating existing categories; Schreier 2012). In order to evaluate the appropriateness of the refined category system, two science education researchers independently used the category system to analyse reactions to anomalous data of one further case. Cohen's kappa ($\kappa = .92$) indicated a very good interrater-agreement and, thus, the

Table 2 Category system, with codes, subcodes, and explanations

Code	Subcode	Explanation
Type 1: inconsistency between observed data and general logic		
Recognition	–	AD (type 1) is recognised
Type 2: inconsistency between observed data and expected data pattern/single datum		
Recognition	–	AD (type 2) is recognised
Explanation	Methodical	Recognised AD is explained with limited own skills, inaccuracies in measurement, or the nature of the black box
	With model	Recognised AD is explained by developing an explanatory model
Change	–	Expectation (concerning data pattern/single datum) is changed based on AD
Type 3: inconsistency between observed data and developed model		
Recognition	–	AD (type 3) is recognised
Explanation	Methodical	Recognised AD is explained with limited own skills, inaccuracies in measurement, or the nature of the black box
	With model	Recognised AD is explained by developing an explanatory model
Change	Modification	Developed model is changed to explain AD
	Rejection	Developed model is rejected

Note: The code descriptions have been shortened. *AD*, anomalous data

category system's appropriateness to objectively categorise pre-service biology teachers' responses to anomalous data during modelling processes. The final category system (Table 2) differentiates between three types of anomalies during the modelling process. For each type, a response to the recognised anomaly was coded, exclusively if the participant previously mentioned an expectation or hypothesis regarding this type of anomaly.

For type 1 only the recognition of the inconsistency is included into the category system, since it represents a general confusion that is not explainable with logical reasoning. Type 2 and type 3 contain the categories explanation and change, because these types of inconsistencies are explainable and the pre-service biology teachers can change their expectations as well as their models.

For data analysis, a three-step strategy of subsumption was used to identify and categorise situations in the transcripts that show responses to anomalous data in modelling processes (Schreier 2012):

1. Situations in the modelling processes were identified in which the subjects recognise the occurrence of anomalous data.
2. Identified situations were categorised using the category system (Table 2) within one of three types of anomaly indicating an inconsistency between observed data and general logic (type 1), expected data pattern/single datum (type 2), or developed model (type 3).
3. Categorised situations are coded by answering two central questions (i) *Are the anomalous data explained (referring to methodical issues or with a model)?* and (ii) *Are the expected data patterns/developed models changed?*

The transcripts were coded by using the MaxQDA software (VERBI Software 2016). After the coding process, disagreements between the two coders were identified and resolved by discussion in order to reach a final consensus for further data analysis.

Table 3 Theoretically expected response patterns to anomalous data during modelling processes

Response pattern	Recognition	Explanation	Reaction
1	No	No	No
2	Yes	No	No
3	Yes	Explanation (method)	No
4	Yes	Explanation (model)	Modelling

In order to compare pre-service biology teachers' responses to anomalous data and group them into response patterns (research question 2), self-generated visualisations (*codelines*) from the data were used (generated with R package ggplot2; Wickham 2016). The grouping of the response patterns was based on literature (Chinn and Brewer 1998; Crujeiras-Pérez and Jiménez-Aleixandre 2019). Table 3 shows the four response patterns that were theoretically expected. The first response pattern is in line with the findings that the recognition of anomaly in self-generated data is difficult (Crujeiras-Pérez and Jiménez-Aleixandre 2019). Without recognising the anomalous data, no explanation and no further reaction by the participant are needed. The second response pattern describes the case that anomalous data are recognised but no explanation is offered for them (*exclusion* in the taxonomy of Chinn and Brewer 1998). The third response pattern bases on findings that recognised anomalous data are mostly explained with methodical issues (*rejection* in the taxonomy of Chinn and Brewer 1998), especially in the case of first-hand data (Crujeiras-Pérez and Jiménez-Aleixandre 2019). Because of this type of explanation, the validity of the data is in question and, therefore, no further reaction regarding participants' expectations and models should occur. The fourth response pattern represents a response to anomalous data that leads to further steps in the modelling process. It is equivalent to *peripheral theory change* and *theory change* in the taxonomy of responses to anomalous data (Chinn and Brewer 1998).

For answering the third research question, we compared the response patterns to anomalous data shown by the pre-service biology teachers with their modelling strategies. The modelling strategies (*exploration only*, *focus on exploration*, *exploration and model development*, *single predicting*, *repeated predicting*) have been identified in the ongoing study from where we selected the sample for the present study (Göhner and Krell 2018). For the identification, an activity-based category system (see supplementary online material, source 1) was used and the transcripts of the modelling processes were independently coded by two researchers resulting in acceptable to very high interrater-agreement $.5 \leq \kappa \leq .9$). Based on this, the codings of modelling activities and responses to anomalous data were superimposed.

Results

Pre-service Biology Teachers Modelling Strategies

For relating pre-service biology teachers' responses to anomalous data to their modelling strategies, the results of the previous mentioned study (Göhner and Krell 2018) are used, which focusses on the identification of modelling strategies by analysing pre-service biology

Table 4 Modelling strategies used by the pre-service teachers in the current sample

Modelling strategy	$n_{participant}$	Description
(3b) Repeated prediction	2	Frida and Jenny explored the black box and developed a model of its inner system on the board. During the process, they further tested their model several times by predicting the black box' behaviour and testing these predictions. Hence, they used their models as epistemic tools to evaluate assumptions about the black box.
(3a) Single prediction	2	Iris and Cynthia explored the black box and developed a model of its inner system on the board. One time during the process, they further tested their model by predicting the black box' behaviour and testing this prediction. Hence, they used their models as epistemic tools to evaluate assumptions about the black box.
(2b) Exploration and model development	2	Jonathan and Floyd explored the black box and developed a model of its inner system on the board. Thereby, they showed activities related to exploration and model development in a rather balanced way that is without a clear focus. Hence, they used their models as representations that is to express their ideas about the black box.
(2a) Focus on exploration	4	Celine, Jim, Lauren, and Angelina explored the black box and developed a model of its inner system on the board. Thereby, most activities were related to the exploration of the black box. Hence, they used their models as representations that is to express their ideas about the black box.
(1) Exploration only	1	Carlo did not develop a model at all; that is, he did not externalise his ideas about the inner system of the black box; he only explored the black box and summarised his observations repeatedly.

teachers' modelling processes during a black box activity. Table 4 gives an overview of the five strategies identified for the present sample and exemplifies how they differ regarding the specific modelling activities. (The resulting sequences of modelling activities are visually represented as codelines in the supplementary online material, source 2.)

Table 5 Absolute frequencies of codes (codings) assigned to situations of recognised anomalous data in the data material

Codes for responses to anomalous data during modelling processes		Codings ($n_{Situations}$)
Type 1: inconsistency between observed data and general logic		
Recognition		4
Type 2: inconsistency between observed data and expected data pattern/single datum		
Recognition		46
Explanation	Methodical	20
	With model	7
Change		2
Type 3: inconsistency between observed data and developed model		
Recognition		14
Explanation	Methodical	1
	With model	6
Change	Modification	6
	Rejection	4

Responses to Anomalous Data During Modelling Processes

Findings from the qualitative content analysis show that all three types of situations, in which anomaly can be recognised, occurred in the transcripts (Table 5). Inconsistencies between observed data and general logic (type 1) occurred rather seldom ($n_{Situations} = 4$). Lauren gives an example of a recognised inconsistency between the observed data and general logic.

> Now, 500 ml are coming out. Or what? Please do not. 500 ml came out. […] That is a bit confusing, because of the system, it cannot change its size. *Lauren*

This participant was confused about a specific output because he observed a different output before and thought that this would not be possible because the black box cannot change its general size.

In the majority of situations, inconsistencies between observed data and the expected data pattern or single datum (type 2) were identified ($n_{Situations} = 46$). The participants mostly explained these situations with methodical issues ($n_{Situations} = 20$), for example Cynthia with imprecise measurement:

> Because I am a bit messy when I measure and that is why it does not fit completely. *Cynthia*

Only in two of the 46 situations in which an inconsistency between the observed data and the expected data pattern was recognised the assumptions about the black box' data pattern was changed (Table 5).

Finally, in 14 situations, the pre-service biology teachers recognised data that were not in line with their previous developed model (type 3). Interestingly, these anomalous data were mostly explained with the model ($n_{Situations} = 6$) rather than with methodical issues ($n_{Situations} = 1$). However, in half of the situations in which this type of inconsistency was recognised, participants gave no explanation at all. Nevertheless, the recognition was often followed by the modification ($n_{Situations} = 6$; see Jonathan) or the rejection ($n_{Situations} = 4$; see Cynthia) of the initial model. In the first example (Jonathan), the earlier 'one-compartment model' was changed into a more complex one to consider the observed data.

> It could be a multi-compartment system. And during the first time I had filled only one compartment. *Jonathan*

> I have to reject the model, because it does not make sense with the other things. *Cynthia*

Grouping Responses to Anomalous Data into Response Patterns

Based on the comparison of the eleven *codelines* (supplementary online material, source 3), we found the theoretically expected response patterns described above (Table 3). However, for two of the four response patterns, we identified sub-patterns that distinguish between no reaction to the recognised anomalous data and the rejection of the model because of anomalous data. Table 6 provides the empirically identified response patterns to anomalous data during modelling processes, including labelling and absolute frequencies.

Table 6 Empirically identified response patterns to anomalous data during modelling processes

Response pattern	Recognition	Explanation	Reaction	*n*
No recognition	No	No	No	1 (Celine)
No explanation	Yes	No	No/rejection	2 (Iris, Jim)
Methodical explanation	Yes	Explanation (only methodical)	No/rejection	3 (Carlo, Cynthia, Lauren)
Model-related explanation	Yes	Explanation (also model-related)	Modify model	5 (Angelina, Floyd, Frida, Jenny, Jonathan)

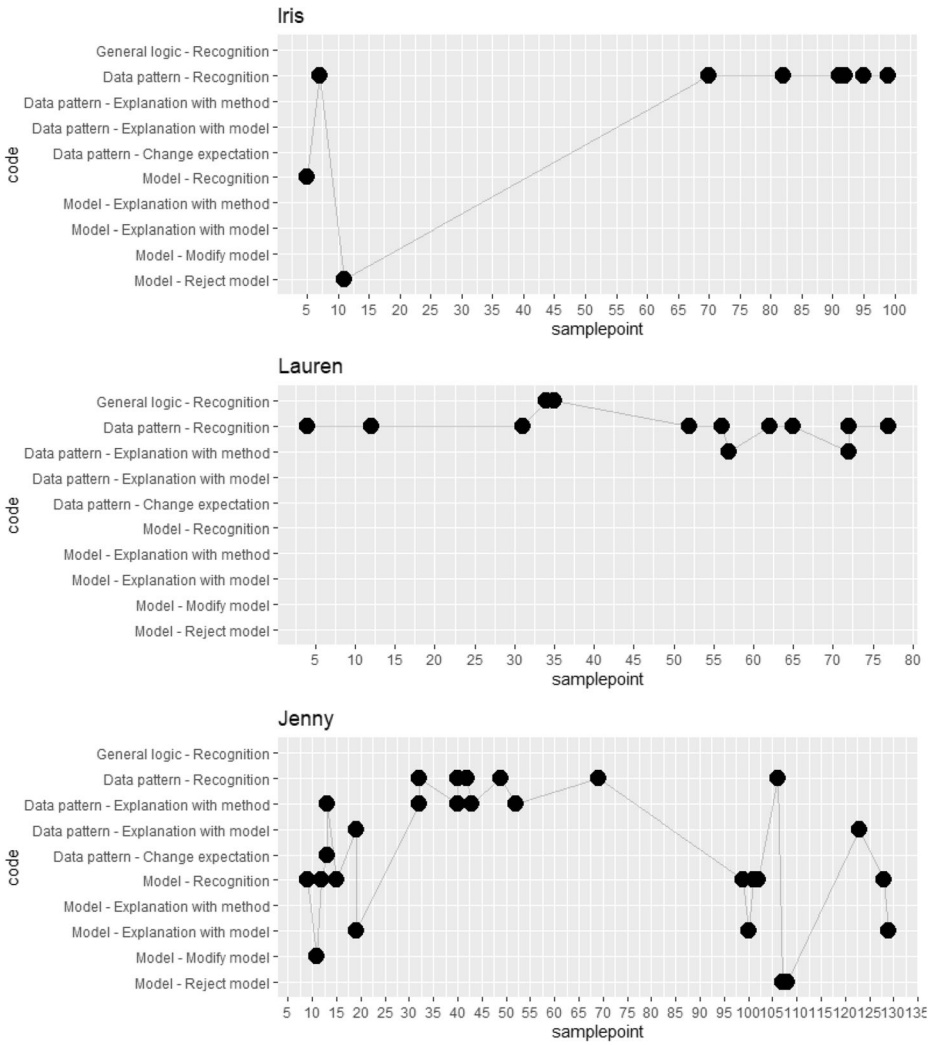

Fig. 2 *Codelines* of three pre-service biology teachers that exemplify the response patterns: *no explanation* (Iris), *methodical explanation* (Lauren), and *model-related explanation* (Jenny)

No recognition of anomalous data was a pattern that was shown by one pre-service biology teacher (Celine). During her modelling process, Celine did not recognise any anomaly, caused by an open exploration of the black box without stating any expectations regarding the output pattern or a hypothetical model. Therefore, no coding with our category system was possible. Figure 2 gives example codelines for the other three response patterns.

The response pattern *no explanation* (Fig. 2, first *codeline*) shows the tendency to recognise anomalous data without offering any explanation for them during the modelling process. However, the recognition of an anomaly can still induce the rejection of the model, since for this kind of reaction, no explanation for the anomalous data is needed.

The response pattern *methodical explanation* (Fig. 2, second *codeline*) is assigned to *codelines* indicating the recognition of anomalous data (mostly of type 2 *inconsistency between observed data and expected data pattern/single datum*) and a tendency to explain these data with methodical issues. Furthermore, subsequent to methodical explanations, modelling processes to explain the anomalous data do not occur in this response pattern or the model is rejected without any explanation (see example in Fig. 2). The rejection of the model is by trend connected with the recognition of type 2 anomalous data (Online source 3).

The pattern *model-related explanation* (Fig. 2, third *codeline*) relates to modelling processes in which the participants provided model-related explanations in addition to methodical explanations and consequently modified the model. Therefore, the modelling process is continued to explain the recognised anomalous data. Generally, the modification of the model is often connected to the recognition of *inconsistency between observed data and developed model* (type 3 anomalous data).

Relation Between Response Patterns to Anomalous Data and Modelling Strategies

Looking for relations between the identified response patterns to anomalous data and the modelling strategies of pre-service biology teachers by contrasting them (Table 7), we found—by tendency—a systematic relation that clusters the pre-service biology teachers in four groups:

1. Celine, Jim, Lauren, and Carlo: focus on exploration in their modelling processes and did not explain anomalous data or explained it mostly methodical;
2. Iris and Cynthia: use developed model for a single prediction and did not explain anomalous data or explained it mostly methodical;

Table 7 The relation between response patterns to anomalous data and modelling strategies

Modelling strategies	Response patterns to anomalous data			
	No recognition	*No explanation*	*Methodical explanation*	*Model-related explanation*
(3b) Repeated prediction				Frida, Jenny
(3a) Single prediction		Iris*	Cynthia*	
(2b) Exploration and model development				Jonathan, Floyd
(2a) Focus on exploration	Celine	Jim	Lauren	Angelina
(1) Exploration only			Carlo	

*These participants reacted to the anomalous data by rejecting their models

3. Frida and Jenny: engage in full cyclic modelling and use model for explanation of recognised anomalous data;
4. Angelina, Floyd, and Jonathan: focus on developing a fitting model and use model for explanation of recognised anomalous data.

Superimposing the codes given in the analysis of responses to anomalous data and the codelines of the modelling processes (see online material source 2) gives further insights into the role of responses to anomalous data during modelling processes for the four groups. Celine, Carlo, Jim, and Lauren generated data, but did not (*exploration only*) or only to a minor extent (*focus on exploration*) engage in modelling by developing and testing a model. These four cases did not use their models to predict the black box' output pattern. The following excerpts exemplify how they recognised anomalous data regarding the expected data pattern during the exploration phase of their modelling processes only:

> Nothing should happen now. Hm, but it is doing something anyway. That's strange. I would estimate, it is something with, that should be 400 ml. […] Okay that is unexpected. *Carlo*

> Yes, okay. Now, there were more than 1000. […] Well, probably it was a mistake earlier. […] Yes, here was a mistake, because I measured this wrong, I think. *Lauren*

However, the pre-service biology teachers in this first group mostly did not explain the anomalous data or if they explained them, they used methodical issues like a lack of experimental skills and did not integrate the anomalous data into their modelling processes.

Iris and Cynthia used their models for stating predictions (*single prediction*) going through the phase exploration, model development, and predicting during their modelling processes. Furthermore, they recognised anomalous data regarding the expected data pattern only while exploring the black box activity, while Cynthia explains these anomalies when developing a model.

> That it is like (…) stairs. I do not know. Like this and it always overflows there. […] But, at this point it does not make sense anymore. That's strange, how can (…). Something has to change inside this, so that no stable pattern, or no stable pattern for me, is perceivable. *Cynthia*

Both recognised anomalies regarding their developed model when testing their stated prediction once.

> I assume, that here the tube goes in. That, so much water is remaining inside, thus the first 400 ml and the tube cannot overrun at the top. That's what I assume, because the second 400 ml flowed through completely. I think when I put 400 ml inside, that 400 will come out again. […] Well, now I'm amazed, because more is coming out than I have expected. Namely a whole litre. *Iris*

Nevertheless, they did not explain this anomaly but still integrated the anomalous data into their modelling processes by rejecting their developed model.

In contrast, Frida and Jenny engaged in a full cyclic modelling process by using their developed models to predict the black box' output pattern (*predicting repeatedly*) and they provided model-related explanations for previous recognised anomalous data. While anomalies regarding the expected data pattern are recognised mostly during the exploration phase of their modelling processes, anomalous data concerning the developed model were recognised during the phases of exploration and prediction.

> Why is nothing coming out now? Why does it store something again? Maybe because it is not full yet. It always empties itself reaching 1,5. And now it has. Before there were five inside. No, we had not nothing inside. Again 1 litre. Now, there would be 1 litre again, theoretically. Now, it should have stored something. But it does not do it. Or? Now, 500 remained inside that means it is not possible that it splits it somehow. *Jenny*

In the modelling processes of these two participants, anomalous data and modelling were intertwined, mostly causing the modification of the model.

However, the other three participants (Angelina, Jonathan, Floyd) that used model-related explanations also integrated the recognised anomalous data into their modelling processes, but only for developing a fitting model and not for testing predictions based on their models. In their modelling processes, this is shown as well by mainly recognising anomalous data regarding the expected data pattern during the modelling phase of exploration. Only Jonathan recognised anomalies that were not in line with his model; however, this was based on verbally externalised analogies and before he tried to develop the model as a graphical representation.

> It could be some kind of overflow valve that flows over when reaching a specific value. […] So, again 400 ml. It should go short over 800.[…] That's interesting. 600 came out. Why are 600 coming out? *Jonathan*

Discussion

The aim of this research is to investigate the role of responses to anomalous data during modelling processes. Previous research in this field mostly focusses on experimentation as a style of scientific thinking, not considering that processes of scientific thinking might differentiate depending on the thinking style in focus (Kind and Osborne 2017). Modelling is a main practice in science (Giere et al. 2006) and is seen as a comprehensive practice that encompasses various aspects of scientific thinking (Lehrer and Schauble 2015). Therefore, modelling as a framework and the used process-based approach to analyse responses to anomalous data provide the opportunity to understand the role of this kind of data on scientific thinking from a broader perspective and not only either on conceptual development (Chinn and Brewer 1998; Hemmerich et al. 2016) or experimentation (Crujeiras-Pérez and Jiménez-Aleixandre 2019; Lin 2007).

For answering our research questions, we analysed how pre-service biology teachers reacted to self-generated anomalous data during modelling processes induced by investigating a water black box (Krell et al. 2017). In the first step of data analysis, we explored the responses to anomalous data, finding that if anomalous data were recognised, the majority of explanations

were based on methodical issues, like imprecise measurement (Table 5). This finding supports the results from previous studies that investigated how undergraduates or students respond to first-hand anomalous data in laboratory settings (Crujeiras-Pérez and Jiménez-Aleixandre 2019; Lin 2007). Furthermore, anomalous data seem to be more salient during the exploration phase of modelling, since the participants mostly recognised anomalous data regarding the expected output pattern of the black box (Table 5). This tendency to focus on the data before a model is developed underlies the importance of knowledge about the phenomenon for modelling processes (Ruppert et al. 2017). Based on the responses to anomalous data, we identified four different response patterns: *no recognition*, *no explanation*, *methodical explanation*, and *model-related explanation*. Similar to previous findings (Crujeiras-Pérez and Jiménez-Aleixandre 2019), one response pattern (*no recognition*) indicates the difficulty to recognise anomalous data in laboratory settings. In contrast to the study by Chinn and Malhotra (2002), this finding might not be influenced by some kind of observation bias (Risinger and Saks 2002) that causes a selective perception during data generation, since the setting of our study with a black box does not induce a specific expectation based on prior knowledge. In the case of our study, this response pattern was caused by not having any expectations or developed models and, therefore, generated data was neither confirming nor anomalous.

In general, our findings regarding the first and second research questions show that responses to anomalous data occur differentially during a databased modelling activity in contrast to analysing products from experimentation processes. Depending on the phase the participants were in the modelling process, either exploration, model development, or prediction (Fig. 1), they responded either by explaining anomalies with methodological issues or they explained it conceptually by using their developed models. Hence, using databased modelling as style of scientific thinking seems to encompass typical responses to anomalous data that have been described from research on first-hand and second-hand anomalous data (Chinn and Brewer 1998; Crujeiras-Pérez and Jiménez-Aleixandre 2019; Hug and McNeill 2008).

Interestingly, we found by tendency a systematic relation between the response patterns to anomalous data and the modelling strategies of the pre-service biology teachers, indicating that model-related explanations for anomalous data are related to more sophisticated modelling strategies (*predicting repeatedly, predicting once*). In these cases, perceived anomalous data is included in the development and modification of the models in order to explain the anomalies. During the exemplary modelling processes of these strategies, handling anomalous data occurred in all three phases of modelling (exploration, development, and prediction, see Fig. 1). However, some participants showing these modelling strategies still connect anomalous data and modelling processes, but by rejecting a model without explaining the anomalous data. In contrast, participants that rely mostly on methodical explanations for anomalous data did indeed develop models but did not connect them with the recognised anomalies. During their modelling processes, this is indicated by handling anomalous data mainly during the modelling phase of exploration. Therefore, explanations based on methodical issues seem to hinder the connection between the evaluation of anomalous data and modelling. This is in line with findings from research on conceptual development showing that the induction of theory change, or in our case model change, by using anomalous data is difficult (Chinn and Brewer 1998). However, from a perspective on experimentation practices, a critical evaluation of methodical causes for anomalous data is important for reflecting aspects of validity and reliability (Gott et al. 2008) and especially first-hand data is qualified for inducing such reflections (Hug and McNeill 2008). Therefore, the reflective use of the response pattern *model-related explanation* that integrates questions of methodical issues may support the connection between the evaluation of

anomalous data and modelling processes. Additionally, difficulties in developing and changing models in order to explain recognised anomalous data might be caused by a lack of knowing alternative theories (Hemmerich et al. 2016) hindering creativity during the modelling process.

Of course, there are some methodologically caused limitations of this study. First, we conducted a qualitative study by analysing responses to anomalous data during modelling processes of a small sample of pre-service biology teachers; therefore, generalisation is strongly limited and further research is needed. Second, our study was contextualised in the rather abstract context of modelling a black box. This has the advantage of reducing the influence of prior knowledge, but, on the other hand, might be criticised as rather artificial (cf. Ruppert et al. 2017). We propose to replicate the findings within more concrete scientific contexts, since this could influence the strength of the expectations on the data and the developed models (Chinn and Brewer 1998; Ruppert et al. 2017). Additionally, considering the influence of experimentation skills like the use of the control of variables strategy might give further insights, especially for responses to anomalous data in laboratory settings. Furthermore, the procedure of this study defines the drawings of the participants as models. Interpreting drawings as externalisations of mental models is discussed as a fruitful approach in addition to the assessment of mental models with interviews or questionnaires (Meister et al. 2018). However, using for example retrospective interviews in a subsequent study would provide further evidence for a valid interpretation of the drawings (Werner et al. 2019).

Implications

Keeping the abovementioned limitations in mind, our study indicates by trend a systematic relation between response patterns to anomalous data and modelling strategies. Consequently, the improvement of responses to anomalous data could be a promising way to foster modelling competencies and vice versa. In order to gain a more holistic understanding on how our findings that focus on competences of scientific modelling can contribute to conceptual science learning, further research that investigates these processes and their interdependence should be conducted using different authentic and content-rich scientific contexts (Ruppert et al. 2017). However, generally our findings suggest to explicitly broach the issue of anomalous data during modelling processes in science education, in order to promote respondents to further evaluate their models and appreciate their role as epistemic tools for sense-making. For example by reflecting about anomalous data and motivate to generate and explain anomalous data with students (Crujeiras-Pérez and Jiménez-Aleixandre 2019). Since explaining anomalous data conceptually and not only by using methodological issues seems to be mostly difficult, providing alternative explanations could be a fruitful approach (Hemmerich et al. 2016). For this, framing handling anomalous data with modelling processes is helpful, since the existence of alternative models for one phenomenon is one of the key aspects of understanding models as epistemic tools (Gilbert and Justi 2016; Schwarz and White 2005). We are convinced that an integrated approach to anomalous data and modelling could lead to deeper insights into the role of data in scientific thinking processes.

Funding Information Open Access funding provided by Projekt DEAL.

References

Campbell, T., Oh, P., Maughn, M., Kiriazis, N., & Zuwallack, R. (2015). A review of modeling pedagogies: Pedagogical functions, discursive acts, and technology in modeling instruction. *Eurasia Journal of Mathematics, Science & Technology Education, 11*, 159–176.

Chinn, C. A., & Malhotra, B. A. (2002). Children's responses to anomalous scientific data: How is conceptual change impeded? *Journal of Educational Psychology, 94*(2), 327–343.

Chinn, C. A., & Brewer, W. F. (1998). An empirical test of a taxonomy of responses to anomalous data in science. *Journal of Research in Science Teaching, 35*(6), 623–654.

Chinn, C. A., & Brewer, W. F. (2001). Models of data: a theory of how people evaluate data. *Cognition and Instruction, 19*(3), 323–393.

Clement, J. (2009). *Creative model construction in scientists and students*. Dordrecht: Springer.

Crujeiras-Pérez, B., & Jiménez-Aleixandre, M. P. (2019). Students' progression in monitoring anomalous results obtained in inquiry-based laboratory tasks. *Research in Science Education, 49*(1), 243–264.

Dounas-Frazer, D., Ríos, L., Pollard, B., Stanley, J., & Lewandowski, H. (2018). Characterizing lab instructors' self-reported learning goals to inform development of an experimental modeling skills assessment. *Physical Review Special Topics - Physics Education Research, 14*, 1–24. https://doi.org/10.1103/PhysRevPhysEducRes.14.020118.

Duncan, R. G., Chinn, C. A., & Barzilai, S. (2018). Grasp of evidence: problematizing and expanding the next generation science standards' conceptualization of evidence. *Journal of Research in Science Teaching, 55*(7), 907–937.

Ericsson, K., & Simon, H. (1998). How to study thinking in everyday life. *Mind, Culture, and Activity, 5*, 178–186.

Fischer, F., Kollar, I., Ufer, S., Sodian, B., Hussmann, H., Pekrun, R., et al. (2014). Scientific reasoning and argumentation: advancing an interdisciplinary research agenda in education. *Frontline Learning Research, 2*(3), 28–45.

Giere, R., Bickle, J., & Mauldin, R. (2006). *Understanding scientific reasoning*. London: Thomson Learning.

Gilbert, J. K., & Justi, R. (2016). *Modelling-based teaching in science education*. Cham: Springer.

Godfrey-Smith, P. (2006). The strategy of model-based science. *Biology and Philosophy, 21*, 725–740. https://doi.org/10.1007/s10539-006-9054-6.

Göhner, M. & Krell, M. (2018). Modellierungsprozesse von Lehramtsstudierenden der Biologie. *Erkenntnisweg Biologiedidaktik, 17,* 45–61.

Göhner, M. & Krell, M. (2019). Entwicklung einer Typologie von Modellierungsstrategien. Paper presented at the 21. Internationale Frühjahrsschule der Fachsektion Didaktik der Biologie im VBIO, 18.-21.03.2019, Köln.

Gott, R., Duggan, S., Roberts, R., & Hussain, A. (2008). *Concepts of evidence*. School of education: University of Durham.

Günther, S. L., Fleige, J., Upmeier zu Belzen, A. & Krüger, D. (2019). Using the case method to foster preservice biology teachers' content knowledge and pedagogical content knowledge related to models and modeling. *Journal of Science Teacher Education 30* (4), 321–343. https://doi.org/10.1080/1046560X.2018.1560208.

Hartmann, S., Upmeier zu Belzen, A., Krüger, D., & Pant, H. (2015). Scientific reasoning in higher education. *Zeitschrift für Psychologie, 223*, 47–53. https://doi.org/10.1027/2151-2604/a000199.

Hemmerich, J. A., Van Voorhis, K., & Wiley, J. (2016). Anomalous evidence, confidence change, and theory change. *Cognitive Science, 40*(6), 1534–1560.

Hetmanek, A., Engelmann, K., Opitz, A., & Fischer, F. (2018). Beyond intelligence and domain knowledge: scientific reasoning and argumentation as a set of cross-domain skills. In *Scientific reasoning and argumentation: The roles of domain-specific and domain-general knowledge* (pp. 203–226). New York: Routledge.

Hug, B., & McNeill, K. L. (2008). Use of first-hand and second-hand data in science: does data type influence classroom conversations? *International Journal of Science Education, 30*(13), 1725–1751.
Jeong, H., Songer, N. B., & Lee, S.-Y. (2007). Evidentiary competence: Sixth Graders' understanding for gathering and interpreting evidence in scientific investigations. *Research in Science Education, 37*(1), 75–97.
Justi, R., & Gilbert, J. K. (2003). Teachers' views on the nature of models. *International Journal of Science Education, 25*, 1369–1386.
Justi, R., & van Driel, J. (2005). The development of science teachers' knowledge on models and modelling: promoting, characterizing, and understanding the process. *International Journal of Science Education, 27*, 549–573.
Khan, S. (2008). What if scenarios for testing student models in chemistry. In J. Clement & M. Rea-Ramirez (Eds.), *Model based learning and instruction in science* (pp. 139–150). Dordrecht: Springer.
Khan, S. (2011). What's missing in model-based teaching. *Journal of Science Teacher Education, 22*, 535–560.
Kind, P. E. R., & Osborne, J. (2017). Styles of scientific reasoning: a cultural rationale for science education? *Science Education, 101*(1), 8–31.
KMK (Ed.). (2005). *Bildungsstandards im Fach Biologie für den Mittleren Schulabschluss [Biology education standards for the Mittlere Schulabschluss]*. München & Neuwied: Wolters Kluwer.
Knuuttila, T. (2011). Modelling and representing: An artefactual approach to model-based representation. *Studies in History and Philosophy of Science, 42*, 262–271.
Krell, M., & Hergert, S. (2020). The blackbox approach: Analyzing modeling strategies. In A. Upmeier zu Belzen, D. Krüger, & J. van Driel (Eds.), Towards a competence-based view on models and modeling in science education. (pp. 147–160). Cham: Springer. https://doi.org/10.1007.
Krell, M., & Krüger, D. (2016). Testing models: A key aspect to promote teaching-activities related to models and modelling in biology lessons? *Journal of Biological Education, 50*, 160–173. https://doi.org/10.1080/00219266.2015.1028570.
Krell, M., Walzer, C., Hergert, S., & Krüger, D. (2017). Development and application of a category system to describe pre-service science teachers' activities in the process of scientific modelling. *Research in Science Education*. https://doi.org/10.1007/s11165-017-9657-8.
Krell, M., Redman, C., Mathesius, S., Krüger, D., & van Driel, J. (2018). Assessing pre-service science teachers' scientific reasoning competencies. *Research in Science Education*. https://doi.org/10.1007/s11165-018-9780-1.
Lederman, N., & Abd-El-Khalick, F. (2002). Avoiding de-natured science. In W. McComas (Ed.), *The nature of science in science education* (pp. 83–126). Dordrecht: Kluwer.
Lehrer, R., & Schauble, L. (2015). The development of scientific thinking. *Handbook of child psychology and developmental science, 2*(7), 671–714.
Lin, J.-Y. (2007). Responses to anomalous data obtained from repeatable experiments in the laboratory. *Journal of Research in Science Teaching, 44*(3), 506–528.
Mason, L. (2001). Responses to anomalous data on controversial topics and theory change. *Learning and Instruction, 11*(6), 453–483.
Meister, S., Zimmerman, C., & Upmeier zu Belzen, A. (2018). Visualizing pre-service biology teachers´ conceptions about population dynamics in ecosystems. Science Education Review Letters (2018), S.7-12. Advance online publication. https://doi.org/10.18452/19423.
NGSS Lead States. (2013). *Next generation science standards: for states, by states*. Washington, DC: The National Academies Press.
Nicolaou, C., & Constantinou, C. (2014). Assessment of the modeling competence. *Educational Research Review, 13*, 52–73.
Oh, P., & Oh, S. (2011). What teachers of science need to know about models. *International Journal of Science Education, 33*, 1109–1130.
Passmore, C., & Svoboda, J. (2012). Exploring opportunities for argumentation in modelling classrooms. *International Journal of Science Education, 34*, 1535–1554.
Passmore, C., Gouvea, J. S., & Giere, R. (2014). Models in science and in learning science: Focusing scientific practice on sense-making. In M. R. Matthews (Ed.), *International handbook of research in history, philosophy and science teaching* (pp. 1171–1202). Dordrecht: Springer.
Passmore, C., Schwarz, C., & Mankowski, J. (2017). Developing and using models. In C. Schwarz, C. Passmore, & B. Reiser (Eds.), *Helping students make sense of the world using next generation science and engineering practices* (pp. 109–134). Arlington: NSTA Press.
Pickering, M., & Monts, D. L. (1982). How students reconcile discordant data: a study of lab report discussions. *Journal of Chemical Education, 59*, 794–796.
Risinger, D. M., & Saks, M. J. (2002). The Daubert/Kumho implications of observer effects in forensic science: Hidden problems of expectation and suggestion. *California Law Review, 90*(1), 1–56.

Rönnebeck, S., Bernholt, S., & Ropohl, M. (2016). Searching for a common ground – a literature review of empirical research on scientific inquiry activities. *Studies in Science Education, 52*(2), 161–197. https://doi.org/10.1080/03057267.2016.1206351.

Ruppert, J., Duncan, R., & Chinn, C. (2017). Disentangling the role of domain-specific knowledge in student modeling. *Research in Science Education, 49*, 921–948. https://doi.org/10.1007/s11165-017-9656-9.

Samarapungavan, A. (2018). Construing scientific evidence: The role of disciplinary knowledge in reasoning with and about evidence in scientific practice. In F. Fischer, K. Englemann, J. Osborne, & C. A. Chinn, (Eds.), *Scientific reasoning and argumentation: The roles of domain-specific and domain-general knowledge* (pp. 66–86). London: Routledge.

Schreier, M. (2012). *Qualitative content analysis in practice*. Thousand Oaks: Sage.

Schwarz, C., & White, B. (2005). Metamodeling knowledge: developing students' understanding of scientific modeling. *Cognition and Instruction, 23*, 165–205.

Svoboda, J., & Passmore, C. (2013). The strategies of modeling in biology education. *Science & Education, 22*, 119–142.

Toplis, R. (2007). Evaluating science investigations at ages 14–16: dealing with anomalous results. *International Journal of Science Education, 29*(2), 127–150.

VCAA [Victorian Curriculum and Assessment Authority]. (2016). Victorian Curriculum: F-10. Melbourne, VIC. Retrieved from http://victoriancurriculum.vcaa.vic.edu.au/science/curriculum/f-10.

VERBI Software. (2016). *MAXQDA Analytics 2018 Plus [computer programme]*. Berlin: VERBI.

Werner, S., Förtsch, C., Boone, W., von Kotzebue, L., & Neuhaus, B. (2019). Investigating how German biology teachers use three-dimensional physical models in classroom instruction: a video study. *Research in Science Education, 49*, 437–463. https://doi.org/10.1007/s11165-017-9624-4.

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. New York, NY: Springer-Verlag.

Affiliations

Sabine Meister[1] **· Moritz Krell**[2] **· Maximilian Göhner**[2] **· Annette Upmeier zu Belzen**[1]

[1] Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany

[2] Freie Universität Berlin, Schwendenerstraße 1, 14195 Berlin, Germany