



Preservice Science Teachers' Strategies in Scientific Reasoning: the Case of Modeling

Maximilian Göhner¹  · Moritz Krell¹ 

Published online: 22 July 2020
© The Author(s) 2020

Abstract

The development of scientific reasoning competencies is a key goal of science education. To better understand the complex construct of scientific reasoning, which includes modeling as one style of reasoning, thorough investigations of the underlying processes are needed. Therefore, in this study, a typology of preservice science teachers' modeling strategies was developed. Thirty-two preservice science teachers were videotaped while engaging in the modeling task of investigating a black box. Following a qualitative content analysis, sequences of modeling activities were identified. By transforming these sequences of modeling activities into state transition graphs, six types of modeling strategies were derived, differing in the homogeneity and complexity of their modeling processes. The preservice science teachers engaged in activities of (1) exploration only; (2a) exploration and development with a focus on development; (2b) exploration and development with a focus on exploration; (2c) exploration and development, balanced; (3a) exploration, development, and drawing predictions from a model once; or (3b) exploration, development, and repeatedly drawing predictions from a model. Finally, this typology is discussed regarding the process of its development and its potential to inform and guide further research as well as the development of interventions aiming to foster competencies in scientific modeling.

Keywords Scientific reasoning · Modeling · Preservice science teachers · Typology · Modeling strategies · Qualitative content analysis

Introduction

The development of scientific reasoning competencies is considered a key goal of twenty-first century education (Osborne 2013), as those competencies are required for active participation in science- and technology-centered societies. Consequently, scientific reasoning

✉ Maximilian Göhner
maximilian.goehner@fu-berlin.de

¹ Biology Education, Freie Universität Berlin, Schwendenerstraße 1, Berlin 14195, Germany

competencies are highlighted in curricular documents worldwide (e.g., BCMOE 2019; KMK 2005; NGSS Lead States 2013; VCAA 2016). To support the development of students' competencies, preservice science teachers need to obtain scientific reasoning competencies as part of their professional competencies (Khan and Krell 2009; Mathesius et al. 2016). These scientific reasoning competencies can be defined as “a complex construct, which encompass[es] the abilities needed for scientific problem solving as well as to reflect on this process at a meta-level” (Krell et al. 2018, p. 2; cf. Morris et al. 2012), including generic cognitive abilities, such as the use of analogies and decision making (Nersessian 2002), and the ability to apply content, procedural, and epistemic knowledge for problem solving (Kind and Osborne 2017). In line with this definition, research on scientific reasoning competencies is a similarly complex matter, leading to a high diversity of studies concentrating on different contexts and foci and, consequently, using different theoretical frameworks and methodical approaches (Koslowski 2013; Rönnebeck et al. 2016). Additionally, it is argued that there is no single form of scientific reasoning; Kind and Osborne (2017), for example, have recently proposed six different styles of scientific reasoning, which include mathematical deduction, experimental evaluation, hypothetical modeling, categorization and classification, probabilistic reasoning, and historical-based evolutionary reasoning. For some styles of scientific reasoning, especially experimental evaluation (Hammann et al. 2008; Schauble et al. 1991a), individuals' processes, strategies, and understandings have been broadly investigated and described in science education research, but others, like hypothetical modeling, lack detailed research (Nicolaou and Constantinou 2014). Hence, a thorough investigation of the processes involved in different styles of scientific reasoning, and their comparison could lead to deeper insights into the structure of the complex construct of scientific reasoning and would help develop a “big picture of what [scientific reasoning] might be” (Kind and Osborne 2017, p. 27). The present study focuses on the style of *hypothetical modeling* and adds to the research on scientific reasoning, by presenting a process-oriented analysis of preservice science teachers' strategies in the process of scientific modeling.

Analyzing Scientific Reasoning Processes

Research on the dynamic and process-oriented aspects of behavior often relies on the systematic observation of predefined behavior codes (Bakeman and Gottman 1997), which are used as the fundamental elements of analysis. In research regarding scientific reasoning, the term “activities” is most commonly used to describe these elements of analysis (e.g., Khan and Krell 2009; Krell et al. 2019). Thus, different frameworks with specific sets of activities can be flexibly utilized to account for varying settings, samples or goals (cf. Rönnebeck et al. 2016), levels of detail (e.g., Chinn and Malhotra 2002; Fischer et al. 2014; Lawson 2004), or theoretical emphases on specific styles of reasoning (e.g., on modeling; Giere et al. 2006). By observing and identifying the different epistemic activities of individuals in their respective order, the sequences of activities can be extracted, which simplify the complex process of scientific reasoning to a level where patterns can be made visible. To date, empirical studies have shown that the activities of scientific reasoning do not follow a strictly predetermined order (e.g., Harwood 2004; Klahr and Dunbar 1988; Reiff et al. 2002), so scientific reasoning can be considered a creative problem-solving process (Zimmerman 2005). However, systematic differences in the reasoning processes of novices and experts have been found (e.g., Čančula et al. 2015; Klahr 2002).

How Typologies of Scientific Reasoning Can Be Beneficial for Science Education

Many studies aim to generalize patterns in scientific reasoning processes by proposing types or classes of reasoning strategies, which can then be used for diagnostic purposes or in targeted interventions (Klahr et al. 1993; Schauble et al. 1991b). One example of an analysis of scientific reasoning processes that can be seen as a central framework in science education research is the work of Klahr and Dunbar (1988), which has been used and adapted in numerous further studies (e.g., Aizpurua et al. 2018; Hammann et al. 2008; Krell 2018; Neumann et al. 2019). Based on case studies, Klahr and Dunbar (1988) developed the *Scientific Discovery as Dual Search* (SDDS) model. They argue that scientific reasoning is a problem-solving task taking place in two problem spaces: the hypothesis space and the experiment space. They distinguish two strategies based on the occurrence of an experimental activity at a specific moment in the scientific reasoning process, classifying subjects into “experimenters” and “theorists.” They further describe the differences regarding the occurrence and frequency of activities assigned to one space or the other between these two classes. When given a scientific problem, “theorists” search for new hypotheses in the hypothesis space, while “experimenters” rely on experimentation to induce regularities without explicitly stating other hypotheses. Studies have indicated that experts rather behave like “theorists” and tend to search the hypothesis space, which could result from differences in prior knowledge (e.g., Rasmussen 1981), allowing them to solve scientific problems more efficiently (Klahr 2002). Based on the SDDS model, Klahr et al. (1993) described commonly found differences in reasoning processes of adults and children. They identified different problems children encounter in the development of scientific reasoning skills. For example, children often have problems constraining their search in the hypothesis or experiment space, leading to the testing of inappropriate hypothesis or design of inappropriate experiments. These findings, which were based on the application of the SDDS model and their typology of “experimenters” and “theorists,” allowed them to draw implications for educational settings (e.g., should educational settings by design be more constrained in regard to the amount of hypotheses and possible experiments if they target children), which were later incorporated into further studies (Klahr and Nigam 2004).

Another study combining the identification of experimentation activities with the generalization of these patterns was implemented by Schauble et al. (1991a). They investigated strategy differences in good and poor learners in an experimental setting, focusing on five “classes of student behavior.” These classes include the general activity level and the four activities of planning, evidence generation, data management, and evidence interpretation. The authors distinguish between an “engineering” and a more advanced “scientific” model of investigating scientific phenomena. Subjects utilizing the engineering model carry out an investigation until a specific effect is reached and then stop, while subjects utilizing the scientific model examine all variables and their influence on a given system. Additionally, good learners use a control of variables strategy more frequently, generate more hypotheses, keep more systematic records, and develop more goal-oriented plans than poor learners. The authors further used the distinction between an engineering and a scientific model of experimentation to develop a learning progression in which students were able to stepwise develop their individual experimentation processes to increasingly resemble the latter (Schauble et al. 1991b). This effect was observed to be stronger if the problems to be solved were introduced in an order of increasing analytical complexity.

Both examples illustrate the fruitfulness of typologies of scientific reasoning for science education; typologies might guide further research and contribute to theory development (e.g., Klahr and Dunbar 1988) as well as be used to develop specific educational settings and interventions (Schauble et al. 1991a, b; Klahr et al. 1993; Klahr and Nigam 2004). However, while different strategies (and related typologies) have been proposed for scientific experimentation, such findings are widely missing for scientific modeling.

Scientific Modeling as One Style of Scientific Reasoning

The relationship between scientific modeling and scientific reasoning is extensively discussed in science education research. As introduced above, six styles of scientific reasoning have been proposed (Kind and Osborne 2017), with each style having a specific set of ontological, procedural, and epistemic resources needed for reasoning. Kind and Osborne (2017) further argue that the six styles of scientific reasoning offer a comprehensive schema for the construct of scientific reasoning. The focus of this study is on one style that is quite established in science education research: scientific modeling (e.g., Clement 2000; Gilbert 2004; Windschitl et al. 2008). Although described as one of the six styles by Kind and Osborne (2017), various other authors specifically emphasize the importance of modeling for reasoning in science; Giere (1999), for example, claims that “scientific reasoning is to a large extent model-based reasoning” (p. 56). Similarly, Lehrer and Schauble (2015) describe scientific reasoning (or scientific thinking, respectively) as a modeling process that encompasses various other practices. Thus, modeling can also be seen as an overarching ability, needed for all processes of scientific reasoning. For this study, however, the specific relationship between modeling and other styles of reasoning in science (i.e., whether modeling is one of various styles or an overarching style encompassing the others) is not of central importance. Therefore, in the present article, scientific modeling will be considered as one style of scientific reasoning and thus needs to be specifically investigated in science education.

From an educational perspective, modeling has the potential to contribute to three of the main objectives of science education (cf. Hodson 2014): *learning science* by learning major models as products of science; *learning to do science* by developing practices to create and evaluate models; and *learning about science* by appreciating the role of models as hypothetical entities and epistemic tools in science (Justi and Gilbert 2003). Consequently, modeling is also highlighted in recent curricular documents as a core scientific practice (e.g., NGSS Lead States 2013), and it has been widely acknowledged that research-based learning essentially consists of the construction, testing, revision, and application of models (e.g., Schwarz et al. 2017; Windschitl et al. 2008). Therefore, scientific modeling has been identified as a potentially useful and relevant example of scientific reasoning and will be the focus of the present article.

Towards a Typology of Modeling Strategies in Science Education

To identify modeling strategies, the present study can build on an extensive body of research on models and modeling in science education, whereby a lot of studies were published already during the 1980s (e.g., Brown and Clement 1989; Gilbert and Osborne 1980; Nersessian 1992) or more recently (e.g., Clement and Rea-Ramirez 2008; Krell et al. 2019; Passmore et al. 2014). In these studies, different facets of knowledge and abilities related to models and modeling of various sample groups have been investigated, such as those of students (e.g., Grosslight et al. 1991; Krell et al. 2014a, b), teachers (van Driel and Verloop 1999; Windschitl

et al. 2008), and experts (e.g., Bailer-Jones 2002; Clement 2008). Other studies have focused on model-based learning and teaching, that is, how to use modeling approaches to enable students to better understand scientific concepts (e.g., Acher et al. 2007; Passmore et al. 2009).

In all these studies, one common finding is that most sample groups seem to have difficulties in understanding the predictive power of models and their role as epistemic tools for scientific reasoning (Krell and Krüger 2016; Passmore et al. 2014). Additionally, even when models are used and understood as epistemic tools, multiple models representing alternative hypotheses are rarely considered (Grosslight et al. 1991). However, it is precisely this characteristic of models as epistemic tools for prediction that is emphasized in science education curricula and standard documents worldwide (e.g., BCMOE 2019; KMK 2005; NGSS Lead States 2013; VCAA 2016). Moreover, reviews propose that research regarding models and modeling in science education mostly focuses on cognitive and metacognitive aspects, so that “meta-modeling knowledge is overevaluated” (Nicolaou and Constantinou 2014, p. 72); detailed and process-oriented analyses of modeling strategies, which could be used for diagnostic purposes or interventions (cf. Klahr et al. 1993; Schauble et al. 1991b), are mostly missing in science education research (Louca and Zacharia 2012; Nicolaou and Constantinou 2014). One exception is the study by Sins et al. (2005), who investigated the types of reasoning, foci of conversation, and types of argumentation used by novices during computer-based modeling. The authors found that more successful students tended to justify their reasoning with prior knowledge and viewed the model as a whole, while the less successful students showed model-fitting behavior (cf. engineering model of experimentation; Schauble et al. 1991a). While the study of Sins et al. (2005) shows how the analysis of sequential activities can be further quantified by using, for example, *z*-scores, they did not distinguish between different strategies or types of modeling processes, which could potentially provide further insights into the nature of different modeling strategies applied by, for example, preservice science teachers.

The present article aims to fill this gap by developing a typology of modeling strategies of preservice science teachers, which can then be further used in science education to distinguish individual modeling processes and enable researchers and teachers to identify points of leverage for interventions on an individual level.

Methodological Frame of the Study and Research Questions

To ensure the quality of the analysis and enable comparisons between the present study and the processes and strategies of reasoning involved in other styles of reasoning, the methods of analysis and the development of the typology and its interpretation must be made as transparent as possible. Therefore, different evaluation criteria for the classification and development of typologies are considered. The analysis is evaluated by the appropriateness of the considered attributes, the minimization of within-group variance, and the maximization of between-group variance (Bailey 1994). The model of empirically grounded type construction (Kluge 2000) is applied during the analysis to maintain methodical flexibility, while achieving a systematic and transparent development process for the typologies. The model describes four steps in the type construction process (Fig. 1). The first three steps can be repeated multiple times to evaluate the appropriateness of possible (combinations of) attributes.

Hence, in the present study, relevant analysis dimensions will be identified (step 1), grouped, and investigated in terms of the empirical regularities (step 2) and meaningful

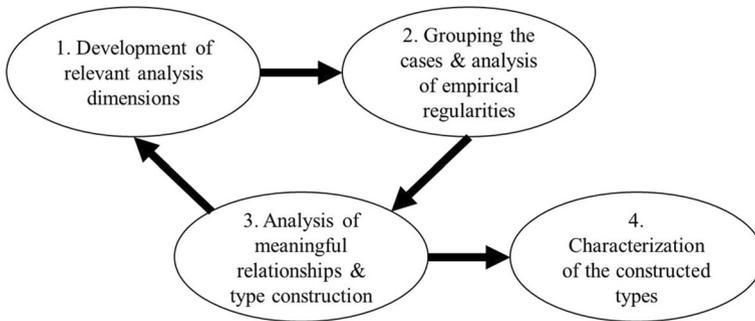


Fig. 1 Model of empirically grounded type construction (Kluge 2000)

relationships (step 3) within modeling processes to identify different types that can be characterized as modeling strategies (step 4). More specifically, the following four research questions will be discussed, whereby each question relates to one step of the model of empirically grounded type construction (Fig. 1):

- 1) Which attributes are appropriate for developing an empirically grounded typology of preservice science teachers' modeling strategies (cf. Fig. 1, step 1)?
- 2) To what extent can modeling processes of preservice science teachers be appropriately grouped based on the selected attributes and consideration of the criteria of minimizing within-group heterogeneity and maximizing between-group heterogeneity (cf. Fig. 1, step 2)?
- 3) To what extent can meaningful relationships be found between cases within the identified groups as well as between the identified groups (cf. Fig. 1, step 3)?
- 4) Which types of preservice science teachers' modeling strategies can be derived from the type construction process (cf. Fig. 1, step 4)?

Methods

Context

The context of this study is the first phase of science teacher education in Germany, which takes place at the university. In this first phase, science teachers are requested to develop basic professional knowledge and competencies that they will need as science teachers (Neumann et al. 2017), including knowledge and competencies related to inquiry and reasoning in science (KMK 2019). In Germany, preservice science teachers usually study two subjects (i.e., two future teaching subjects) within a bachelor's (six semesters) and a subsequent master's program (four semesters) before leaving university and entering the second phase of teacher education (internship). Existing empirical studies propose a significant positive development of German preservice science teachers' scientific reasoning competencies over their course of studies (e.g., Hartmann et al. 2015; Krüger et al. 2020). Students in later semesters are, therefore, expected to have advanced competencies related to inquiry and reasoning in science.

Setting

To induce scientific reasoning processes, a black box setting was applied (Lederman and Abd-El-Khalick 2002). The black box approach is established in science education research to study processes of scientific thinking and modeling (Khan and Krell 2019; 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. It was shown that black box approaches are suitable to elicit modeling processes, with models used as epistemic tools to discover the black box (Krell et al. 2019; Passmore and Svoboda 2012). In this study, the black box approach was preferably chosen over an authentic and content-rich scientific problem, in order to reduce the influence of prior knowledge on the preservice science teachers' modeling processes. However, it is recognized that this might be one limitation of the study as some authors emphasize the important role of content knowledge for scientific modeling (e.g., Ruppert et al. 2019). In the present article, a water black box was used. It can be explored by filling the black box with water (input), which then results in measurable outputs of water (see Krell et al. 2019 for a detailed description of the black box).

Participation in the black box activity was voluntary and not connected to any university courses or obligatory parts of the curriculum. Researchers and participants had no formal relationships to one another. Each participant was briefly introduced to the study and signed an informed consent. To gain additional insights into the scientific reasoning processes, all participants were asked to concurrently think aloud (Leighton and Gierl 2007), which was practiced with three short questions. After these preparations, the participants were brought into a room equipped with three video cameras, the black box, some pre-filled glasses of water, and a chalkboard. The researcher briefly explained the basic functionality of the black box using a prepared script and provided the following task: "Draw a model of the inside of the black box." Participants were informed that there are no time constraints (mean length of the activity for all participants: 1 h and 11 min). The first author stayed in the room during the process of modeling to prevent any technical errors and, if necessary, to remind the participants to concurrently think aloud; otherwise, he did not intervene.

Sampling

Data from 32 participants were collected and analyzed for the present study. The participants' academic progress ranged from first (i.e., bachelor) to tenth semester (i.e., master). The participants were aged between 17 and 39 (mean age was 25). To further maximize the heterogeneity of participants and make the observation of different strategies more likely, theoretical sampling was used. Preservice science teachers were selected based on two criteria. The first criterion was the preservice science teachers' achievement in an established multiple-choice test instrument, which has been previously developed to assess scientific reasoning competencies and encompasses tasks about scientific investigations and scientific modeling (Krüger et al. 2020). The second criterion was the preservice science teachers' achievement in the so called I-S-T 2000R instrument (Liepmann et al. 2007). This paper-pencil instrument assesses general cognitive abilities. It also includes tasks of spatial transformation, which is being argued to be one ability needed for modeling (Nersessian 2002). Both criteria were assessed before participants were invited to the black box activity, and only those preservice science

teachers differing one standard deviation from the tested population's mean scores on both tests were asked to participate.

Data Analysis

The participants' videos were transcribed verbatim, including all verbalizations as well as behavioral aspects deemed important for the analysis (these included making an input or observing an output, going to the board, or cleaning the board). The videos and corresponding transcripts were analyzed by the first author and one additional rater (trained student assistant) following a qualitative content analytic approach (Schreier 2012). To improve the comparability of the results and enhance the workflow, the transcripts were coded watching the videos simultaneously.

Each participant's modeling process was split into individual activities of scientific modeling. For this purpose, a framework consisting of 19 different activities of modeling was utilized (Table 1, i.e., category system), which was developed in a previous study (Khan and Krell 2019; Krell et al. 2019) and integrates frameworks used to describe the processes of analogical reasoning (Brown and Clement 1989) and scientific reasoning through models (Giere et al. 2006). This framework also allows the assignment of every activity to one of the three modeling phases of exploration, model development, or making predictions. While this previously developed category system includes all activities of modeling that have been identified within the whole sample of the previous study (Khan and Krell 2019; Krell et al. 2019), the present study further applies this category system in order to find patterns and identify types for individual preservice teachers' modeling strategies based on the sequence of these modeling activities. Additionally, it is worth mentioning that analogies (Code 6) can be activated throughout the whole process wherein participants are able to evaluate their ideas and mental models verbally; however, the methodical approach in the present study requires an externalized model in form of a drawing.

Table 1 Category system with 19 modeling activities (Krell et al. 2019)

Number	Code	Phase
1	Perception of a phenomenon	Exploration
2	Exploration of the system: input/output (exploratory)	
3	Exploration of the system: summarizing/describing observations	
4	Exploration of the system: input/output (pattern detection)	
5	Exploration of the system: recognizing patterns	
6	Activation of analogies and experiences	Development
7	Development of model: graphically develop model	
8	Development of model: change model to optimize consistency	
9	Development of model: change model to optimize representation	
10	Development of model: reject model due to poor consistency/representation	
11	Evaluation of consistency and representation: evaluate consistency	
12	Evaluation of consistency and representation: evaluate representation	
13	Finding of consistency and representation	Prediction
14	Deduction of predictions	
15	Evaluation of predictions: input/output (to test predictions)	
16	Evaluation of predictions: confirmation of prediction	
17	Evaluation of predictions: falsification of prediction	
18	Modification/rejection of model: change model due to falsified predictions	
19	Modification/rejection of model: reject model due to falsified predictions	

Table 2 exemplifies how the transcripts were divided into single activities for a short transcript excerpt. This excerpt covers minutes 24 to 30 of one respondent’s 66-min-long modeling process. Hence, the excerpt begins in the middle of the modeling process, after an initial exploration phase and the development of an initial model.

To ensure the reliability of the analysis (intrarater agreement), every video and corresponding transcript was coded twice by the first author. To ensure the intersubjectivity of the analysis (interrater agreement), the videos and transcripts were then coded again by the trained student assistant, followed by a discussion and resolution of any disagreements, which might have occurred. Cohen’s Kappa suggests acceptable to high agreements (Table 3). Disagreements were discussed and resolved between both raters. The identified sequences of activities were then used for further analysis related to the steps of the model of empirically grounded type construction (Fig. 1). The respective approaches to data analysis are, therefore, explained in the sections below.

Developing a Typology of Modeling Strategies

In the following, the findings are presented according to the research questions, which relate to the steps of the model of empirically grounded type construction. Two cycles will be presented

Table 2 An excerpt of the transcript of one participant, exemplarily showing how the transcripts were divided into single activities

Time [min:s]	Paragraph	Text	Code
24:46	29	Let us see how much water I get back. [input and observing the output] Ok, the missing 100 came out, after I put in 2900 ml in total	2 exploration of the system: input/output (exploratory)
25:14	30	Now it should be empty again	3 exploration of the system: summarizing/describing observations
25:58	31	[input] I put in 400 again to fill up the first glass [observes output]	4 exploration of the system: input/output (pattern detection)
26:19	32	Ok, this could be what happened	5 exploration of the system: Recognizing patterns
26:43	33	[going to the board] Ok, now I will try to draw this step by step. At first, there is a chamber containing 400 ml. [leaves the board]	7 development of model: graphically develop model
27:11	34	I will put in 100 ml, to see if I can get it to overflow.	14 deduction of predictions
27:53	35	[input of 100 ml] Yes, something is overflowing, let us see if these are 100 ml. [observing the output]	15 evaluation of predictions: input/output (to test predictions)
28:24	36	Ok, either the chamber fits 50 more or the 50 ml went somewhere else	17 evaluation of predictions: Falsification of prediction
28:52	37	I will try again with the 100, let us see how much comes out. [input, observing the output] It is 50 again	2 exploration of the system: input/output (exploratory)
29:13	38	[going to the board] I get nearly half of it back, which would suggest that the chamber fits 400 and when it overflows one half is going out and another half in a different chamber. [leaving the board]	18 modification/rejection of model: change model due to falsified predictions

Transcribed behavioral aspects (taken from the video) are added in brackets

Table 3 Overview of Cohen’s Kappa calculations

	Min	Max	Mean
Intrarater-agreement	0.67	0.95	0.84
Interrater-agreement	0.45	0.93	0.77

focusing on the occurrence of activities (cycle 1) and the transitions between activities (cycle 2), respectively.

Cycle 1, Step 1: Occurrence of Activities as a Relevant Analysis Dimension

Based on the studies described above, the occurrences of activities (Klahr and Dunbar 1988; Schauble et al. 1991a) were identified as the first relevant dimension of analysis. In the first cycle of the development of the typology, the sequence of activities conducted by the

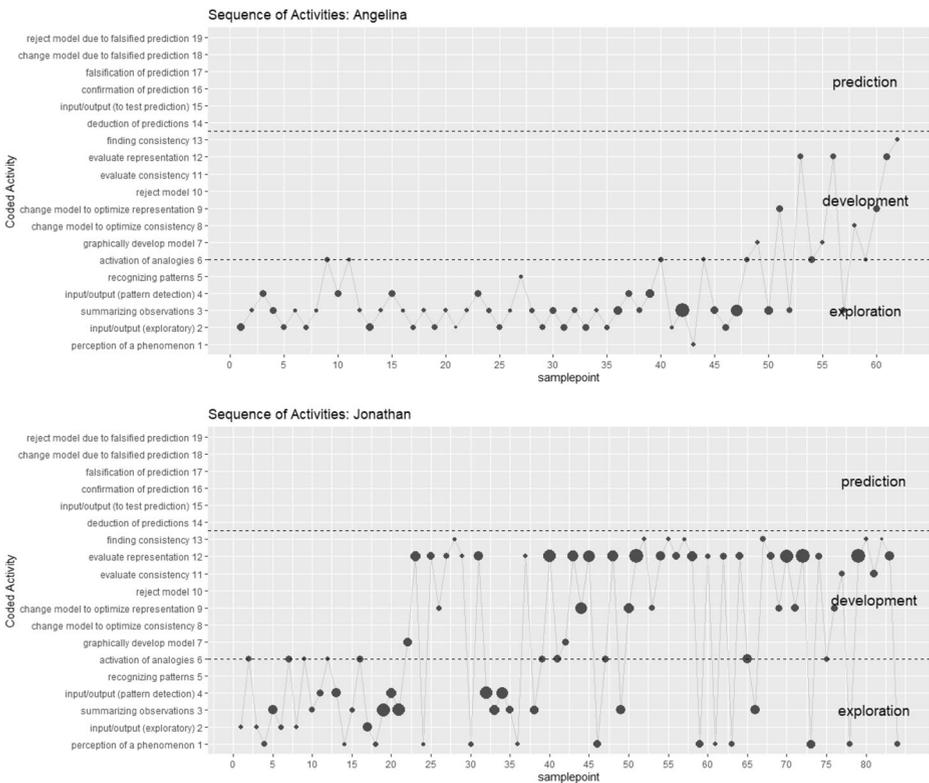


Fig. 2 Codelines of Angelina (top) and Jonathan (bottom). The conducted activities are shown in chronological order, whereby the size of each data point corresponds to the time spent on each activity. The dotted lines indicate the assignment of the activities to the three phases of exploration, model development, and prediction. While both participants engage in activities of exploration and model development, Angelina mostly focuses on activities of exploration, with activities of model development at the end. In contrast, Jonathan frequently switches between activities of exploration and model development. However, both Angelina and Jonathan conduct no activities of prediction

participants was visualized as codelines (see, e.g., Fig. 2), which is common practice in science education research (e.g., Louca and Zacharia 2015).

Cycle 1, Step 2: Grouping the Cases and Analysis of Empirical Regularities Based on the Occurrence of Activities

Based on the visual assessment of these codelines, three groups of participants can be easily distinguished:

- (1) One single participant conducted only activities of exploration and did not draw a model at all.
- (2) Seventeen participants conducted activities of exploration and model development but did not derive any predictions from their models (cf. Fig. 2).
- (3) Fourteen participants conducted activities of exploration, model development, and prediction.

As the affiliation of the participants to their group can be directly identified from the occurrence of activities, no further interpretation was needed, suggesting a sufficient degree of between-group heterogeneity. Comparing the sequences of activities of participants in the same group, however, suggests rather low within-group homogeneity, as the participants show major differences in the proportion of activities conducted as well as their sequential order. For example, Angelina shows a long sequence of activities of exploration, developing a model in the very end (Fig. 2 top), while Jonathan shows activities of model development throughout, alternating with activities of exploration (Fig. 2 bottom). Consequently, a further differentiation of the groups seemed to be needed, and no further analysis of meaningful relationships and type construction based on the occurrence of activities was conducted (cf. Fig. 1; step 3).

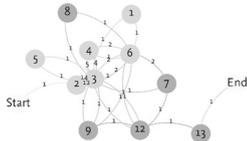
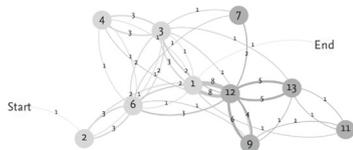
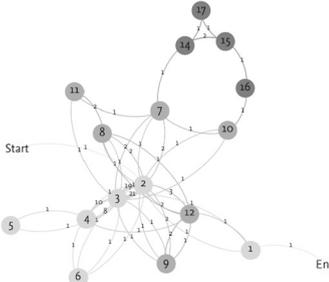
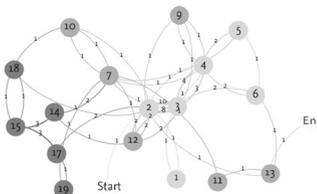
Cycle 2, Step 1: Transitions Between Activities as a Relevant Analysis Dimension

Because a quantitative analysis such as clustering would only account for the amount and time the activities were conducted, the transitions between single modeling activities were selected as relevant attributes to take the sequential order of activities into account in a second cycle of developing the typology (cf. Sins et al. 2005). To visualize the transitions, the codelines depicting the sequence of activities performed by every participant were transformed into state transition graphs (Andrienko and Andrienko 2018). These state transition graphs show all nineteen identified activities as knots that are connected by directed edges of different weights accounting for the quantity of transitions between these knots. Additionally, two knots for “start” and “end” are appended. All state transition graphs are automatically laid out in Gephi (Bastian et al. 2009) using the force-directed graph layout algorithm ForceAtlas2 (Jacomy et al. 2014).

Cycle 2, Step 2: Grouping the Cases and Analysis of Empirical Regularities Based on the Transitions Between Activities

The resulting graphs were then grouped by two raters in a first visual analysis. Six groups could be identified (Table 4), which further differentiate the three initial groups from step 1. Participants engaged in activities of modeling by the following:

Table 4 Examples of state transition graphs for each identified group

Modeling strategy	Example	Pseudonym
(1) exploration only		Carlo
(2a) exploration and development with a focus on development		Kara
(2b) exploration and development with a focus on exploration		Angelina
(2c) exploration and development, balanced		Jonathan
(3a) exploration, development and drawing predictions from a model once		Ben
(3b) exploration, development and drawing predictions from a model repeatedly		Raphael

The numbers in the circles illustrate the respective activities (Table 2), and the numbers on the lines show how often the activities were done sequentially. Activities of exploration, model development, or prediction are highlighted in different shades of gray

(1) Exploration only ($n = 1$)

(2a) Exploration and development with a focus on development ($n = 1$)

(2b) Exploration and development with a focus on exploration ($n = 11$)

- (2c) Exploration and development, balanced ($n = 5$)
- (3a) Exploration, development, and drawing predictions from a model once ($n = 9$)
- (3b) Exploration, development, and repeatedly drawing predictions from a model ($n = 5$)

Discussion of the results between both raters showed an agreement in the grouping of 23 out of 27 cases (85%), suggesting high between-group heterogeneity. Moreover, in the discussion of the grouping process, both raters considered size, complexity (number of knots and edges), the homogeneity of the edges, and edge weights in their analysis of the state transition graphs. Because complexity is mostly represented by the occurrence of the different activities, which was the attribute of analysis in cycle 1, homogeneity turned out to be the refining factor in cycle 2, which further differentiated the typology, improving within-group homogeneity.

Cycle 2, Step 3: Analysis of Meaningful Relationships and Type Construction Based on the Transitions Between Activities

To evaluate the visual assessment, various quantifiable graph metrics were considered (cf. Andrienko and Andrienko 2018), representing the attributes complexity and homogeneity. State transition graphs with lower complexity show a higher number of communities as activities not conducted by the participant and, therefore, unconnected to the rest of the graph, are counted as one community each (Bastian et al. 2009). To represent complexity, the number of communities was inverted by subtraction with 14 (maximum number of communities—1). To quantify the homogeneity, the reciprocal centrality of the graph was chosen based on the sum of three centrality measures (closeness centrality, degree centrality, and betweenness centrality) to account for structural outliers (Ronqui and Travieso 2015). The mean values of each group in terms of complexity and homogeneity were consistent with the visual assessment (Table 5).

Participants in groups (3a) and (3b) show higher scores for complexity compared with the other groups, as they conduct a broader range of different activities because they use their models to predict the black box behavior. Participants in groups (2c) and (3b) show higher scores for homogeneity compared with groups (2a), (2b), and (3a), as they tend to conduct the different activities more equally. The single participants in groups (1) and (2a) show a complexity score of 0, as they carry out only a few different activities, resulting in the highest number of distinct communities, which were used to offset all calculations. Together with the

Table 5 Mean values of the calculated complexity and homogeneity metrics for each group

Group	<i>n</i>	Mean complexity	Mean homogeneity
(1) Exploration only	1	0.0	0.81
(2a) Exploration and development with a focus on development	1	0.0	1.39
(2b) Exploration and development with a focus on exploration	11	2.9	1.37
(2c) Exploration and development, balanced	5	4.2	2.08
(3a) Exploration, development, and drawing predictions from a model once	9	7.0	1.25
(3b) Exploration, development, and repeatedly drawing predictions from a model	5	8.4	1.67

participants' individual scores, defined areas can be observed that support the visual grouping and that are in line with expectations (Fig. 3).

However, two participants (Sabrina and James) showed a statistically higher homogeneity in comparison to their assigned groups and were therefore investigated further (cf. Fig. 1, step 3). Both participants invested less time in the investigation of the black box compared with their respective groups: while Sabrina and James engaged in the black box activity for about half an hour, the other participants in groups (3a) and (3b) invested an hour more on average. The shorter time of investigation then leads to less repetitive actions and a higher homogeneity score than for the other participants.

Discussion and Conclusion

Modeling is seen as a central practice of scientific reasoning (e.g., Clement 2000; Gilbert 2004) and is highlighted in curricular documents as a core scientific practice (e.g., NGSS Lead States 2013). However, research on scientific modeling widely lacks detailed investigations of modeling processes (Nicolau and Constantinou 2014). Therefore, this study aimed to contribute to filling this gap in science education research by developing a typology of preservice teachers' modeling strategies. This typology has the potential to inform and guide further research as well as the development of interventions aiming to foster competencies in scientific modeling. The proposed typology is based on the homogeneity and complexity of modeling processes (Fig. 3) and includes the following strategies: (1) exploration only; (2a) exploration and development with a focus on development; (2b) exploration and development with a focus on exploration; (2c) exploration and development, balanced; (3a) exploration, development,

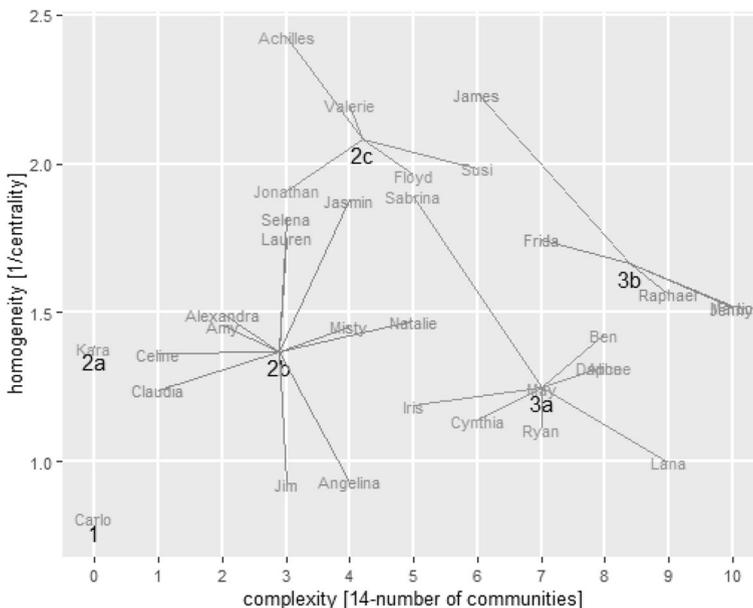


Fig. 3 Star plot illustrating the calculated metrics of each participant's state transition graph in conjunction with the mean values of his or her respective group (Table 4)

and drawing predictions from a model once; and (3b) exploration, development, and drawing predictions from a model repeatedly (Table 4).

Comparing the developed types of modeling strategies with strategies identified in previous studies of scientific reasoning processes, many parallels can be drawn, indicating that at least the strategies of modeling and experimental evaluation are quite similar: Participants of type (2b), who conduct activities of exploration and model development with a focus on exploration, show similarities to the participants of Schauble et al. (1991a), who utilize an engineering model of scientific investigation when conducting experiments. The participants carry out an investigation; in this case, the exploration of the black box, until a specific effect, for example, a repeating pattern, is reached. Then, the participants end the modeling process by following the initial task, drawing their model of the black box. In contrast, participants of type (2a) and (2c) perform activities of model development more frequently, expressing behavior similar to that previously described as model fitting (Sins et al. 2005). Moreover, in contrast to previous empirical studies suggesting that understanding the predictive use of models is challenging for most sample groups (Krell and Krüger 2016; Passmore et al. 2014), half of the cases we observed could be assigned to type (3a) or (3b) and included activities of prediction in their modeling processes. However, even participants of type (3b), who drew predictions from their model repeatedly, ended their modeling process when they found a possible solution. Participants did not further evaluate the developed models by trying to falsify the model or by systematically testing alternative hypotheses, although this is suggested in other studies as being important for scientific modeling (Grosslight et al. 1991; Louca and Zacharia 2015).

Methodologically, the identification of single activities of scientific reasoning, in this case modeling, built the foundation for a further analysis of scientific reasoning processes. Based on the occurrence of activities and the transitions between them, a typology of modeling strategies was successfully derived. The occurrence of activities alone was not regarded as a sufficient attribute, since the resulting groups did not show a sufficient degree of within-group homogeneity. By additionally considering the order of the sequence of activities in the form of their transitions, the typology was further refined, improving within-group homogeneity while maintaining between-group heterogeneity (Bailey 1994). Visualizing the modeling processes of each participant in different forms that emphasized the selected attributes, that is, codelines and state transitions graphs, allowed a fast, visual grouping process, helped reduce the complexity of the rich datasets, and improved their comparability. Grouping by multiple raters enabled the evaluation of interrater agreement and a reflection on the grouping process and the decisions involved. This led to the identification of more abstract attributes, such as the complexity and homogeneity of state transition graphs, that could be further quantified and that supported the visual assessment. Multiple cycles of typology development were necessary that were, in the present study, performed iteratively following the model of empirically grounded type construction (Kluge 2000). This model proved helpful in evaluating and refining the typology, as well as in avoiding errors in the assignment of a specific type to an individual case, as the model supported the systematic selection and evaluation of attributes and the resulting groups. We further argue that explicitly describing the different steps of empirically grounded type construction greatly improves transparency, as the selection of attributes and grouping processes have rarely been defined in other studies.

Regarding practical implications, we are aware that the black box setting and the sample size may limit the generalizability of the results. Also, participants' mental models (cf. Johnson-Laird 1983) were not considered in the analysis, as the operationalization of the present study focused on externalized models. Further research is needed that adopts the described methods to other settings and contexts to replicate the proposed typology.

Consequently, at this point, the integration of these findings into teaching environments and the development of interventions that foster competencies in scientific modeling still seem beyond reach. However, with the present findings extending the category system of Krell et al. (2019), preservice science teachers' individual modeling processes can be rapidly assigned one of the proposed types, highlighting the differences between preservice science teachers and reducing the complexity of the analysis. Identifying the type of modeling strategy, a preservice science teacher uses in a specific setting then might yield potential points of leverage on an individual level and could be used to guide this preservice science teacher to understand and use models as epistemic tools for scientific reasoning. Therefore, the proposed typology of modeling strategies might guide further research and theory development (cf. Klahr and Dunbar 1988) and potentially contribute to the development of specific educational interventions (cf. Schauble et al. 1991a, b; Klahr et al. 1993; Klahr and Nigam 2004). One example would be a learning progression similar to Schauble et al. (1991a, b), in which the proposed typology could be used as levels of the learning progression, which are encouraged through modeling tasks with increasing degrees of complexity. Herein, the black box setting might be considered rather complex, since participants are less likely to apply their background knowledge and the black box is nontransparent and dynamic (Betsch et al. 2011).

Outlook

Our sample will be enlarged, and a possible improvement of within-group homogeneity and between-group heterogeneity will be further investigated. Also, aiming to automate the type assignment process, other statistical measures for calculating differences between sequences will be considered and possibly integrated into the typology as additional dimensions, including z -scores (cf. Sins et al. 2005) and Yule's Q (cf. Lämsä et al. 2020). To gain an understanding of how successful a specific type of modeling strategy is, we plan to investigate the relationship between a participant's type of modeling strategy and the quality of the participant's modeling product (i.e., the final model or models throughout the process), as specific strategies are assumed to be used by less successful participants (Schauble et al. 1991a; Sins et al. 2005). Finally, the constructed types will be characterized, including all assessed background variables and the quality of the developed models. We hope that this will help to develop an understanding of how preservice science teachers' model and which factors impact their modeling and, by doing so, provide valuable information for those who are aiming to develop preservice competencies of scientific modeling.

Funding information Open Access funding provided by Projekt DEAL. We thank Jonna Kirchhoff and Maïke Barnebeck who contributed significantly to the qualitative part of the data analysis, and the German Research Foundation (DFG) for funding the project TypMoL (project No. 327507949). Any opinions, findings, conclusions, or recommendations expressed in this article are those of the authors and do not necessarily reflect the views of the DFG.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Acher, A., Arcá, M., & Sanmartí, N. (2007). Modeling as a teaching learning process for understanding materials: A case study in primary education. *Science Education*, 91(3), 398–418.
- Aizpurua, A., Lizaso, I., & Iturbe, I. (2018). Learning strategies and reasoning skills of university students. *Revista de Psicodidáctica (English ed.)*, 23(2), 110–116. <https://doi.org/10.1016/j.psicoe.2018.02.002> .
- Andrienko, N., & Andrienko, G. (2018). State transition graphs for semantic analysis of movement behaviours. *Information Visualization*, 17(1), 41–65. <https://doi.org/10.1177/1473871617692841> .
- Bailer-Jones, D. (2002). Scientists' thoughts on scientific models. *Perspectives on Science*, 10(3), 275–301. <https://doi.org/10.1162/106361402321899069> .
- Bailey, K. (1994). *Typologies and taxonomies: An introduction to classification techniques* (Vol. 102). Thousand Oaks, CA: Sage.
- Bakeman, R., & Gottman, J. M. (1997). *Observing interaction: An introduction to sequential analysis*. Cambridge, NY: Cambridge University Press.
- Bastian M, Heymann S, Jacomy M (2009). Gephi: An open source software for exploring and manipulating networks. *International AAAI Conference on Weblogs and Social Media*, 8(2009), 361–362.
- Betsch, T., Funke, J., & Plessner, H. (2011). *Denken–Urteilen, Entscheiden, Problemlösen*. Berlin/Heidelberg: Springer Nature.
- British Columbia Ministry of Education (BCMOE). (2019). Curriculum redesign. Retrieved from <https://curriculum.gov.bc.ca/rethinking-curriculum>
- Brown, D., & Clement, J. (1989). Overcoming misconceptions via analogical reasoning: Abstract transfer versus explanatory model construction. *Instructional Science*, 18(4), 237–261. <https://doi.org/10.1007/bf00118013>
- Čančula, M. P., Planinšič, G., & Etkina, E. (2015). Analyzing patterns in experts' approaches to solving experimental problems. *American Journal of Physics*, 83(4), 366–374. <https://doi.org/10.1119/1.4913528> .
- Chinn, C. A., & Malhotra, B. A. (2002). Epistemologically authentic inquiry in schools: A theoretical framework for evaluating inquiry tasks. *Science Education*, 86(2), 175–218. <https://doi.org/10.1002/sce.10001> .
- Clement, J. (2000). Model based learning as a key research area for science education. *International Journal of Science Education*, 22(9), 1041–1053.
- Clement, J. (2008). *Creative model construction in scientists and students: The role of imagery, analogy, and mental simulation*. Springer Science & Business Media.
- Clement, J., & Rea-Ramirez, M. A. (2008). *Model based learning and instruction in science (model based learning and instruction in science)*. Dordrecht: Springer.
- 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. N. (1999). *Science without laws*. Chicago, IL: The University of Chicago Press.
- Giere, R. N., Bickle, J., & Mauldin, R. (2006). *Understanding scientific reasoning* (5th ed.). Belmont: Thomson Wadsworth.
- Gilbert, J. (2004). Models and modelling: Routes to more authentic science education. *International Journal of Science and Mathematics Education*, 2(2), 115–130. <https://doi.org/10.1007/s10763-004-3186-4> .
- Gilbert, J., & Osborne, R. J. (1980). The use of models in science and science teaching. *European Journal of Science Education*, 2(13), 3–13. <https://doi.org/10.1080/0140528800020103> .
- Grosslight, L., Unger, C., Jay, E., & Smith, C. L. (1991). Understanding models and their use in science: Conceptions of middle and high school students and experts. *Journal of Research in Science Teaching*, 28(9), 799–822.
- Hammann, M., Phan, T. H., & Bayrhuber, H. (2008). Experimentieren als Problemlösen: Lässt sich das SDDS-Modell nutzen, um unterschiedliche Dimensionen beim Experimentieren zu messen? [experimentation as problem-solving: Can the SDDS model be used to measure different dimensions of experimentation?]. In M. Prenzel, I. Gogolin, & H.-H. Krüger (Eds.), *Kompetenzdiagnostik: Zeitschrift für Erziehungswissenschaft* (pp. 33–49). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Hartmann, S., Upmeyer zu Belzen, A., Krüger, D., & Pant, H. A. (2015). Scientific reasoning in higher education: Constructing and evaluating the criterion-related validity of an assessment of preservice science teachers' competencies. *Zeitschrift Für Psychologie*, 223(1), 47–53. <https://doi.org/10.1027/2151-2604/a000199> .
- Harwood, W. S. (2004). A new model for inquiry: Is the scientific method dead? *Journal of College Science Teaching*, 33(7), 29–33.
- Hodson, D. (2014). Learning science, learning about science, doing science: Different goals demand different learning methods. *International Journal of Science Education*, 36, 2534–2553. <https://doi.org/10.1080/09500693.2014.899722> .

- Jacomy, M., Venturini, T., Heymann, S., & Bastian, M. (2014). ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PLoS One*, *9*(6), e98679.
- Johnson-Laird, P. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge: Cambridge U.P.
- Justi, R., & Gilbert, J. K. (2003). Teachers' views on the nature of models. *International Journal of Science Education*, *25*, 1369–1386.
- Khan, S., & Krell, M. (2019). Scientific reasoning competencies: A case of preservice teacher education. *Canadian Journal of Science, Mathematics and Technology Education*, (19), 446–464. Retrieved from <https://link.springer.com/article/10.1007/s42330-019-00063-9>.
- Kind, P., & Osborne, J. (2017). Styles of scientific reasoning: A cultural rationale for science education? *Science Education*, *101*(1), 8–31. <https://doi.org/10.1002/sce.21251>.
- Klahr, D. (2002). *Exploring science: The cognition and development of discovery processes*. Cambridge, MA: MIT press.
- Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science*, *12*(1), 1–48.
- Klahr, D., & Nigam, M. (2004). The equivalence of learning paths in early science instruction: Effects of direct instruction and discovery learning. *Psychological Science*, *15*(10), 661–667. <https://doi.org/10.1111/j.0956-7976.2004.00737.x>.
- Klahr, D., Fay, A. L., & Dunbar, K. (1993). Heuristics for scientific experimentation: A developmental study. *Cognitive Psychology*, *25*(1), 111–146.
- Kluge, S. (2000). Empirically grounded construction of types and typologies in qualitative social research. *Forum, Qualitative Social Research / Forum, Qualitative Sozialforschung*, *1*(1). <http://nbn-resolving.de/urn:nbn:de:0114-fqs0001145>.
- KMK. (2005). *Bildungsstandards im Fach Biologie für den Mittleren Schulabschluss [Biology education standards for the Mittlere Schulabschluss]* (5th ed.). München & Neuwied: Wolters Kluwer.
- KMK. (2019). *Ländergemeinsame inhaltliche Anforderungen für die Fachwissenschaften und Fachdidaktiken in der Lehrerbildung [Common federal requirements regarding scientific disciplines and science education in teacher preparation programs]*. Berlin: Sekretariat der Kultusministerkonferenz
- Kosłowski, B. (2013). Scientific reasoning: Explanation, confirmation bias, and scientific practice. In G. J. Feist & M. E. Gorman (Eds.), *Handbook of the psychology of science* (pp. 151–192). New York, NY: Springer.
- Krell, M. (2018). Schwierigkeitserzeugende Aufgabenmerkmale bei Multiple-Choice-Aufgaben zur Experimentierkompetenz im Biologieunterricht: Eine Replikationsstudie. *Zeitschrift Für Didaktik Der Naturwissenschaften*, *24*. <https://doi.org/10.1007/s40573-017-0069-0>.
- 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*(2), 160–173. <https://doi.org/10.1080/00219266.2015.1028570>.
- Krell, M., Upmeyer zu Belzen, A., & Krüger, D. (2014a). Context-Specificities in Students' Understanding of Models and Modelling: E-Book Proceedings of the ESERA 2013 Conference: Science Education Research For Evidence-based Teaching and Coherence in Learning. Nicosia, Cyprus: *European Science Education Research Association*.
- Krell, M., Upmeyer zu Belzen, A., & Krüger, D. (2014b). Students' levels of understanding models and modelling in biology: Global or aspect-dependent? *Research in Science Education*, *44*(1), 109–132. <https://doi.org/10.1007/s11165-013-9365-y>.
- 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>.
- Krell, M., Walzer, C., Hergert, S., & Krüger, D. (2019). Development and application of a category system to describe pre-service science teachers' activities in the process of scientific modelling. *Research in Science Education* *49*(5), 1319–1345. <https://doi.org/10.1007/s11165-017-9657-8>.
- Krüger, D., Hartmann, S., Nordmeier, V., & Upmeyer zu Belzen, A. (2020). Measuring scientific reasoning competencies. In O. Zlatkin-Troitschanskaia, H. Pant, M. Toepper, & C. Lautenbach (Eds.), *Student learning in German higher education* (pp. 261–280). Wiesbaden: Springer.
- Lämsä, J., Hämäläinen, R., Koskinen, P., Viiri, J., & Mannonen, J. (2020). The potential of temporal analysis: Combining log data and lag sequential analysis to investigate temporal differences between scaffolded and non-scaffolded group inquiry-based learning processes. *Computers & Education*, *143*, 103674. <https://doi.org/10.1016/j.compedu.2019.103674>.
- Lawson, A. E. (2004). The nature and development of scientific reasoning: A synthetic view. *International Journal of Science and Mathematics Education*, *2*(3), 307–338.
- Lederman, N. G., & Abd-El-Khalick, F. (2002). Avoiding de-natured science: Activities that promote understandings of the nature of science. In W. F. M. (Ed.), *The Nature of Science in Science Education* (Vol. 5). Dordrecht: Science & Technology Education Library: Springer.

- Lehrer, R., & Schauble, L. (2015). The development of scientific thinking. *Handbook of child psychology and developmental science*, 2(7), 671–714.
- Leighton, J., & Gierl, M. (2007). *Cognitive diagnostic assessment for education: Theory and applications*. Cambridge: Cambridge University Press.
- Liepmann, D., Beauducel, A., Brocke, B., & Amthauer, R. (2007). *Intelligenz-Struktur-Test 2000 R: IST 2000 R* (2nd ed.). Göttingen: Hogrefe.
- Louca, L. T., & Zacharia, Z. C. (2012). Modeling-based learning in science education: Cognitive, metacognitive, social, material and epistemological contributions. *Educational Review*, 64(4), 471–492. <https://doi.org/10.1080/00131911.2011.628748> .
- Louca, L. T., & Zacharia, Z. C. (2015). Examining learning through modeling in k-6 science education. *Journal of Science Education and Technology*, 24(2), 192–215. <https://doi.org/10.1007/s10956-014-9533-5> .
- Mathesius, S., Hartmann, S., Upmeier zu Belzen, A., & Krüger, D. (2016). Scientific reasoning as an aspect of pre-service biology teacher education: Assessing competencies using a paper-pencil test. In T. Tal & A. Yarden (Eds.), *The future of biology education research* (pp. 93–110). Israel: Haifa.
- Morris, B. J., Croker, S., MASNICK, A. M., & Zimmerman, C. (2012). The emergence of scientific reasoning. In H. Kloos, B. J. Morris, & B. J. Amaral (Eds.), *Current topics in children's learning and cognition*. IntechOpen. <https://doi.org/10.5772/53885>.
- Nersessian, N. (1992). How do scientists think? Capturing the dynamics of conceptual change in science. In R. N. Giere & H. Feigl (Eds.), *Cognitive Models of Science* (Vol. 15, pp. 3–45). Minneapolis, MN: University of Minnesota Press.
- Nersessian, N. (2002). The cognitive basis of model-based reasoning in science. In P. Carruthers, S. Stich, & M. Siegal (Eds.), *The cognitive basis of science* (pp. 133–153). New York: Cambridge University Press.
- Neumann, K., Hätig, H., Harms, U., & Parchmann, I. (2017). Science teacher preparation in Germany. In J. E. Pedersen, T. Isozaki, & Hirano T. (Eds.), *Model science teacher preparation programs: An international comparison of what works*. Charlotte, NC: Information Age.
- Neumann, K., Schecker, H., & Theyßens, H. (2019). Assessing complex patterns of student resources and behavior in the large scale. *The Annals of the American Academy of Political and Social Science*, 683(1), 233–249. <https://doi.org/10.1177/0002716219844963> .
- NGSS Lead States. (2013). *Next generation science standards: For States, by States*. Washington, DC: The National Academies Press.
- Nicolau, C. T., & Constantinou, C. P. (2014). Assessment of the modeling competence: A systematic review and synthesis of empirical research. *Educational Research Review*, 13, 52–73. <https://doi.org/10.1016/j.edurev.2014.10.001> .
- Osborne, J. (2013). The 21st century challenge for science education: Assessing scientific reasoning. *Thinking Skills and Creativity*, 10, 265–279.
- Passmore, C. M., & Svoboda, J. (2012). Exploring opportunities for argumentation in modeling classrooms. *International Journal of Science Education*, 34(10), 1535–1554.
- Passmore, C., Stewart, J., & Cartier, J. (2009). Model-based inquiry and school science: Creating connections. *School Science and Mathematics*, 109(7), 394–402.
- 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.
- Rasmussen, J. (1981). Models of mental strategies in process plant diagnosis. In J. Rasmussen & W. B. R (Eds.), *Human Detection and Diagnosis of System Failures, NATO conference series* (Vol. 15, pp. 241–258). Boston: Springer.
- Reiff, R., Harwood, W. S., & Phillipson, T. (2002) A scientific method based upon research scientists' conceptions of scientific inquiry. In *Proceedings of the Annual International Conference of the Association for the Education of Teachers in Science*, Charlotte, NC: Association for the Education of Teachers in Science.
- 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> .
- Ronqui, J. R. F., & Travieso, G. (2015). Analyzing complex networks through correlations in centrality measurements. *Journal of Statistical Mechanics: Theory and Experiment*, 2015(5), P05030.
- Ruppert, J., Duncan, R. G., & Chinn, C. A. (2019). 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>.
- Schauble, L., Glaser, R., Raghavan, K., & Reiner, M. (1991a). Causal models and experimentation strategies in scientific reasoning. *Journal of the Learning Sciences*, 1(2), 201–238. https://doi.org/10.1207/s15327809jls0102_3 .

- Schauble, L., Klopfer, L. E., & Raghavan, K. (1991b). Students' transition from an engineering model to a science model of experimentation. *Journal of Research in Science Teaching*, 28(9), 859–882.
- Schreier, M. (2012). *Qualitative content analysis in practice*. London: Sage Publications.
- Schwarz, C. V., Passmore, C., & Reiser, B. J. (Eds.). (2017). *Helping students make sense of the world using next generation science and engineering practices*. Arlington: NSTA Press.
- Sins, P. H., Savelsbergh, E. R., & van Joolingen, W. R. (2005). The difficult process of scientific modelling: An analysis of novices' reasoning during computer-based modelling. *International Journal of Science Education*, 27(14), 1695–1721.
- van Driel, J. H., & Verloop, N. (1999). Teachers' knowledge of models and modelling in science. *International Journal of Science Education*, 21, 1141–1153. <https://doi.org/10.1080/095006999290110> .
- 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>.
- Windschitl, M., Thompson, J., & Braaten, M. (2008). How novice science teachers appropriate epistemic discourses around model-based inquiry for use in classrooms. *Cognition and Instruction*, 26(3), 310–378. <https://doi.org/10.1080/07370000802177193> .
- Zimmerman, C. (2005). The development of scientific reasoning skills. *Developmental Review*, 20(1), 99–149.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.