Self-evaluative Scientific Modeling in an Outreach Gene Technology Laboratory

Tamara Roth[1] · Franz-Josef Scharfenberg[1] · Julia Mierdel[1] · Franz X. Bogner[1]

Abstract

The integration of scientific modeling into science teaching is key to the development of students' understanding of complex scientific phenomena, such as genetics. With this in mind, we conducted an introductory hands-on module during an outreach gene technology laboratory on the structure of DNA. Our module examined the influence of two model evaluation variants on cognitive achievement: Evaluation 1, based on students' hand-drawn sketches of DNA models and two open questions, and evaluation 2, based on students' own evaluations of their models in comparison to a commercially available DNA model. We subsequently subdivided our sample ($N = 296$) into modellers-1 ($n = 151$) and modellers-2 ($n = 145$). Analyses of cognitive achievement revealed that modellers-2 achieved higher scores than modellers-1. In both cases, low achievers, in particular, benefitted from participation. Assessment of modellers-2 self-evaluation sheets revealed differences between self-evaluation and independent reassessment, as non-existent model features were tagged as correct whereas existent features were not identified. Correlation analyses between the models' assessment scores and cognitive achievement revealed small-to-medium correlations. Consequently, our evaluation-2 phase impacted students' performance in overall and model-related cognitive achievement, attesting to the value of our module as a means to integrate real scientific practices into science teaching. Although it may increase the workload for science teachers, we find that the potential scientific modeling holds as an inquiry-based learning strategy is worth the effort.

Keywords Cognitive knowledge · Scientific models and modeling · Gene technology outreach learning · Science education

Introduction

Hands-on experiments in the classroom are key for conveying knowledge of complex natural phenomena that might otherwise remain inaccessible to students (Kind 2015). Yet, for many years, a clear focus on instructional and theoretical teaching has dominated the educational landscape. Today, new national educational standards (KMK 2005) replace these long-established methods and raise hope of more practical approaches to science teaching. However, one must bear in mind that some schools may lack the financial means to build and stock laboratories (e.g., Raviv et al. 2019) which may require expensive and specialist equipment.

✉ Tamara Roth
  tamara.roth@uni-bayreuth.de

[1] Centre of Math & Science Education, Department of Biology Education, University of Bayreuth, Universitätsstr. 30 NW I, 95447 Bayreuth, Germany

For schools without the necessary means, outreach laboratories (e.g., at universities) offer opportunities for students to experience science in authentic settings. Many such laboratories have been developed over the last two decades and offer opportunities to foster students' interest and motivation (Glowinski and Bayrhuber 2011). Hereby, solution-oriented approaches to scientific problems and an adequate learning environment are important (e.g., Nasir et al. 2006, Brody et al. 2007). Our outreach program, *Simply inGEN(E)ious! DNA as a carrier of genetic information*, meets these requirements: Designed as a 1-day-long module, it involves hands-on experiments, prediction of potential results, and a modeling phase (Mierdel and Bogner 2019a).

In our previous study, we compared two modeling variants: model viewers and modelers. Model viewers worked "with a commercially available school model of DNA structure" while modelers "were required to generate a DNA model using assorted handcrafting materials" (Mierdel and Bogner 2019a, p. 1). Most of the existing literature suggests that hands-on modeling contributes to cognitive achievement (e.g., Jackson et al. 2008, Passmore et al. 2009), yet we found that model viewers had a higher mid-term increase in

knowledge. Thus, "we dispensed a more detailed model evaluation for modellers" (Mierdel and Bogner 2019a, p. 12). To fill this gap in our research, we selected two variants related to students' model evaluation: modellers-1, who only participated in one evaluation phase, and modellers-2, who additionally completed a second evaluation phase. We based the latter on our model-viewing approach.

Our study aims to observe the potential influences of both model evaluation variants on cognitive achievement. We first outline the relevant theoretical background, including knowledge of learning and model learning which influenced the way we developed, conducted, and evaluated our intervention. We then provide a brief explanation of the phases of our intervention before we present our research questions.

Theoretical Background

Learning

Despite numerous suggestions, a universally accepted definition of "learning" has not yet been developed (e.g., Domjan 2015, Haselgrove 2016). Learning is often summarized as "a change in behavior [depending on] experience" (De Houwer et al., 2011, p. 631, Thorpe 1943), a view which is supported by constructivists (Hodson 1998; Lachman 1997). Thereby, students autonomously verify or falsify hypotheses and modify preconceptions accordingly or develop new ideas (Schwarz et al. 2009). These competencies also reflect (Kolb's 1984, Abdulwahed and Nagy 2013) categorization of learners' abilities: concrete experience ability, reflective observation ability, abstract conceptualization ability, and active experimentation ability. Hence, the construction of knowledge from chunks of information requires experiential processes to transform such information into a coherent mental model.

Real experiments in class or outreach settings provide the necessary basis for students to build up knowledge and allow them to manage scientific equipment. Such first-hand experience is vital to understanding the nature of the physical world and science, where experimental outcomes are not always as expected (Loveys and Riggs 2019; Olympiou and Zacharia 2012). Solomon (1980, p.13) even described hands-on experimenting as key to increasing scientific knowledge: "Science teaching must take place in the laboratory; science simply belongs there as naturally as cooking belongs in the kitchen and gardening in the garden." Different levels of scientific knowledge can, thus, be achieved by combining hands-on experimentation with the development of scientific thinking skills. This approach significantly enhances skills set by the national standards (KMK 2005), such as the understanding of scientific processes. Together with an open and flexible learning environment, critical thinking skills and self-

evaluation may be fostered. This is especially important when it comes to reassessing results. An effective method for encouraging self-evaluation is reflective writing (Kovanović et al. 2018) about particular experimental phases. This approach can encourage students to rethink and reassess certain steps in the experimental process (Scharfenberg and Bogner 2013; for details, see below). In this "self-directive process," students "transform their mental abilities into academic skills" (Zimmerman, 2002, p. 65).

Model Learning

Models are tools commonly used to visualize and explain phenomena (Krajcik and Merritt 2012). Windschitl and Thompson (2006, p. 796), for instance, described models as "hypothesized relationships among objects, processes, and events". Thereby, models can reveal underlying mechanisms, show causal links, raise questions, and test multiple hypotheses. Whenever a prediction proves incorrect or new evidence emerges, a model can be adapted and refined (e.g., Passmore et al. 2009, Carpenter et al. 2018). Models are regarded as the cornerstones of every scientific discipline, combining theory with modeled processes and functions to explain how things really are (Ates and Erylmaz 2011). Thus, modeling in science classes is invaluable for student learning (e.g., Jackson et al. 2008, Passmore et al. 2009). Although natural phenomena are often difficult to observe, models can provide an authentic alternative experience, are often easier to understand, and do not require exhaustive preparatory work (e.g., Windschitl and Thompson 2006, Carpenter et al. 2018). Students, interalia, gained better cognitive achievement scores when applying three-dimensional DNA models rather than two-dimensional DNA models (e.g., Rotbain et al. 2006, Saka et al. 2006).

Modeling may lead to more productive student activity in the classroom, emphasizing vital scientific practices, including the need to "engage in inquiry other than controlled experiments, [to] use existing models in their inquiries, [to] engage in inquiry that leads to revised models, [to] use models to construct explanations, [to] use models to 'unify' their understanding, and [to] engage in argumentation" (Passmore et al. 2009, p. 397). We decided to include these activities in our three-dimensional modeling approach (for details, see below).

Intervention Phases

We divided our 1-day module into four different intervention phases (Table 1).

Table 1 Quasi-experimental study design

Intervention phases	Evaluation variants	
	Modellers-1 ($n = 151$)	Modellers-2 ($n = 145$)
Pre-lab	+	+
DNA-related theoretical and experimental phases		
DNA relevance	+	+
Hands-on isolation of DNA	+	+
Gel electrophoresis of DNA	+	+
Model-related phases		
Mentally modeling: text analysis	+	+
DNA modeling with craft materials	+	+
Model evaluation-1: drawing a paper and pencil version of the crafted model and answering questions	+	+
Model evaluation-2: comparing the crafted model with a scientific demonstration model	–	+
Interpretation	+	+

Pre-lab Phase

As many students are not familiar with laboratory equipment or scientific work, we provide an appropriate pre-lab phase. This phase addresses three scientific aspects, in particular: affective dimensions, introduction to laboratory techniques, and introduction to theoretical concepts (e.g., Sarmouk et al. 2019). Students, thus, learn how to handle the equipment necessary for experimentation, such as micropipettes and centrifuges. Teachers act as external guides and introduce theoretical concepts via presentations and demonstrations, which are key to understanding the subsequent experiments. This phase should also prevent students from feeling overwhelmed in response to the experimental situation (e.g. Kalyuga 2009).

DNA-Related Theoretical and Experimental Phases

Theoretical Phases Having provided a real-life background to the subsequent experiments (DNA relevance; Table 1), the underlying scientific basics of DNA isolation and gel electrophoresis were introduced. Students were then encouraged to connect new information with prior knowledge and hypothesize about the experiments' potential outcomes (Mierdel and Bogner 2019a). Our decision to provide information in a series of theoretical phases ensures that students are not overwhelmed by too much new information at once and allows them to focus on the experimental phase ahead. This is particularly important when explaining the difficult concept of gel electrophoresis which not only involves chemical knowledge about the DNA's contents but also abstract thinking in order to imagine the processes involved.

Experimental Phases Ours is an evidence-based, two-step approach (Scharfenberg and Bogner 2013), in which students answer questions in their individual workbooks and think about subsequent experimental procedures. They work in pairs to discuss every step before carrying out experiments that effectively combine hands-on and minds-on activities and require students to do more than simply follow instructions (Scharfenberg and Bogner 2013).

Model-Related Phases

Both model-related phases directly followed the experimental, DNA-related phases. As we regard models as vital to visualizing and explaining phenomena in science and science education (Krajcik & Merritt 2012), we subdivided our model-related phases into a mental modeling phase involving text analysis, a modeling phase involving craft materials, a model evaluation-1 phase, and a model evaluation-2 phase. Only modellers-2 participated in all four phases.

As in our previous study, we based our model-related phases on the four main stages of the *Model of Modelling* (Justi and Gilbert 2002, p. 370 ff.): (1) "experience[s] of the phenomenon being modelled", (2) "forming a mental model", (3) "decision … about the mode of representation in which it is to be expressed", and (4) "testing … scope and limitations of the model". Mental modeling is, thereby, key to providing a theoretical basis for experimental findings. According to Franco and Colinvaux (2000), building mental models involves reasoning about previously obtained knowledge to make predictions and derive new ideas from it. A text about the discovery of the DNA's structure (Usher 2013), thus, provides fundamental knowledge about the necessary components of the DNA, which will be applied to the students'

simplified mental model (e.g., Mierdel and Bogner 2019b, Franco & Colinvaux 2000).

After determining the DNA's representation and building the handcrafted model, model evaluation-1 phase was organized as a reciprocal self-evaluation. As combining sketching and handcrafting models proved to be important (e.g., Prabha 2016, Orhan and Sahin 2018), modellers-1 and modellers-2 evaluated their handcrafted DNA model based on their earlier paper-and-pencil version. Another effective method for encouraging self-evaluation is reflective writing (Kovanović et al. 2018). Open-ended questions about model-related components encourage students to rethink and reassess certain steps and decisions in their development of the mental model into its physical counterpart (Mierdel and Bogner 2019a). Modellers-2 also assessed their handcrafted DNA models using a comparison-based self-evaluation with a commercially available DNA demonstration model. In other words, we based this second evaluation phase on our previous model-viewing approach.

Interpretation Phase

Here, students compared their hypothesis about the outcome of their experiments with the gel electrophoresis' images. They also discussed their individual models in class and compared these to the molecular DNA model by Watson and Crick which, in most cases, differed from the students' models (Mierdel and Bogner 2019b).

Objectives of the Study

The present study tries to answer the following research questions:

1 How does the application of one or two model evaluation phases influence overall cognitive achievement and, in particular, the model-related knowledge developed during the hands-on laboratory?
2 How does the additional evaluation-2 phase of modellers-2's three-dimensional models influence their overall learning?

Thus, we had three specific objectives:

- to assess students' overall cognitive achievement and the model-related knowledge of modellers-2, which we would compare with that of modellers-1
- to determine the quality of the evaluation-2 phase and the DNA component that modellers-2 correctly identified in their handcrafted models

- to examine the potential correlations between modellers-2's performance in the model evaluation-2 phase and modellers-2's cognitive achievement

Materials and Methods

After a brief introduction to our educational intervention, its design, and the independent variable, we explain the modeling phases and the model evaluation phases. We then describe the students' sample, discuss the dependent variables, and outline the statistical methods applied.

Educational Intervention, Design, and Independent Variable

The 1-day, hands-on module offered inquiry-based learning activities focused on the structure of DNA, aimed at ninth graders. Students worked in pairs to complete their tasks with guidance provided in a workbook (for a detailed module description, see Mierdel and Bogner 2019a). The content of the module is in line with the state's syllabus and follows the national competency requirements (KMK 2005).

We conducted two versions of the intervention which differed concerning students' evaluations of their models. Since often only quasi-experimental designs are feasible for students in intact class groups (Cook and Campell 1979), student classes were randomly assigned to each of the evaluation variants. The students were, thus, divided into modellers-1 and modellers-2, respectively, as the independent variable (Table 1).

Both versions of the intervention began with a pre-lab phase (50 min) wherein students were familiarized with the lab equipment and relevant working techniques. Thereafter, the criminological relevance of DNA was introduced to contextualize the two main DNA-related experimental phases: DNA isolation from oral mucosal cells (60 min) and agarose gel electrophoresis (85 min). Both were connected to model-related phases (60 min). Students then tried to retrace Watson and Crick's research to solve the molecular puzzle of the DNA's structure.
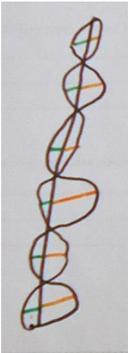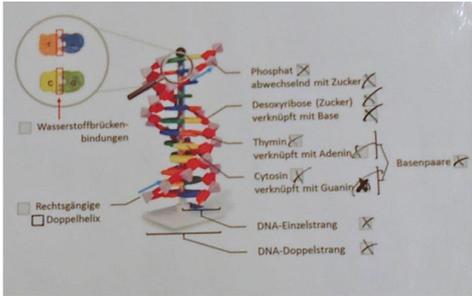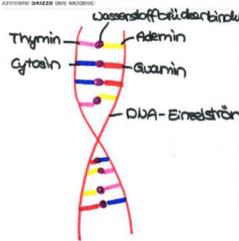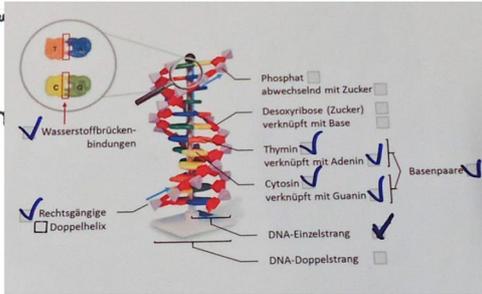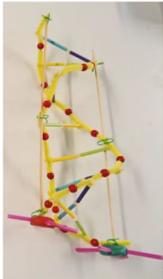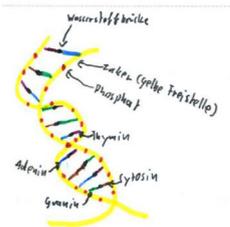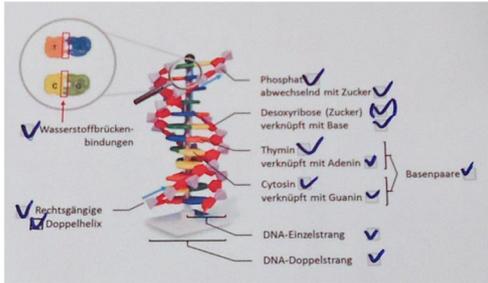
The Modeling Phases These phases (Table 1) were key in providing a theoretical basis for experimental findings. Having read about the discovery of the structure of DNA (Usher 2013), students discussed and answered questions in their workbooks (e.g., "DNA's backbone: Label its components and describe their set up"; for details, see Electronic Supplementary Material [ESM] 1 as an online resource). The aim was to enable the students to internalize key aspects of the text which would later allow them to mentally model the

DNA's structure. To this end, the text included references to all key components of the DNA (e.g., base pairing).

In the next stage, students transformed their mental DNA models into physical, handcrafted DNA models (Table 1) using a DNA-modeling kit containing crafting materials (e.g., colored beads and pipe cleaners; for three model examples; Table 2, 1st column).

The Model Evaluation Phases These phases (Table 2, 2nd and 3rd columns) were organized as a reciprocal self-evaluation of students' models. Modellers-1 and modellers-2 all evaluated their handcrafted DNA model against their paper-and-pencil sketch of the model (evaluation-1). Modellers-2 then conducted an additional, comparison-based assessment of their handcrafted DNA

Table 2 Assessment of modellers-2's evaluation phases

Phase			Assessment (model[a] / sketches[b] / self-evaluation sheet[c])
Modelling	Model evaluation		
Hand-crafted model	Evaluation-1: sketches	Evaluation-2: comparing and ticking the self-evaluation sheet	
			6 / 6 / 6
			10 / 11 / 7
			13 / 17 / 13

[a] Score maximal 14 points

[b] Score maximal 14 points

[c] Score maximal the model score; see [a]

models, in which they were asked to draw a paper-and-pencil model with labels to identify the model's components (Table 2, 2nd column) and the elements of their handcrafted models. This was an opportunity to reflect on the process of building the model and on its informative value. In the meantime, ideas about the accuracy of their models could be exchanged. While modeling, students might have become aware of differences between their own and their classmates' models. Once they had completed a sketch of their models, students answered two open-ended questions on their worksheets: "Which features of the original DNA molecule are simplified in your model" and "Explain why one might create different models of one biological original (in our case, the structure of the DNA)?" Thus, students were required to consider the scope and limitations of their models.

Modellers-2 also completed evaluation-2 (Table 1), wherein they were asked to compare and contrast their handmade models with a commercially available DNA model. A self-evaluation sheet (ESM 2) displaying the image of this scientific model served as the basis of their assessments. Each component included in both models was tagged to assist students' self-evaluations of their models.

In both versions, students' findings from the model phase were integrated into a final interpretation phase. Here, students discussed experimental results of the gel electrophoresis, which they compared with previous hypotheses (Table 1).

Participants

Altogether, 296 ninth-graders (higher secondary school) participated in our study (girls 52.0%, boys 48.0%; $M_{Class\ size} = 22.8$, $SD = 6.2$; $M_{Age} = 14.6$, $SD = 0.8$). Six classes took part as modellers-1 ($n = 151$) and seven classes as modellers-2 ($n = 145$). Modellers-1 teamed up in 77 groups (75 2-person groups and two students working individually due to illness), modellers-2 in 73 groups (72 2-person groups and one student working individually due to illness). To avoid bias, we compared students' prior knowledge of biology with the respective biology grades. We found no significant difference (Mann-Whitney U test [MWU]: $Z = -1.144$, $p = .253$). Moreover, we compared individual students' prior in-class experience of modeling (3-item scale, adapted from Authors 2007; Cronbach's Alpha .62) and did not discover a significant difference (MWU: $Z = -0.859$, $p = .390$).

Participation was voluntary. Written parental consent was given prior to students' participation in our study, although the data collection was pseudo-anonymous and students could not be identified. The study was designed in accordance with the Declaration of Helsinki (2013), and the state ministry approved the questionnaires used.

Dependent Variables

As dependent variables, we examined students' knowledge in a repeated measurement design: a pre-test (T0) 2 weeks before the intervention, a post-test (T1) after the module, and a retention test 6 weeks thereafter (T2). We examined students' sketches and their responses to the open questions from the evaluation-1 phase; for modellers-2, we additionally assessed the evaluation-2 phase. Throughout the entire intervention, students were unaware of any testing schedules.

Students' Knowledge We applied an ad hoc knowledge test comprising 30 multiple-choice items: 12 items (examples; Table 3) assessed knowledge of the DNA-related phases (DNA relevance, hands-on isolation, and gel electrophoresis of DNA; Table 1) and 18 items (Table 3) analyzed knowledge related to the model phases (Table 1).

Content validity was given as the items were consistent with the state syllabus. Regarding construct validity, inter-item correlations below 0.20 (T0 = 0.08; T1 = 0.19; T2 = 0.18) confirmed that each item referred to different knowledge facets. Furthermore, the items' heterogeneity concerning complex constructs, such as cognitive achievement, emphasizes the given construct validity (Rost 2004). Cronbach's alpha values of 0.71 (T0), 0.64 (T1), and 0.70 (T2) indicate acceptable internal consistency when the value exceeds 0.70. According to Lienert and Raatz (1998), values between 0.50 and 0.70 allow for the differentiation of groups. Item difficulties (percentage of correct answers, Bortz and Döring 1995) ranged between 7% (high difficulty) and 88% (low difficulty). In the case of our intervention, the item difficulties reversed from T0 to T1 (Fig. 1) and generally decreased.

We calculated the students' scores and analyzed these with regard to increases in knowledge (T1 minus T0) and retention rate (T2 minus T0). However, these different variables do not reflect actual knowledge growth. Thus, we calculated the actual learning success with respect to the maximal attainable score (30 correct answers): $(T1 - T0) \times (T1/30)$ and the persistent learning success $(T2 - T0) \times (T2/30)$ (Scharfenberg et al. 2007). Increased knowledge is, hence, weighted according to the students' actual knowledge, making it possible to compare cognitive achievement despite some students exhibiting a huge increase in knowledge yet low final scores, and vice versa.

We also calculated correlations between biology grades and post-test (T1) scores for overall and model-related knowledge items using Spearman-Rho (Field 2012).

Evaluation-1 Phase In order to compare the evaluation variants for the evaluation-1 phase, we assessed students' model sketches (changed after Langheinrich and Bogner

Table 3 Knowledge item examples related to the DNA-related phases (A) and the modeling phase (B)

Item relation	Item difficulty[a] (T0/T1)	Item example
DNA-related (A)[b]	63.7/85.5	A positively charged particle migrates through the electrical field … (Item 1) (a) Between both poles (b) To the positive pole (c) To the negative pole* (d) Not at all
	40.3/60.4	The DNA encodes the genetic information of … (Item 24) (a) All organisms* (b) Great apes (c) All organisms but bacteria (d) Vertebrates
	7.4/15.1	With the aid of gel electrophoresis, you get information about … (Item 6) (a) The molecular mass* (b) The number of bindings of a molecule (c) The components of a molecule (d) The atoms of a molecule
Modelling (B)[c]	48.9/82.7	Which of the following is not part of the DNA? (Item 2) (a) Adenine (b) Ribose* (c) Guanine (d) Deoxyribose
	24.8/64.7	In 1962, James Watson and Francis Crick were honoured with the Nobel prize for the discovery of … (Item 4) (a) Of DNA located in cell nuclei (b) Components that form the DNA (c) Gel electrophoresis (d) The DNA double helix structure*
	11.6/15.1	The analysis of a DNA-section revealed a proportion of guanine with 30%. Following, the proportion of adenine is … (Item 26) (a)… not determinable. (b)…also 30%. (c) 70%. (d) 20%.*

Correct answer marked with asterisk

[a] Percentage of correct answers

[b] Sum 12 items

[c] Sum 18 items

2015; for definitions, examples, and frequencies, see ESM 3). We randomly selected 26 out of 150 drawings for a second scoring (17.3%). Cohen's kappa coefficient (Cohen 1968) scores of 0.88 and 0.82 for intra-rater and inter-rater reliability showed an "almost perfect" rating (Wolf 1997, p. 964). To avoid bias, sketches of both variants were compared and no significant difference could be identified (MWU: $Z = -0.745$, $p = .456$).

Using content analysis (Bos and Tarnai 1999), we iteratively categorized the statements that students made in response to the open questions. For the first question "Which features of the original DNA molecule are simplified in your model", four categories were employed: *level of DNA*, *level of substance*, *level of particles*, and *level of structure* (for definitions, examples, and frequencies, see ESM 4). We randomly selected 58 out of 384 statements for a second scoring

(15.1%). We computed Cohen's kappa coefficient (Cohen 1968) scores of 0.98 and 0.78 for intra-rater and inter-rater reliability, which showed a "substantial" to almost perfect rating (Wolf 1997, p. 964). For the second question, "Explain why one might create different models of one biological original (in our case, the structure of the DNA)?", we applied the adapted category system of Mierdel and Bogner (2019b) and identified five categories: *individuality of DNA*, *different interpretation*, *different model design*, *different focus*, and *different research state* (for definitions, examples, and frequencies, see ESM 5). We randomly selected 27 out of 150 statements for a second scoring (18.0%). We computed Cohen's kappa coefficient (Cohen, 1968) scores of 0.75 and 0.70 for intra-rater and inter-rater reliability, which showed a substantial rating (Wolf 1997, p. 964). To avoid bias, we compared the two evaluation variants in terms of category

Fig. 1 **a** Repeated measurement design of both instructional variants in the outreach lab. **b** Item-difficulties for T0 and T1. DNA-related items shortened "D", model-related items shortened "M". Item examples shown in Table 3 highlighted in light grey (note the shift between the pre- and post-schedule)
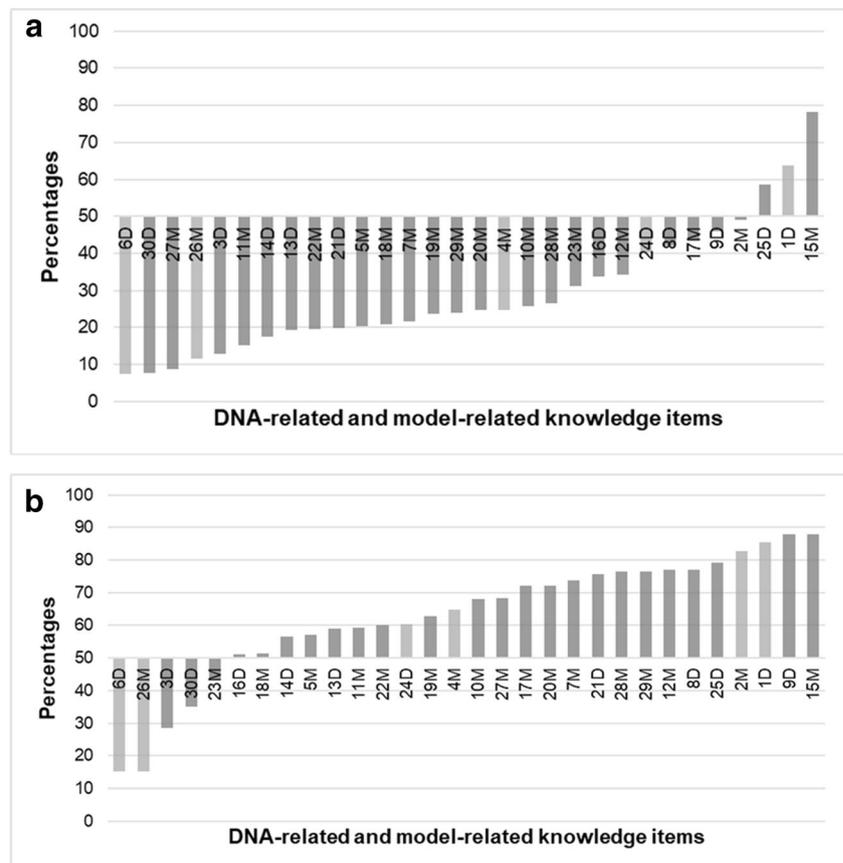
frequencies of responses to both open questions. We did not find any significant contingencies (adjusted Pearson's $C \leq .192$; $p \geq .065$; Pearson 1900).

Evaluation-2 Phase A three-step approach was applied to evaluate the evaluation-2 phase (Table 4):

- Documentation of the students' self-evaluation: We counted each box that students had tagged on their self-evaluation sheet as one point (maximal score 14 points)
- Assessment of students' self-evaluation sheets: We analyzed the tagged boxes' conformity on the self-evaluation sheets using the respective models. Appropriate tags received one point each. If students tagged all their boxes correctly, they would reach the maximal core as was recorded on their self-evaluation sheets
- Assessment of students' models: We independently assessed the models. Correct features each received one point whether or not they had been identified by the students (maximal score 14 points)

A comparison between the documented boxes and the assessment of the self-evaluation sheets enabled us to determine the extent to which students had correctly

evaluated their models. Lower scores on the self-evaluation sheet indicate students' mistakes when assessing the quality of their models. However, a lower model score also indicates that a student may have documented model features that were not given. By contrast, higher model scores indicate model features that the student did not identify as such.

Statistical Analysis

We applied nonparametric methods due to an abnormal distribution of variables (the Kolmogorov-Smirnov test (Lilliefors modification): partially $p < .001$), and, consequently, use boxplots to illustrate our results. Intra-group differences over the three test dates were analyzed using the Friedman test (F) in combination with a pairwise analysis from T0 to T1 and T2, and from T1 to T2, using the Wilcoxon (W) signed-rank test. The Mann-Whitney U tests (MWU) were used to evaluate inter-group differences. Due to multiple testing, we applied a Bonferroni correction (Field 2012). In the case of significant results, effect sizes r (Lipsey and Wilson 2001) were calculated with small (> 0.1), medium (> 0.3), and large (> 0.5) effect sizes. For correlation analyses, we applied Spearman's rank correlations and report Spearman's Rho values.

Table 4 Assessment of the modellers-2 evaluation-2 phase

Analysis sector	Description (score)	Maximum	Students' self-evaluation sheet[a]	Scores Assessment students' self- evaluation sheet[a]	Assessment students model[a]
Bases	- Four bases are indicated (2) - Base pairs correctly indicated (3) - Hydrogen bonds correctly indicated, differing in G/C and A/T (1)	6	5.1 (2.0/6.0)	4.3 (0/5.0)	5.0 (5.0/5.0)
Deoxyribose	- Deoxyribose indicated (1) - Deoxyribose linked to base (1)	2	1.3 (0/2.0)	1.0 (0/2.0)	1.5 (1.0/2.0)
Phosphate	- Phosphate indicated (1) - Phosphate and deoxyribose alternately arranged (1)	2	1.4 (0/2.0)	1.2 (0/2.0)	1.7 (1.0/2.0)
Primary structure	- Single strand visible (1) - Double strand visible (1)	2	1.6 (0.3/2.0)	1.6 (1.0/2.0)	2.0 (2.0/2.0)
Secondary structure	- Double helix visible (1) - Right-handed double helix visible (1)	2	1.1(0/2.0)	0.6 (0/1.8)	1.2 (0/2.0)
Sum		14	10.6 (6.0/13.0)	8.7 (2.5/11.0)	10.8 (8.0/13.0)

[a] Grouped median, 25th and 75th percentiles in brackets

Results

We first provide an overview of our intra-group and inter-group analyses with regard to overall and model-related knowledge. This is followed by a detailed assessment of the evaluation-2 phase.

Intra-Group Analyses of Cognitive Achievement

Intra-group analysis (F and W tests; Table 5) revealed significant changes for modellers-1 and modellers-2, in terms of both overall and model-related knowledge: They initially increased their knowledge at both levels, which then dropped between T1 and T2, but not below prior levels (T0). This suggests that students gain short-term and mid-term knowledge throughout the intervention (Table 6).

Inter-Group Analyses of Cognitive Achievement

To account for differences in students' prior knowledge (Table 6, superscript a), we calculated difference variables for short-term increases in knowledge and mid-term retention rates and learning success variables to assess inter-group differences. These were the only variables taken into account. Based on sum scores, increases in knowledge (T1-T0) and retention rate (T2-T0) were calculated (Field 2012) for overall (30 items) and model-related (18 items) knowledge scores (Table 6).

Overall Knowledge Scores of the overall knowledge test, which included both DNA-related and model-related knowledge items, showed differences in increases in knowledge and retention rates between modellers-1 and modellers-2. Modellers-2 scored significantly higher in increased knowledge and retention rate than modellers-1, with a medium-to-large effect size (Table 6, superscripts d/e).

As different variables do not display the actual cognitive achievement, we analyzed learning success variables (see above). For short-term actual and mid-term persistent learning success, significant differences in overall knowledge were identified with medium-to-large effect sizes (Table 6; superscripts h/i). Compared to modellers-1, modellers-2 achieved higher scores in terms of actual and persistent learning success (Fig. 2).

Table 5 Cognitive achievement of the students' sample as a whole and model understanding items only

Items	Friedman test			Wilcoxon signed-rank test					
				T0/T1		T0/T2		T1/T2	
	Chi-square	df	p	Z	p	Z	p	Z	p
Model-related knowledge	275.65	2	< .001	− 12.98	< .001	− 8.64	< .001	− 9.92	< .001
Overall knowledge	259.03	2	< .001	− 12.95	< .001	− 9.61	< .001	− 9.08	< .001

Table 6 Dependent variables for both modellers-1 and modellers-2, analyzed with regard to knowledge scores, difference variables, and learning success. Overall knowledge items and model-related items only were differentiated

Dependent variable	Knowledge			
	Overall knowledge items (max. 30)		Model-related knowledge items (max. 18)	
	Modellers-1	Modellers-2	Modellers-1	Modellers-2
	Mdn (25th/75th *P*)	*Mdn* (25th/75th *P*)	*Mdn* (25th/75th *P*)	*Mdn* (25th/75th *P*)
T0[a]	10.3 (8.0/13.0)	6.3 (4.0/8.0)	5.7 (4.0/8.0)	3.8 (2.0/5.0)
T1[b]	20.6 (18.0/23.0)	19.4 (16.0/22.0)	13.9 (12.0/15.0)	12.8 (11.0/15.0)
T2[c]	16.9 (13.0/20.0)	17.2 (15.0/19.0)	10.9 (9.0/13.0)	10.5 (9.0/12.0)
Difference variables				
	Overall knowledge Items (max. 30)[j]		Model-related knowledge Items (max. 18)[j]	
Increase in knowledge[d]	9.9 (7.0/12.3)	13.1 (9.0/16.0)	7.8 (6.0/10.0)	9.1 (7.0/11.0)
Retention rate[e]	6.9 (4.0/9.0)	10.4 (8.0/13.0)	5.4 (3.0/7.0)	6.6 (4.0/8.0)
Learning success				
	Overall knowledge items (max. 30)		Model-related knowledge items (max. 18)[j]	
Actual learning success[f]	6.7 (4.2/8.6)	8.7 (5.1/11.9)[h]	3.4 (2.4/4.7)	3.9 (2.4/5.5)
Persistent learning success[g]	3.7 (1.8/5.9)	6.0 (3.7/7.9)[i]	1.9 (1.0/3.2)	2.3 (1.2/3.3)

[a] MWU test overall knowledge: $Z = -7.90$; $p < .001$; $r = .552$ and model-related items $Z = -5.95$; $p < .001$; $r = .439\%$)

[b] MWU test overall knowledge: $Z = -2.26$; $p = .024$; $r = .185$ and model-related items $Z = -2.92$; $p = .004$; $r = .224$

[c] MWU test overall knowledge: $Z = -0.271$; $p = .786$ and model-related items $Z = -1.59$; $p = .111$

[d] Score: T1–T0; MWU test overall knowledge: $Z = -4.99$; $p < .001$; $r = .386$ and model-related items $Z = -2.93$; $p = .003$; $r = .270$

[e] Score: T2–T0; MWU test overall knowledge: $Z = -6.41$; $p < .001$; $r = .499$ and model-related items $Z = -2.82$; $p = .005$; $r = .304$

[f] Actual learning success: $(T1 - T0) \times (T1/30)$ for overall knowledge and $(T1 - T0) \times (T1/18)$ for model-related knowledge items

[g] Persistent learning success: $(T2 - T0) \times (T2/30)$ for overall knowledge and $(T2 - T0) \times (T2/18)$ for model-related knowledge items

[h] MWU test overall knowledge items: $Z = -3.36$; $p = .001$; $r = .293$

[i] MWU test overall knowledge items: $Z = -5.23$; $p < .001$; $r = .436$

[j] Percentage share of correctly answered items: increase in knowledge overall modellers-1 and modellers-2 (33%, 44%) < model-related modellers-1 and modellers-2 (43%, 51%); retention rate overall modellers-1 and modellers-2 (23%, 35%) < model-related modellers-1 and modellers-2 (30%, 37%)

Fig. 2 Changes in overall knowledge scores for both modellers-1 and modellers-2. All participating groups increased their actual learning success scores: $(T1 - T0) \times (T1/30)$. Persistent learning success scores: $(T2 - T0) \times (T2/30)$, however, dropped in comparison to actual learning success scores
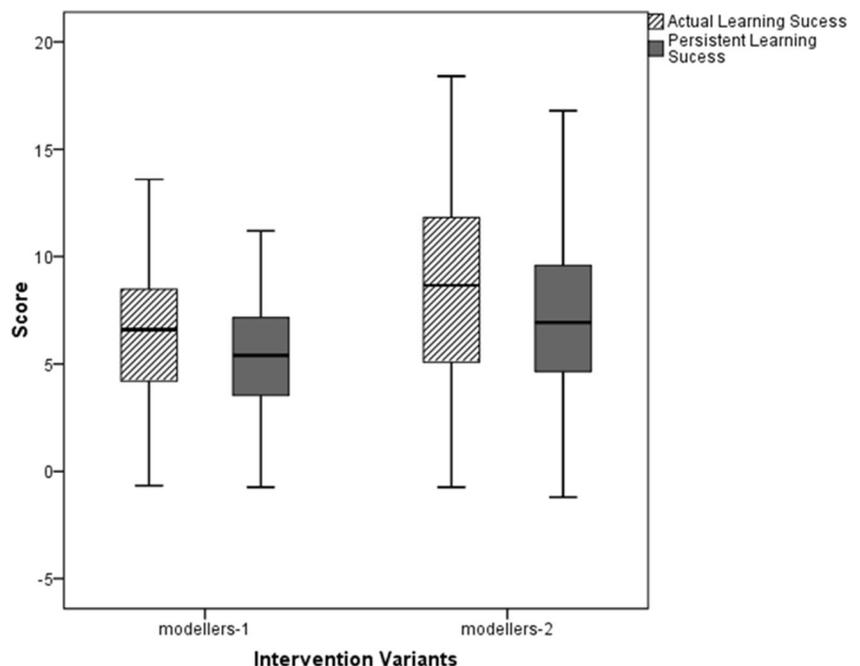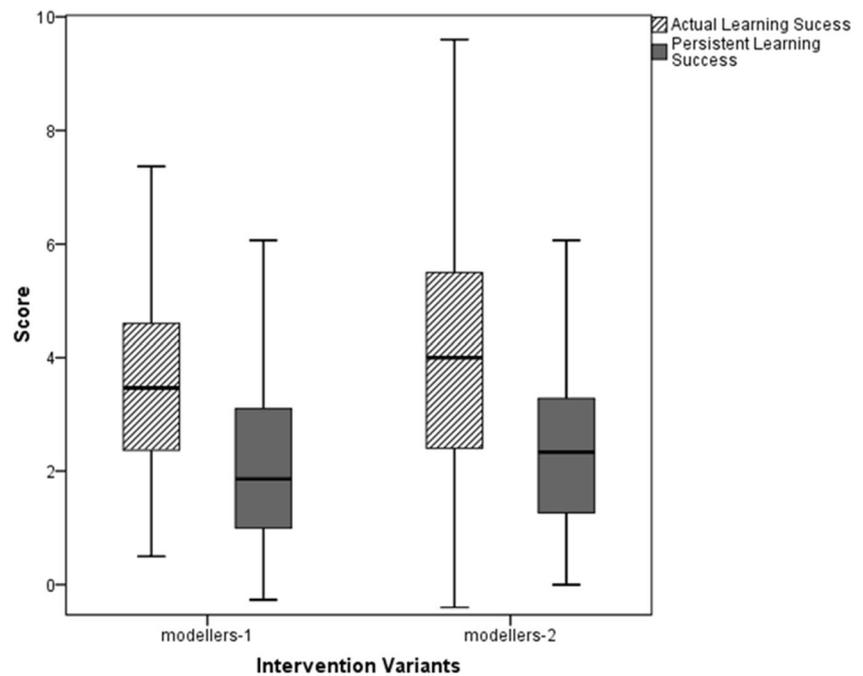
Fig. 3 Changes in all 18 model-related knowledge items for both modellers-1 and modellers-2. All participating groups increased their actual learning success scores: (T1 − T0) x (T1/18). Persistent learning success scores: (T2 − T0) x (T2/18), however, dropped in comparison to actual learning success scores

Model-Related Knowledge Regarding scores based on 18, model-related knowledge items, modellers-2 achieved significantly higher increases in knowledge and retention rates with a medium effect size (Table 6, superscripts d/e). At the level of learning success, modellers-2 achieved higher scores (Fig. 3). However, model-related cognitive achievement exceeded that of overall knowledge (Table 6, superscripts j).

Correlation Biology Grades Biology grades and post-test (T1) scores displayed a weak negative correlation ($r_S = -.209$, $p < .001$). We obtained similar results by correlating students' biology grades and post-tests for model-related knowledge ($r_S = -.202$, $p = .001$). Splitting our evaluation variants, modellers-1 did not reveal any negative correlations ($p \geq .190$); modellers-2, however, displayed significant negative correlations of $r_S = -.235$ ($p = .006$) for overall and of $r_S = -.196$ ($p = .026$) for model-related knowledge. Thus, students with lower grades tend to achieve higher scores in overall and model-related knowledge post-tests than students with better grades. This was particularly evident in our evaluation-2 variant.

Assessment of Evaluation-2 Phase

Intra-group analyses of modellers-2 revealed the differences between students' self-evaluations and our assessment of their self-evaluation sheets and their model (Table 4; F: chi-square 59.531, $df = 2$; $p < .001$). Pairwise analysis revealed lower scores for their assessed self-evaluation sheets with a large effect (W: $Z = -6.220$, $p < .001$; $r = .728$). Thus, students

identified as correct features that were not given in their model. This discrepancy was evident across all analyses sectors (Table 4; F: chi-square ≥ 16.919, $df = 2$, $p < .001$, in each case; W: $Z \leq -2.715$, $p \leq .007$, in each case; $r \geq .318$). In contrast—and also to a significant effect—some of the assessed models scored higher than the assessed self-evaluation sheets (W: $Z = -5.804$, $p < .001$; $r = .684$). Thereby, the students did not identify all of the correctly modeled features. This phenomenon was also evident across all analyses sectors (Table 4; W: $Z \leq -3.938$, $p \leq .001$, in each case; $r \geq .464$).

Correlation analysis revealed small-to-medium correlations between the models' assessment scores, reflecting the models' quality, the actual learning success ($r_S = .260$, $p = .003$), and the persistent learning success ($r_S = .251$, $p = .005$). Further correlations between the abovementioned variable and model-related actual learning success ($r_S = .215$, $p = .015$) and the model-related persistent learning success ($r_S = .218$, $p = .014$) were of particular importance. Thereby, models' assessment scores showed a small correlation with students' prior in-class experience of modeling ($r_S = .217$, $p = .014$).

Discussion

Our aim was to examine potential differences between participation in the two evaluation variants. The data suggest that students' short-term and mid-term DNA- and model-related knowledge was improved by the additional evaluation-2 phase.

Cognitive Achievement

Overall Knowledge The effects of an additional model evaluation-2 phase on the actual and persistent learning success indicate that this is a positive approach that supports effective learning. According to Ainsworth (2008), modeling affects three dimensions of learning which enable students to abstract, extend, and relate their knowledge; to identify familiar concepts; and to appropriately display multiple aspects of the respective phenomenon. Altogether, this may lead to a deeper understanding of the subject (Oh and Oh, 2011). In our previous study, the model evaluation-1 phase had already suggested that a 1-day intervention combining model-related activities and hands-on experimentation could encourage learning in the abstract field of molecular genetics (Mierdel and Bogner 2019a). The effective integration of difficult scientific theory and working techniques into biology classrooms is, thus, a promising approach, as has been argued by Peel et al. (2019). Still, there was room for improvement which is why we conducted our present intervention with the additional model evaluation-2 phase. Compared to evaluation-1 variant (modellers-1), those participating in evaluation-2 (modellers-2) achieved higher scores regarding increases in knowledge and retention rate and received higher actual and persistent learning success scores (Table 6). Students not only developed an initial sketch of their mental DNA model but also handcrafted a three-dimensional model which they assessed in a comparison-based self-evaluation during which they worked in pairs to discuss and review their approach. In science, this type of evaluation is vital to assess the adequacy of previously developed models and make efforts to improve upon them. In our study, these discussions required students to focus on providing consistent explanations and balanced assessments of their models (e.g., Passmore et al. 2009, Schwarz et al. 2009).

Model-Related Knowledge Specific analysis of model-related knowledge items revealed a higher increase in knowledge and retention rate, for modellers-2 than for modellers-1. It also revealed higher actual and persistent learning success scores for modellers-2 (Table 6). Thus, our additional evaluation phase seems to impact model-based knowledge items. Schwarz et al. (2009) and Bryce et al. (2016) suggest a link between modeling practices in the classroom and students' learning success. Our measured cognitive achievement in model-related items (Table 6) might, then, be due to a "progression in knowledge and skills required for modelling, necessarily [entailing] progression in knowledge about the nature of models" (Gilbert and Justi 2016, p. 195). This possibility is also reflected in correlation analysis, with small-to-medium correlations between models' assessment scores and actual as well as persistent learning success. Thus, we can conclude that understanding of

models is connected to knowledge about models (Peel et al. 2019).

Modeling also helped students to organize information about DNA as a model. It ultimately contributed to a deeper understanding of the learning content (e.g., Bryce et al. 2016; Grünkorn et al. 2014) and, in our case, supported students' understanding of DNA as a model.

Evaluation-2 Phase

Although actual and persistent learning success differed in both evaluation variants, there were also quantifiable differences in modellers-2's handcrafted models (for example, Table 2). Some students correctly identified and labeled the DNA's different components, while some students only identified a few and others only modeled its basic structure. This result is in line with Howell et al. (2019) and Kim et al. (2015), who described students' difficulties in understanding DNA's structure-function relationships. Other studies about scientific modeling do not focus on DNA but, for instance, on the acoustic properties of materials (Hernández et al. 2015), natural selection and antibiotic resistance (Peel et al. 2019), or general classroom examples in biology, geography, and physics (Schwarz et al. 2009). It could also be discussed that students were, to a certain extent, unable to transfer their previously obtained factual knowledge into the new form of a model. Yet, although hardly any of the students had sufficient prior experience, they nonetheless found a way to excel—an achievement particularly notable among those who were considered, grade-wise, to be low achievers (Bamberger & Davis 2011). Our own prior research suggests that most students find models (and other types of images: charts, graphs, diagrams, etc.) very difficult to engage with (personal research). Moreover, our results are in line with the findings of Quigley et al. (2017) who used the EcoSurvey tool to assess different kinds of models across different classrooms and correlated these with learning success. The authors suspected differences in modeling experience to be the underlying cause of success or failure. We also found a small correlation between the students' model quality and their prior experience in modeling at school. In our case, individual students stood out from the crowd if they either had a better understanding of models, processed new information more effectively, or read the text more carefully. Analysis of answers to DNA model–related questions on the students' worksheet from model evaluation-1 (Table 1) supports this approach (ESM 4, ESM 5). Sample answers showed a broad understanding of DNA as a model; for instance, at particle level, that "the sugar and phosphate molecules are simplified (they usually consist of various atoms)" (ESM 4). This supports the claim that modeling is closely connected to students' academic performance (Quigley et al. 2017), as long as the modeling is not only about "doing school" but has scientific relevance (Schwarz et al. 2009, p.652).

Methodological Aspects

Firstly, as noted by Hernández et al. (2015, p. 257), several "recurring cycles of generation, evaluation and modification" would have been helpful to enhance model quality and deepen model understanding. Schwarz et al. (2009) suggest a four-step approach to successfully promote progressive understanding of models and scientific modeling. Yet the limited time available in our 1-day outreach laboratory meant that this intervention was not suitable for such extensive modeling activities and we could not offer several cycles of improvement even though this might have positively influenced learning success scores (Louca et al. 2011). Such cycles would instead require regular in-class modeling and evaluation phases (e.g., José et al. 2015). To compensate for the lack of time, a more extensive pre-modeling phase prior to the outreach teaching unit could be included, during which students would be directly introduced to scientific modeling.

Second, our knowledge items about DNA and analytical methods in gene technology laboratories only provide information about the development of students' factual knowledge. Thus, more open-ended, conceptual questions, such as those in our surveys, would give a deeper insight into student learning.

Conclusion

Although models are already used in biology lessons to encourage scientific reasoning, their effectiveness is rarely scrutinized (Werner et al. 2017). Yet, certain levels of complexity in experimentation and model design are required to adequately support scientific reasoning. Rinehart et al. (2016), therefore, have argued that all *cookbook* laboratories should be replaced with authentic, epistemic, scientific practices. Maintaining authenticity is, thereby, mandatory, as merely constructing models of scientific phenomena for the sake of modeling would miss the mark. Classroom activities that explicitly introduce students to the nature of models would be far more beneficial (José et al. 2015). Moreover, continuous in-class reflection and discussion are vital for retaining scientific authenticity and familiarizing students with real-life scientific practice (Acevedo 2008). Therefore, we consider our additional evaluation phase to be another valuable approach to integrating real scientific practices into science teaching. Using outreach laboratories, we demonstrated the impact of model-supported teaching on cognitive achievement. Based on our intervention variants, modellers-2 proved to be more effective than modellers-1 when focusing on DNA structure. Our intervention also confirms the effectiveness of research-based laboratory practice and active-learning protocols for cognitive achievement. Every student approaches new learning contents differently, and our gene technology laboratory offers the required flexibility for differentiated teaching,

addressing all types and speeds of learning (e.g., Mierdel and Bogner 2019a, b; Chen et al. 2016). Thus, teachers can apply our model evaluation and active-learning approach in the classroom and in other science subjects (e.g., in chemistry education for modeling protein structure (Torres & Correia 2007)). This will, of course, require science teachers to create new materials suitable for complex, inquiry-based lessons. Nonetheless, the modeling has the potential to encourage students to hypothesize, assess the accuracy of explanations, and identify knowledge gaps and is, thus, worth the effort (Svoboda and Passmore 2013). In future studies, extending the modeling phase to include several model-evaluation cycles, as suggested by Hernández et al. (2015), and assessing the impact on actual and persistent learning success would be of interest.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

Ethical Statement

1. Hereby, we, the authors, consciously assure that for the manuscript "Self-evaluative scientific modeling in an outreach gene technology laboratory" the following is fulfilled:

   (a) This material is the authors' own original work, which has not been previously published elsewhere.
   (b) The paper is not currently being considered for publication elsewhere.
   (c) The paper reflects the authors' own research and analysis in a truthful and complete manner.
   (d) The paper properly credits the meaningful contributions of co-authors and co-researchers.
   (e) The results are appropriately placed in the context of prior and existing research.
   (f) All the sources used are properly disclosed (correct citation). Literally copying of text must be indicated as such by using quotation marks and giving proper reference.
   (g) All the authors have been personally and actively involved in substantial work leading to the paper and will take public responsibility for its content.

2. Moreover, "all the procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards."

Consent Statement "Informed consent was obtained from all individual participants included in the study."

References

Abdulwahed, M., & Nagy, Z. K. (2013). Applying Kolb's experiential learning cycle for laboratory education. *J Eng Educ, 98,* 283–294.

Acevedo, J. A. (2008). El estado actual de la naturaleza de la ciencia en la didáctica de lasciencias [The state of the art on nature of science in Science Education]. *Revista Eureka sobre Enseñanza y Divulgación de las Ciencias, 5*(2), 134–169.

Ainsworth, S. (2008). The educational value of multiple-representations when learning complex scientific concepts. In J. K. Gilbert, M. Reiner, & M. Nakhleh (Eds.), *Visualization: Theory and practice in science education* (pp. 191–208). London: Springer.

Ates, Ö., & Erylmaz, A. (2011). Effectiveness of hands-on and minds-on activities on students' achievement and attitudes towards physics. *Asia-Pacific Forum on Science Learning and Teaching, 12,* 1–23.

Bamberger, Y. M., & Davis, E. A. (2011). Middle-school science students' scientific modelling performances across content areas and within a learning progression. *Int J Sci Educ, 35*(2), 213–238.

Bortz, J., & Döring, N. (1995). Forschungsmethoden und Evaluation [research methods and evaluation] (2nd ed.) springer, Berlin.

Bos, W., & Tarnai, C. (1999). Content analysis in empirical social research. *Int J Educ Res, 31*(8), 659–671.

Brody, M., Bangert, A., & Dillon, J. (2007). Assessing learning in informal science contexts. National Research Council for Science Learning in Informal Environments Committee. http://www7.nationalacademies.org/bose/Brody_Commissioned_Paper.pdf. .

Bryce, C. M., et al. (2016). Exploring models in the biology classroom. *Am Biol Teach, 8,* 35–42.

Carpenter, S. L., et al. (2018). Models are a "metaphor in your brain": How potential and preservice teachers understand the science and engineering practice of modeling. *Sci Educ, 119,* 275–286.

Chen, W., Shah, U., & Brechtelsbauer, C. (2016). The discovery laboratory – A student-centred experiential learning practical: Part I – Overview. *Educ Chem Eng, 17,* 44–53.

Cohen, J. (1968). Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychol Bull, 70*(4), 213–220.

Cook, T. D., & Campell, D. (1979). *Quasi-experimentation. Design & analysis issues for field settings*. Chicago: Rand McNally College Publishing Company.

Declaration of Helsinki. (2013). World Medical Association. *Journal of the American Medical Association, 310,* 2191–2194.

De Houwer, J., Barnes Holmes, D., & Moors, A. (2011). Why the cognitive approach in psychology would profit from a functional approach and vice versa. *Perspect Psychol Sci, 6*(2), 202–209.

Domjan, M. (2015). *The principles of learning and behavior* (7th ed.). Stamford, CT: Cengage Learning.

Field, A. (2012). *Discovering statistics using IBM SPSS statistics* (4th ed.). London: SAGE.

Franco, C., & Colinvaux, D. (2000). Grasping mental models. In J. K. Gilbert & C. J. Boulter (Eds.), *Developing models in science education (pp. 93–118)*. Springer.

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

Glowinski, I., & Bayrhuber, H. (2011). Student labs on a university campus as a type of out-of-school learning environment: Assessing the potential to promote students' interest in science. *International Journal of Environmental and Science Education, 6,* 371–392.

Grünkorn, J., Upmeier zu Belzen, A., & Krüger, D. (2014). Assessing students' understandings of biological models and their use in science to evaluate a theoretical framework. *Int J Sci Educ, 36*(10), 1651–1684.

Haselgrove, M. (2016). *Learning: A very short introduction*. Oxford: Oxford University Press.

Hernández, M. I., Couso, D., & Pintó, R. (2015). Analyzing students' learning progressions throughout a teaching sequence on acoustic properties of materials with a model-based inquiry approach. *Journal of Science Education Technology, 24*(2-3), 356–377.

Hodson, D. (1998). *Teaching and learning science - A personalized approach*. Buckingham: Open University Press.

Howell, M. E., Booth, C. S., Sikich, S. M., Helikart, T., Roston, R. L., Couch, B. A., & van Dijk, K. (2019). Student understanding of DNA structure–function relationships improves from using 3D learning modules with dynamic 3D printed models. *Biochem Mol Biol Educ, 47*(3), 303–317.

Jackson, J., Dukerich, L., & Hestenes, D. (2008). Modeling instruction: An effective model for science education. *Sci Educ, 17,* 10–17.

José, M. O., Del Mar Aragón, M., & Cuesta, J. (2015). The competence of modelling in learning chemical change: A study with secondary school students. *Int J Sci Math Educ, 13,* 751–791.

Justi, R. S., & Gilbert, J. K. (2002). Modelling, teachers' views on the nature of modelling, and implications for the education of modellers. *Int J Sci Educ, 24*(4), 369–387.

Kalyuga, S. (2009). Knowledge elaboration: A cognitive load perspective. *Learn Instr, 19*(5), 402–410.

Kim, et al. (2015). Cycles of exploration, reflection, and consolidation in model-based learning of genetics. *J Sci Educ Technol, 24*(6), 789–802.

Kind, V. (2015). Preservice science teachers' science teaching orientations and beliefs about science. *Sci Educ, 100,* 122–152.

KMK (2005). Beschlüsse der Kultusministerkonferenz – Bildungsstandards im Fach Biologie für den Mittleren Bildungsabschluss [resolution of the standing conference of the ministers of education and cultural Affairs of the Länder in the Federal Republic of Germany - standards of biology education for secondary school]. Munich: Luchterhand.

Kolb, D. A. (1984). *Experiential learning: Experience as the source of learning and development*. Englewood Cliffs, NJ: Prentice-Hall.

Kovanović, V., et al. (2018). Understanding the relationship between technology-use and cognitive presence in MOOCs. In *Proceedings of the Seventh International Conference on Learning Analytics and Knowledge, 582–583*. New York: Association for Computing Machinery.

Krajcik, J., & Merritt, J. (2012). Engaging students in scientific practices: What does constructing and revising models look like in the science classroom? *Sci Child, 49,* 10–13.

Lachman, S. J. (1997). Learning is a process: Toward an improved definition of learning. *J Psychol, 131*(5), 477–480.

Langheinrich, J., & Bogner, F. X. (2015). Student conceptions about the DNA structure within a hierarchical organizational level: Improvement by experiment- and computer-based outreach learning. *Biochemistry and Molecular Biology Education, 43*(6), 393–402.

Lienert, G. A., & Raatz, U. (1998). *Testaufbau und Testanalyse [test setup and test analysis]* (6th ed.). Psychologie Verlags Union: Weinheim.

Lipsey, M. W., & Wilson, D. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage Publications.

Louca, L. T., Zacharia, Z. C., & Constantinou, C. P. (2011). In quest of productive modelling-based learning discourse in elementary school science. *J Res Sci Teach, 48*(8), 919–951.

Lovey, B. R., & Riggs, K. M. (2019). Flipping the laboratory: Improving student engagement and learning outcomes in second year science courses. *Int J Sci Educ, 41*(1), 64–79.

Mierdel, J., & Bogner, F.X. (2019a). Investigations of modellers and model viewers in an out-of-school gene technology laboratory. *Research in Science Education*, https://doi.org/10.1007/s11165-019-09871-3.

Mierdel, J., & Bogner, F. X. (2019b). Comparing the Use of Two Different Model Approaches on Students' Understanding of DNA Models. *Education Sciences, 9*(115), 1–18.

Nasir, N. S., Rosebery, A. S., Warren, B., & Lee, C. D. (2006). Learning as a cultural process: Achieving equity through diversity. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 489–504). Cambridge: Cambridge University Press.

Oh, P. S., & Oh, S. J. (2011). What teachers of science need to know about models: An overview. *Int J Sci Educ, 33*(8), 1109–1130.

Olympiou, G., & Zacharia, Z. C. (2012). Blending physical and virtual manipulatives: An effort to improve students' conceptual understanding through science laboratory experimentation. *Sci Educ, 96*(1), 21–47.

Orhan, T. Y., & Sahin, N. (2018). The impact of innovative teaching approaches on biotechnology knowledge and laboratory experiences of science teachers. *Education Sciences, 8*(4), 213–237.

Passmore, C., Stewart, J., & Cartier, J. (2009). Model-based inquiry and school science: Creating connections. *Sch Sci Math, 109*(7), 394–402.

Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philos Mag, 50*, 157–175.

Peel, A., Zangori, L., Friedrichsen, P., Hayes, E., & Sadler, T. (2019). Students' model-based explanations about natural selection and antibiotic resistance through socio-scientific issues-based learning. *Int J Sci Educ, 41*(4), 510–532.

Prabha, S. (2016). Laboratory experiences for prospective science teachers: A meta-analytic review of issues and concerns. *Eur Sci J, 12*(34), 235–250.

Quigley, D., Ostwald, J., & Sumner, T. (2017). Scientific modelling: Using learning analytics to examine student practices and classroom variation. Proceedings of the Seventh International Learning Analytics and Knowledge Conference, 329–338. New York: Association for Computing Machinery.

Raviv, A., Cohen, S., & Aflalo, E. (2019). How should students learn in the school science laboratory? The benefits of cooperative learning. *Res Sci Educ, 49*(2), 331–345.

Rinehart, R., Duncan, R., Chinn, C., Atkins, T., & DiBenedetti, J. (2016). Critical design decisions for successful model-based inquiry in science classroom. *International Journal of Design for Learning, 7*, 17–40.

Rost, J. (2004). *Lehrbuch Testtheorie–Testkonstruktion [textbook test theory–test construction]* (2nd ed.). Bern: Hans Huber.

Rotbain, Y., Marbach-Ad, G., & Stavy, R. (2006). Effect of bead and illustrations models on high school students' achievement in molecular genetics. *J Res Sci Teach, 43*(5), 500–529.

Saka, A., Cerrah, L., Akdeniz, A. R., & Ayas, A. (2006). A cross-age study of the understanding of three genetic concepts: How do they imagine the gene, DNA and chromosome? *J Sci Educ Technol, 15*(2), 192–202.

Sarmouk, C., et al. (2019). Pre-laboratory online learning resource improves preparedness and performance in pharmaceutical sciences practical classes. Innovations in Education and Teaching International, online first, 1–12.

Scharfenberg, F. J., Bogner, F. X., & Klautke, S. (2007). Learning in a gene technology laboratory with educational focus: Results of a teaching unit with authentic experiments. *Biochemistry and Molecular Biology Education, 35*(1), 28–39.

Scharfenberg, F. J., & Bogner, F. X. (2013). Teaching Gene Technology in an Outreach Lab: Students' Assigned Cognitive load clusters and the clusters' relationships to learner characteristics, laboratory variables, and cognitive achievement. *Research in Science Education, 43*, 141–161.

Schwarz, C., et al. (2009). Developing a learning progression for scientific modeling: Making scientific modeling accessible and meaningful for learners. *J Res Sci Teach, 46*(6), 632–654.

Solomon, J. (1980). *Teaching children in the laboratory*. London: Crom Helm.

Svoboda, J., & Passmore, C. (2013). The strategies of modeling in biology education. *Sci & Educ, 22*(1), 119–142.

Thorpe, W. H. (1943). Types of learning in insects and other arthropods. *Br J Psychol, 33*, 220–234.

Torres, B. B., & Correia, P. R. M. (2007). Using denatured egg white as a macroscopic model for teaching protein structure and introducing protein synthesis for high school students. *J Chem Educ, 84*(12), 1941–1944.

Usher, S. (2013). *Letters of note. Correspondence deserving of a wider audience*. Edinburgh: Canongate.

Werner, S., et al. (2017). Investigating how German biology teachers use three-dimensional physical models in classroom instruction: A video study. *Res Sci Educ, 49*, 1–27.

Windschitl, M., & Thompson, J. (2006). Transcending simple forms of school science investigation: The impact of preservice instruction on teachers' understandings of model-based inquiry. *Am Educ Res J, 43*(4), 783–835.

Wolf, R. (1997). Rating scales. In J. Keeves (Ed.), *Educational research, methodology and measurement: An international handbook* (pp. 958–965). Oxford: Elsevier.

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory Pract, 41*(2), 64–70.