

Article

Preservice Biology Teachers' Scientific Reasoning Skills and Beliefs about Nature of Science: How Do They Develop and Is There a Mutual Relationship during the Development?

Daniela Mahler ^{1,2,*} , Denise Bock ²  and Till Bruckermann ^{3,*} ¹ Biology Education, Freie Universität Berlin, 14195 Berlin, Germany² IPN—Leibniz Institute for Science and Mathematics Education, 24118 Kiel, Germany; bock@leibniz-ipn.de³ Institute of Education, Leibniz University Hannover, 30167 Hannover, Germany

* Correspondence: daniela.mahler@fu-berlin.de (D.M.); till.bruckermann@iew.uni-hannover.de (T.B.)

† Both authors contributed equally.

Abstract: Scientific reasoning (SR) skills and nature of science (NOS) beliefs represent important characteristics of biology teachers' professional competence. In particular, teacher education at university is formative for the professionalization of future teachers and is thus the focus of the current study. Our study aimed to examine the development of SR skills and NOS beliefs and their mutual relationship during teacher education. We applied paper-and-pencil tests to measure SR skills and NOS beliefs of 299 preservice biology teachers from 25 universities in Germany. The results of linear mixed models and planned comparisons revealed that both SR skills and NOS beliefs develop over the course of the study. Nevertheless, the development of SR skills and multiple aspects of NOS beliefs proceeds in different trajectories. Cross-lagged models showed a complex picture concerning the mutual relationship between SR skills and NOS beliefs during their development (both positive and negative). The current study contributes to the existing research because it is based on longitudinal data and allows—in contrast to cross-sectional research—conclusions about the *development* of SR skills and NOS beliefs.

Keywords: scientific reasoning; nature of science; preservice teachers; longitudinal study; cross-lagged panel



Citation: Mahler, D.; Bock, D.; Bruckermann, T. Preservice Biology Teachers' Scientific Reasoning Skills and Beliefs about Nature of Science: How Do They Develop and Is There a Mutual Relationship during the Development? *Educ. Sci.* **2021**, *11*, 558. <https://doi.org/10.3390/educsci11090558>

Academic Editors: Eila Jeronen, Moritz Krell, Andreas Vorholzer and Andreas Nehring

Received: 6 August 2021

Accepted: 13 September 2021

Published: 18 September 2021

Corrected: 30 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Fostering the scientific literacy of students is one of the core aims of science education in schools (e.g., [1] [Germany]; [2] [U.S.]). Science teachers' scientific reasoning (SR) skills and their beliefs about nature of science (NOS) represent key domains of science teachers' professional competence regarding science as inquiry. Science teachers with higher proficiency in SR skills are more likely to promote inquiry-based learning of their students [3,4]. Furthermore, science teachers need adequate NOS beliefs to integrate NOS teaching practices in their classrooms [5,6]. Accordingly, SR skills and NOS beliefs should be considered equally important as knowledge of other science concepts [7]. In different countries, standard documents of university teacher education, therefore, include SR skills and NOS beliefs (e.g., [8] [Germany]; [9] [U.S.]).

Because teacher education at the university is one of the most formative phases of the professionalization of teachers and the development of their professional competence [10], it should also be considered an important starting point for the development of SR skills and NOS beliefs. Previous research shows that preservice teachers' SR skills and NOS beliefs improve, at least to some degree, during teacher education at university, and that this development is related to appropriate learning opportunities provided in science education courses [11–13]. Regarding SR skills, preservice teachers are more skilled in graduate science education courses than students in graduate courses that did not explicitly

reflect upon scientific inquiry [12]. The more courses in university teacher education referred to NOS concepts, the higher is preservice teachers' understanding of NOS [11]. Nevertheless, empirical evidence based on longitudinal data allowing statements about the development of SR skills and NOS beliefs is rare (see [13], for an overview). In response to this desideratum, the first objective of the present study is to investigate the development of preservice science teachers' SR skills and NOS beliefs throughout university teacher education.

Although SR skills and NOS beliefs are two separate domains of teachers' professional competence, their interplay is important: Within teachers' professional competence, SR skills and NOS beliefs may be closely related because they both concern the development of scientific knowledge [14]. Whereas SR skills reflect preservice teachers' procedural knowledge about scientific inquiry processes (i.e., knowing how), NOS beliefs reflect their evaluations of how scientific knowledge comes into being (i.e., knowing why; [15]). Most research to date, however, was only able to correlate SR skills and NOS beliefs in cross-sectional designs [14,16]. Thus, the mutual relationship between SR skills and NOS beliefs and their development over time remains unexplored. For example, it is unclear whether NOS beliefs are beneficial for SR skills or vice versa (see [17], for an overview). Our second objective addresses this research gap by exploring the mutual relationship between SR skills and NOS beliefs during their development. With our results, we aim to contribute to understanding both constructs and their relationship, and the improvement in teacher education at university regarding SR skills and NOS beliefs.

1.1. Scientific Reasoning Skills and Beliefs about Nature of Science as Domains of Science Teachers' Professional Competence

To foster students' scientific literacy, science teachers need specific characteristics that are located within the concept of professional competence [18]. Professional competence covers different aspects of successful teachers' professional knowledge, motivational orientations, beliefs, values, and goals, and self-regulative skills [18]. These aspects are a critical resource for teachers to promote student learning [19–21]. One of the core assumptions of professional competence is that future teachers do not enter their careers with all of the desired characteristics but acquire them over time. This assumption implies that the aspects of teachers' professional competence are, in principle, learnable [22].

1.2. Scientific Reasoning Skills

SR skills are a procedural facet of teachers' content knowledge. Content knowledge represents the content-related domain of professional knowledge [18,23]. More specifically, SR skills represent the "knowing how" (i.e., about scientific inquiry processes; [15,24,25]). SR skills refer to an individual's ability to solve problems scientifically, that is, to a domain-specific set of knowledge and skills for scientific inquiry processes, which differ from domain-general cognitive strategies [26,27]. SR skills are different from domain-general cognitive abilities with which they only have medium but positive correlations [28]. SR skills comprise several subskills, such as formulating hypotheses, planning investigations, and analyzing and interpreting data. Furthermore, SR skills comprise different dimensions related to methods such as observing, investigations, and modeling [12,29,30]. The theoretical framework corresponding to the current study more precisely defines four subskills of SR (i.e., "formulating questions", "generating hypotheses", "planning investigations", "analyzing data and drawing conclusions") in the dimension of "conducting scientific investigations" [31] (p. 264). We refer to "skills" because they are—other than intelligence—trainable and also distinct from conceptual knowledge (see [32], for a similar account).

Regarding the assessment of SR skills, previous research noted a low validity of questionnaires based on ill-defined constructs of SR skills [25,33]. There have been few attempts to establish the psychometric quality of the questionnaires (e.g., [34]; see [32], for an overview). One notable exception was the development of a multiple-choice questionnaire for science teachers by theory-based item selection (see [31], for an overview), and testing of its validity [12,30] and applicability in different countries [29,35].

1.3. Beliefs about Nature of Science

Whereas scientific reasoning reflects the different activities in scientific inquiry (e.g., forming hypotheses, planning investigations, and analyzing and interpreting data), NOS reflects the epistemological basis of scientific inquiry and knowledge [5]. NOS beliefs, therefore, are different from knowledge of scientific inquiry. The location of NOS beliefs within the teachers' professional competence model reflects this conceptual difference [18], because NOS beliefs conceptually belong to teachers' beliefs [36]. Teachers' beliefs are defined as "psychologically held understandings and assumptions about phenomena or objects of the world that are felt to be true, have both implicit and explicit aspects and influence people's interactions with the world" [36] (p. 250). More precisely, NOS beliefs conceptually belong to teachers' epistemological beliefs related to the nature of knowledge or a particular science (see, [37] for mathematics education; see [38], for an overview). In science education, NOS beliefs reflect preservice teachers' evaluations of the characteristics of scientific knowledge and its production [5]. Despite the ongoing debate about the general aspects conceptualization of NOS, science education researchers, to a certain degree, agree on the inclusion of seven to ten aspects in the NOS conceptualization (i.e., the consensus view on aspects to be taught in schools; [39]). Previous research aligned the following aspects from different NOS conceptualizations: tentativeness; observations and inferences; creativity and imagination; subjectivity and objectivity; social and cultural embeddedness; diversity of scientific methods; and scientific theories and laws [39,40]. Recently, a comprehensive account of professional competence that teachers need for effective NOS instruction [38] added to the description of preservice teachers' NOS beliefs during teacher education [13] and provided a more prescriptive framework for teacher education.

The assessment of preservice teachers' NOS beliefs with questionnaires reflects the general aspects conceptualization of NOS. Questionnaires to assess NOS beliefs follow qualitative approaches, such as the Views of Nature of Science Questionnaire (VNOS; [41]), or quantitative Likert-type approaches, such as the questionnaire Student Understanding of Science and Scientific Inquiry (SUSI; [42,43]). The Likert-type SUSI questionnaire is based on the aspects from the VNOS [43]. Although Likert-type NOS questionnaires have been criticized [41], they are especially useful in research that assesses larger samples or repeatedly tests individuals and investigates the relationship between NOS beliefs and other constructs [44].

1.4. Development during Teacher Education

Kunter et al. [10] describe the first phase of teacher education at university as incredibly formative for the professionalization of teachers. Neumann and colleagues [45] provide a concise overview of the German teacher education system (in which our study is situated). Prospective teachers can choose from different teacher education programs to qualify for different school tracks (primary school, non-academic track, or academic track). Typically, they study two subjects. For prospective science teachers, it is essential to mention that they can study the separate science disciplines (biology, chemistry, and physics). On average, teacher education programs last five years (three years for the bachelor's phase and two years for the master's phase).

Both preservice teachers' SR skills and NOS beliefs profit from learning opportunities in science education courses during teacher education at university [11–13]. According to previous research, academic training generally appears to promote the development of SR skills [12,26,46]. Other research, however, suggests that the development of SR skills is more pronounced when courses promote explicit reflection on scientific inquiry [47–49]. Therefore, the development of preservice teachers' SR skills may vary throughout teacher education at university. In a cross-sectional study, university courses that require explicit reflection are part of the postgraduate phase of university teacher education [12]. The authors of this study assume that explicit reflection improved science teacher students' SR skills that were higher than those of natural science students, and they suggest further exploring this assumption in longitudinal studies. Adding to the cross-sectional findings,

one study provides evidence for the—at least moderate—development of SR skills during university teacher education in a longitudinal study on preservice teachers from two universities [31]. The development of SR skills was evident from four time points: the first and fourth semester of their undergraduate studies, and the first and fourth semester of their postgraduate studies (i.e., the 7th and 10th semesters in total).

Most research on preservice teachers' NOS beliefs stems from cross-sectional analysis and repeatedly shows that they do not possess what is considered to be adequate beliefs about NOS (e.g., [43]). Cross-sectional findings highlight that (future) science teachers often have an exclusively positive or idealistic image of science, even when the researchers accounted for the number of teaching years, the type of teacher education program, and the discipline (see [5], for an overview). Other research focused on how to promote adequate NOS beliefs and highlighted the effectiveness of explicit and reflective instruction [50–52]. Nonetheless, some aspects of NOS beliefs are more difficult to change than others (e.g., differences between scientific laws and theories; [52]). Less research focuses on how preservice teachers' NOS beliefs develop over time during university teacher education [11,53]. One study found a decline in adequate NOS beliefs in a sample of Turkish preservice teachers, although this study was cross-sectional [53]. Another study showed that adequate NOS beliefs increase with learning opportunities provided during university teacher education [11]. This study, however, also took a cross-sectional approach and did not consider how NOS beliefs of individual preservice teachers develop over time. Thus, longitudinal studies of preservice teachers' NOS beliefs are needed, particularly to explore how different aspects of NOS beliefs develop in relation to their difficulty [13].

1.5. The Interplay between SR Skills and NOS Beliefs

The two constructs may be related to each other because SR skills reflect the knowledge of how to pose questions scientifically, whereas NOS beliefs reflect the knowledge of why scientific inquiry proceeds in specific ways [14,54]. Both SR skills and NOS beliefs can be enhanced by appropriate instruction, such as explicit reflection about scientific inquiry [47,55]. Therefore, explicit teaching about scientific inquiry may lead to more appropriate NOS beliefs, for example, that theories are subject to change. Conversely, more appropriate NOS beliefs may promote more profound SR skills, such as drawing valid conclusions from data (see [16], for a similar account on the nature of scientific inquiry). Most research, however, either studied the development of SR skills (e.g., [12]) or NOS beliefs [11,53]. Other research that assessed both SR skills and NOS beliefs neither investigated their relationship (e.g., [56]) nor established a theoretical framework of how specific beliefs and skills may be related (e.g., [16]). Recently, the theoretical ScieNo-framework was developed [14]. In their framework, the authors assume that specific beliefs about the nature of scientific inquiry are related to specific SR skills: for example, SR skills for observations are related to views about the role of theory in observations [14,54]. Beyond the above-mentioned theoretical assumptions, there is also empirical evidence indicating a relationship between SR skills and NOS beliefs. The two studies that examined the relationship between SR skills and NOS beliefs [14,16], however, were cross-sectional, and they were not able to investigate the relation between two constructs during development. Furthermore, a longitudinal framework enables the investigation of the directions of the effects. Therefore, we suggest testing which NOS beliefs are related to SR skills in a longitudinal study.

1.6. The Current Research

In the current research, we investigated the development of SR skills and NOS beliefs in a longitudinal study with preservice biology teachers from German universities. Our study extends previous studies that investigated preservice teachers' SR skills (e.g., [12]) and NOS beliefs (e.g., [11,53]) only in cross-sectional designs and, therefore, were not able to describe the development using a longitudinal approach. In line with previous cross-sectional studies, we expected preservice teachers' SR skills and NOS beliefs to increase.

Furthermore, the current research investigated the mutual relationship between preservice teachers' SR skills and NOS beliefs during university teacher education. Our study extends previous research on the relationship between SR skills and NOS beliefs using a longitudinal approach to discern causal relationships between the constructs through a cross-lagged panel design. In line with previous studies that highlighted small to medium positive correlations between SR skills and NOS beliefs [14,16], we expected a positive mutual relationship between preservice teachers' SR skills and NOS beliefs. However, we did not assume a specific effect from one construct on another, because previous studies have only shown positive correlations. The following research questions guided our study:

1. How do preservice biology teachers' SR skills and their NOS beliefs develop over time during university teacher education?
2. How are preservice biology teachers' SR skills and NOS beliefs related to each other during the course of university teacher education?

2. Materials and Methods

2.1. Study Framework and Participants

This study was conducted in the longitudinal KeiLa project (Development of professional competence in science and mathematics teacher education). In KeiLa, preservice science and mathematics teachers from 25 universities in Germany attended up to four 4 h paper-and-pencil assessments between 2014 and 2017. The surveys took place independently of specific learning opportunities or courses. Instead, extra appointments were offered to participate in the study. We did this to obtain a general overview of the development of SR skills and NOS beliefs over the course of the study rather than to examine the effectiveness of specific learning opportunities.

In the current study, we refer to data of 299 preservice teachers (76% female; $M_{\text{age}} = 21.36$ years at first attendance, $SD_{\text{age}} = 2.59$). In the KeiLa project, a sequential-cohort design was conducted. We obtained annual data of preservice teachers enrolled in four consecutive semesters of semesters 1 to 11 throughout the four measurement points of the sequential-cohort design. All preservice teachers gave their informed consent for inclusion before they participated in the study. The study was conducted in accordance with the Declaration of Helsinki, and no approval of the protocol by the local Ethics Committee was necessary. The reason for this is that the testing was carried out anonymously and proceeded in the familiar surroundings of university lecture halls, therefore causing no distress to the participating preservice teachers.

2.2. Instruments

2.2.1. Scientific Reasoning Skills

We assessed the SR skills of preservice biology teachers with 12 items developed in the Ko-WADiS project [12,31]. The single-choice items cover four subskills of SR with three items each: (1) formulating questions, (2) generating hypotheses, (3) planning investigations, and (4) analyzing data and drawing conclusions (see Table 1, for means and standard errors). We report a one-dimensional model based on dimensionality tests (see Section 2.3.2. Preliminary Analyses). The reliability of the scale is sufficient ($EAP/PV_{\text{Rel}} = 0.54$; based on concurrent calibration).

Table 1. Means and standard errors for scientific reasoning (SR) and nature of science (NOS) subscales of semesters 1 to 7.

Scale	1	3	5	7
	M^a (SE)	M (SE)	M (SE)	M (SE)
SR skills ^b	0.00 (0.07)	0.15 (0.08)	0.38 (0.07)	0.54 (0.08)
NOS beliefs ^c				
Observations and inferences	3.59 (0.05)	3.70 (0.06)	3.76 (0.05)	3.77 (0.06)
Tentativeness	3.84 (0.05)	3.89 (0.05)	3.94 (0.04)	4.10 (0.05)

Table 1. *Cont.*

Scale	1	3	5	7
	<i>M</i> ^a (SE)	<i>M</i> (SE)	<i>M</i> (SE)	<i>M</i> (SE)
Scientific theories and laws	2.67 (0.05)	2.74 (0.06)	2.80 (0.05)	2.84 (0.06)
Social/cultural embeddedness	3.46 (0.06)	3.65 (0.07)	3.72 (0.06)	3.80 (0.07)
Creativity and imagination	3.26 (0.06)	3.47 (0.07)	3.37 (0.06)	3.39 (0.07)
Scientific methods	3.67 (0.04)	3.71 (0.04)	3.74 (0.04)	3.87 (0.05)

^a Estimated marginal means and respective standard errors are based on linear mixed models; ^b Mean values of SR skills are based on WLE scores and can take any values centered around zero; ^c Mean values of NOS beliefs are based on Likert-type scales which range from 1 to 5.

2.2.2. Nature of Science Beliefs

We measured NOS beliefs with the “Student Understanding of Science and Scientific Inquiry” [42]. This contains 24 items that were assessed on 5-point Likert scales comprising 1 (*does not apply at all*), 2 (*does rather not apply*), 3 (*uncertain*), 4 (*largely applies*), and 5 (*fully applies*). The six NOS subscales (1) observations and inferences, (2) tentativeness, (3) scientific theories and laws, (4) social and cultural embeddedness, (5) creativity and imagination, and (6) scientific methods, were assessed with four items each (see Table 1, for means and standard errors). Ranges of the reliabilities (Cronbach’s α) of the subscales were as follows throughout the four semesters: from 0.45 to 0.68 for observations and inferences, from 0.50 to 0.59 for tentativeness, from 0.19 to 0.30 for scientific theories and laws, from 0.69 to 0.78 for social and cultural embeddedness, from 0.53 to 0.69 for creativity and imagination, and from 0.28 to 0.51 for scientific methods. They were calculated in R [57] with the “psych” package [58].

2.3. Analyses

2.3.1. Data Preparation

In our data set, we included participants from semesters 1 to 7 ($n_1 = 141$, $n_3 = 101$, $n_5 = 155$, $n_7 = 101$) because sample sizes in semesters 9 and 11 were too small for our analyses ($n_9 = 48$, $n_{11} = 8$). In our data set, preservice teachers were assigned to the respective semesters independent of the measurement points. Thereby, data were reshaped from the sequential-cohort design of the study to a longitudinal design.

2.3.2. Preliminary Analyses

In a first step, we conducted confirmatory factor analyses in Mplus [59] to check the assumed dimensionality of the constructs based on the subscales of the instruments (NOS beliefs: [42]; SR skills: [31]). Results revealed that a six-dimensional NOS model and a one-dimensional model of SR fitted the data significantly better than a one-dimensional model and a four-dimensional model, respectively (see Table 2).

Additionally, we calculated weighted likelihood estimation (WLE; [60]) scores for SR skills based on a one-parameter logistic item response theory model with concurrent calibration in R [57] using the “TAM” package [61].

Table 2. Chi-square difference ($\Delta\chi^2$), degrees of freedom difference (Δdf) and *p*-value of model comparison for one- and four-dimensional models of scientific reasoning (SR), and one- and six-dimensional models of nature of science (NOS) for semesters 1 to 7.

Semester	SR Skills			NOS Beliefs		
	$\Delta\chi^2$	Δdf	<i>p</i>	$\Delta\chi^2$	Δdf	<i>p</i>
1	7.44	6	0.283	181.44	15	<0.001
3	^a			161.61	15	<0.001
5	4.50	6	0.609	191.93	15	<0.001
7	10.11	6	0.120	142.83	15	<0.001

^a The four-dimensional model did not converge in semester 3.

2.3.3. Analyses concerning the Development

We chose a linear mixed model approach (LMM: [62]) to test if time (i.e., semesters) has a significant effect on preservice teachers' SR skills and NOS beliefs; that is, if SR skills and NOS beliefs develop over time. LMMs extend simple linear models by allowing both fixed and random effects. These models show various advantages. First, unlike in a repeated-measures analysis of variance, missing values can be easily handled (e.g., with restricted maximum likelihood [REML] estimation). Second, LMMs enable us to take the nested structure of our data (repeated observations nested in participants) into account. Thus, we can control for unobserved, time-invariant differences between participants. Third, LMMs allow us to control for specific autocorrelation structures, which can occur in repeated measures.

All LMMs were computed separately for each subscale of NOS beliefs and for SR skills. We fixed the correlations between time points to zero because previous checks revealed no critical autocorrelation structure to be considered. In our models, we treated the semester variable as a numeric fixed effect and the participants' ID as a grouping variable for the random effect, and we applied REML estimation. In addition to p -values, we computed the variance that is explained by all fixed effects (i.e., marginal R^2) and by fixed and random effects (i.e., conditional R^2) [63] because the trustworthiness of p -values provided for LMMs is the object of ongoing statistical discussions [64].

Finally, we examined planned comparisons of time points to further examine between which semesters significant mean changes for SR skills and NOS beliefs occur. First, we contrasted semesters 1 and 7 for SR skills and each subscale of NOS beliefs. In a second step, we compared consecutive semesters (i.e., semesters 1 vs. 3, 3 vs. 5, 5 vs. 7). p -values of the multiple comparisons were Bonferroni–Holm adjusted. We additionally calculated effect sizes (Cohen's d) based on the t -statistics for every comparison [65]. These are generally interpreted as small ($d = 0.2$), medium ($d = 0.5$), and large ($d = 0.8$; [66]).

We used R [57] with the "nlme" package for LMMs [67], the "MuMIn" package for R^2 values [68], the "emmeans" package for planned comparisons [69], and the "effectsize" package for effect sizes d [70].

2.3.4. Analyses concerning the Mutual Relationship

We used a cross-lagged panel design with four waves (semesters 1, 3, 5, and 7) and specified the respective path models to examine the interactions of SR skills and NOS beliefs during teacher education. In the cross-lagged models, estimates of a later time point from one construct can directly be regressed on values of the previous time point from another construct (i.e., cross-lagged paths), and vice versa. Furthermore, we allowed parallel time points to be correlated and autoregressive paths. We computed a single model for each NOS subscale and its relationship with SR skills, including all four time points. Cross-lagged models were computed in R [57] with the "lavaan" package [71].

3. Results

3.1. Development of Scientific Reasoning Skills

Based on the linear mixed model (LMM), the semester has a significant effect on scientific reasoning values ($B = 0.09$, $SE = 0.02$, $t(198) = 5.43$, $p < 0.001$) with the marginal $R^2_m = 0.05$ and the conditional $R^2_c = 0.37$ (see Appendix A Table A1, for detailed LMM results).

The direct comparison of semester 1 with semester 7 shows that the mean of semester 7 is significantly higher than the mean of semester 1 (Estimate = 0.54, $SE = 0.11$, $t = 5.00$, $p < 0.001$, $d = 0.36$). Comparisons between sequential semesters yield no significant differences (Table 3; see Appendix A Table A2, for detailed comparison results).

Table 3. Estimates (Est.), standard errors (SE), and effect sizes (*d*) for subsequent time point comparisons of scientific reasoning and nature of science subscales.

Scale	1 vs. 7		1 vs. 3		3 vs. 5		5 vs. 7	
	Est. (SE)	<i>d</i>	Est. (SE)	<i>d</i>	Est. (SE)	<i>d</i>	Est. (SE)	<i>d</i>
Scientific reasoning skills	0.54 *** (0.11)	0.36	0.15 (0.10)	0.11	0.23 (0.10)	0.16	0.16 (0.10)	0.11
Nature of science beliefs								
Observations and inferenc.	0.17 * (0.07)	0.17	0.11 (0.06)	0.12	0.05 (0.06)	0.06	0.01 (0.06)	0.01
Tentativeness	0.26 *** (0.06)	0.28	0.05 (0.06)	0.06	0.05 (0.06)	0.06	0.16 * (0.06)	0.20
Scientific theories and laws	0.18 * (0.07)	0.17	0.07 (0.07)	0.08	0.06 (0.07)	0.06	0.04 (0.07)	0.05
Social and cultural embed.	0.34 *** (0.08)	0.29	0.19 * (0.08)	0.18	0.07 (0.08)	0.07	0.08 (0.07)	0.08
Creativity and imagination	0.14 (0.09)	0.11	0.22 * (0.08)	0.19	−0.10 (0.08)	−0.09	0.03 (0.08)	0.02
Scientific methods	0.20 ** (0.06)	0.23	0.05 (0.06)	0.06	0.03 (0.06)	0.04	0.12 (0.06)	0.15

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

3.2. Development of Nature of Science Beliefs

Based on the LMMs, the semester has a significant effect on the subscales observations and inferences ($B = 0.03$, $SE = 0.01$, $t(199) = 2.70$, $p = 0.008$, $R^2_m = 0.01$, $R^2_c = 0.59$), tentativeness ($B = 0.04$, $SE = 0.01$, $t(199) = 3.81$, $p < 0.001$, $R^2_m = 0.02$, $R^2_c = 0.58$), scientific laws and theories ($B = 0.03$, $SE = 0.01$, $t(199) = 2.60$, $p = 0.010$, $R^2_m = 0.01$, $R^2_c = 0.44$), social and cultural embeddedness ($B = 0.06$, $SE = 0.01$, $t(199) = 4.30$, $p < 0.001$, $R^2_m = 0.03$, $R^2_c = 0.53$), and scientific methods ($B = 0.03$, $SE = 0.01$, $t(199) = 3.07$, $p = 0.002$, $R^2_m = 0.01$, $R^2_c = 0.42$; see Appendix A Tables A3–A8, for detailed LMM results).

When comparing semester 1 and semester 7, we found a significant increase for five NOS subscales, that is, for observations and inferences (Estimate = 0.17, $SE = 0.07$, $t = 2.44$, $p = 0.011$, $d = 0.17$), tentativeness (Estimate = 0.26, $SE = 0.06$, $t = 4.00$, $p < 0.001$, $d = 0.28$), scientific laws and theories (Estimate = 0.17, $SE = 0.07$, $t = 2.41$, $p = 0.016$, $d = 0.17$), social and cultural embeddedness (Estimate = 0.34, $SE = 0.08$, $t = 4.12$, $p < 0.001$, $d = 0.29$), and scientific methods (Estimate = 0.20, $SE = 0.06$, $t = 3.22$, $p = 0.001$, $d = 0.23$). Sequential comparisons show that no consistent significant change from semester to semester occurs (Table 3; see Appendix A Tables A9–A14, for detailed comparison results).

3.3. The Mutual Relationship between SR Skills and NOS Beliefs

We found no significant cross-lagged paths between SR skills and the NOS subscales observations and inferences (all B s $< |0.13|$, all SE s < 0.16 , all p s > 0.05), scientific theories and laws (all B s $< |0.24|$, all SE s < 0.17 , all p s > 0.05), or creativity and imagination (all B s $< |0.15|$, all SE s < 0.15 , all p s > 0.05). However, we found significant relations in the other cross-lagged models, which are described in the following: A positive influence was found of NOS beliefs subscale tentativeness at semester 1 on SR skills at semester 3 ($B = 0.48$, $SE = 0.13$, $p < 0.001$). Another positive influence was found for social and cultural embeddedness beliefs at semester 5 on SR skills at semester 7 ($B = 0.30$, $SE = 0.12$, $p = 0.014$). Additionally, SR skills at semester 3 negatively influence scientific methods beliefs at semester 5 ($B = -0.22$, $SE = 0.07$, $p = 0.002$; see Figure 1).

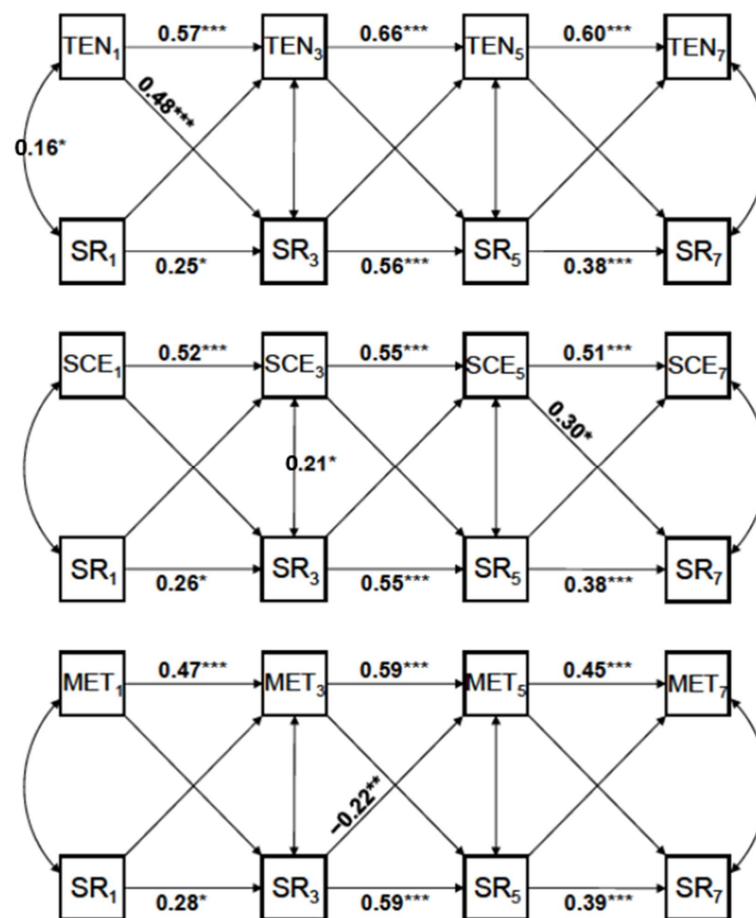


Figure 1. Unstandardized regression weights for (auto-)regression paths of cross-lagged models computed for scientific reasoning and nature of science subscales; tentativeness, TEN; social and cultural embeddedness, SCE; and scientific methods, MET; numbers in subscript refer to the respective semester; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

4. Discussion

Science teachers' SR skills and NOS beliefs are essential characteristics when teaching science as inquiry [3–6]. To date, there are few longitudinal findings on either construct that allow making statements about their development during teacher education (e.g., SR skills: [31]). Thus, previous research left a gap concerning the development of individual preservice teachers over more extended periods of university teacher education [13]. Our study aimed to close this gap by taking a longitudinal approach to the development of preservice biology teachers' SR skills and NOS beliefs throughout university teacher education, and their mutual relationship during development. In line with evidence from cross-sectional studies, we assumed that both preservice teachers' SR skills (e.g., [12]), and NOS beliefs (e.g., [61]; cf. [62]), improve throughout teacher education at the university. Furthermore, we assumed a positive relationship between SR skills and NOS beliefs. However, we did not assume any direction in their relationship because previous findings were based on cross-sectional studies [14,16].

First, our results indicate that both preservice teachers' SR skills and NOS beliefs improved over the semesters of teacher education at university. The linear mixed models (LMM) revealed a positive impact of the semester variable on preservice teachers' SR skills, and on five of six NOS subscales, that is, observations and inferences, tentativeness, scientific theory and laws, social and cultural embeddedness, and scientific methods, but not creativity and imagination. Thus, our longitudinal study provides evidence that both develop throughout teacher education. When we account for the semesters in the

LMMs, they explain at least a slight variance (SR skills: 5% variance explained; NOS beliefs: 1–3% variance explained). We argue that the amount of explained variance appears plausible with regard to cross-sectional findings from other research. Our results support a previous study that found a comparable amount of variance of preservice teachers' NOS beliefs explained by semesters (i.e., 4%; [11]). Although our results indicate that preservice teachers' development of SR skills and NOS beliefs depends to some degree on their attendance of consecutive semesters, numerous other factors besides the semester remain unexplored. In this regard, the relatively high conditional R^2 values refer to a large amount of variance explained by time-invariant differences between participants, that is, differences that do not change over the considered period. These differences could be, for example, individual prerequisites such as a previously acquired degree (for SR skills: [35]) or the respective subject area of preservice teachers (for SR skills: [12]; for NOS beliefs: [51]).

Second, we explored when preservice teachers' SR skills and NOS beliefs develop throughout university teacher education by planned comparisons between the semesters 1 and 7 or 1 and 3, 3 and 5, and 5 and 7. For comparing semesters 1 and 7, we found that both SR skills and five of six aspects of NOS beliefs show a small to moderate increase in our study. Our results extend previous results from cross-sectional SR research [12,46] by using a longitudinal approach showing that preservice teachers' SR skills increase during semesters 1 to 7. The magnitude of this increase is comparable—at least at a descriptive level—to that in longitudinal data of semesters 1 to 7 from a previous study [31]. Although our results strengthen cross-sectional findings of the general development of NOS beliefs during teacher education (e.g., [11,52]; cf. [50,53]), most NOS beliefs' means at semester 7 still range between 3 (*uncertain*) and 4 (*largely applies*) on the Likert scale. The mean of the NOS subscale scientific theories and laws even remains below 3 throughout its development. Thus, we cannot assume that preservice teachers' development leads to *informed* views of NOS at the sample level [42,43,72].

Our more detailed analyses of the in-between semesters (1 vs. 3, 3 vs. 5, and 5 vs. 7), however, show that preservice teachers' SR skills and NOS beliefs developed in differing trajectories that are inconsistent in three ways: (1) they do not significantly improve during the in-between semesters, that is, for preservice teachers' SR skills and NOS beliefs about observations and inferences, scientific theories and laws, and scientific methods, only the comparison between semester 1 and 7 is significant; (2) some NOS beliefs do not steadily improve, such as tentativeness and social and cultural embeddedness, but show a significant increase with a small effect size ($d > 0.19$) only between two of the four consecutive semesters; (3) a slight decrease follows after initial positive development of NOS beliefs about creativity and imagination. These results suggest that the development does not just happen incidentally but that something must happen during teacher education that triggers this inconsistent picture of different trajectories. In principle, other studies also found that—at least for NOS beliefs—a decrease throughout teacher education is also possible [53]. Previous research highlighted that not all NOS aspects are equally changeable [13,52], which matches our result that only five aspects develop throughout teacher education. Our results complement prior research that showed that creativity and imagination, for example, are more likely to change [52], in that our results show that this belief changes mainly in the first few semesters and subsequently stagnates. Therefore, we assume that, in addition to the complexity of some NOS aspects [13] and individual differences among preservice teachers [50], learning opportunities in each semester should also be considered [11,12]. We assume that the uneven development of preservice teachers' SR skills and NOS beliefs depends on university teacher education's different learning opportunities. Previous research supports our assumption by indicating that learning opportunities are not equally distributed across teacher education at the university regarding pedagogical knowledge [73] and content knowledge, such as SR skills and NOS beliefs [11,12]. Thus, preservice teachers' SR skills and NOS beliefs are more likely to develop through multiple, interacting learning opportunities than through linearly cumulative learning opportunities (see [50], for a similar account). To further

understand the inconsistent picture, studies addressing learning opportunities during teacher education may help. A cross-sectional study found that the number of NOS-related learning opportunities is positively related to preservice biology teachers' NOS beliefs [11]. Both SR skills [12,74] and NOS beliefs [50–52] benefit from explicit and reflective learning opportunities. Accordingly, to further understand our results (e.g., why there are phases during the course of study that are more important for the development of SR skills and NOS beliefs compared to others), a closer examination of the number of learning opportunities, their distribution throughout teacher education, and their type (i.e., implicit vs. explicit) would have been helpful.

Third, we explored the mutual relationship between preservice teachers' SR skills and NOS beliefs during their development. We found that preservice teachers' NOS beliefs about tentativeness and social and cultural embeddedness positively influenced their SR skills: Less naïve NOS beliefs about tentativeness in semester 1 led to more profound SR skills in semester 3. Furthermore, less naïve NOS beliefs about the social and cultural embeddedness in semester 5 led to more profound SR skills in semester 7. Our longitudinal results align with previous research that established a positive correlation between NOS beliefs and SR skills in cross-sectional studies [14,16]. Furthermore, a mutual relationship between the SR skills and NOS beliefs appears plausible because both constructs refer to knowledge of scientific inquiry (knowing how and knowing why; [15]), and they can be improved through similar instructional approaches [47,55]. However, our longitudinal results also extend previous findings because our results revealed a much more inconsistent relationship that was limited to only some NOS beliefs and not stable across all semesters. The ScieNo-framework [14] may help to understand the inconsistent picture: skills for specific inquiry methods (such as conducting investigations and using models; i.e., dimensions of SR: [31]) are related to specific beliefs that are conceptually close (e.g., skills for observations and beliefs about the role of theory in observations). Therefore, we suggest that not all NOS beliefs are equally related to the SR skills of conducting investigations. In particular, NOS beliefs about tentativeness and the social and cultural embeddedness are challenging to grasp for preservice teachers [13,52], so that more adequate beliefs may have enhanced their SR skills in the following semesters. Other NOS beliefs that may be learned more easily probably do not positively influence SR skills. Future research should test our assumption that those NOS beliefs, in particular, that are more difficult to learn positively influence SR skills. Although we could separate six different NOS beliefs, we cannot relate them at this level of detail to different subskills of SR (e.g., formulating hypotheses) because we could not empirically separate the subscales for the SR skills. Furthermore, we used a short questionnaire that comprised 12 items of the dimension conducting scientific investigations from the whole item set that also includes the dimension of using models (see [31], for an overview). Thus, we suggest further research exploring the mutual relationship between different dimensions of SR skills, that is, for observing, experimenting, and modeling, and NOS beliefs in longitudinal studies.

Interestingly, we also found that preservice teachers' SR skills at semester 3 negatively influence their NOS beliefs about scientific methods at semester 5. Preservice teachers with more profound SR skills later had more naïve beliefs about scientific methods. We suspect that this negative effect may be related to a distortion in their beliefs about scientific methods when preservice teachers learn about methods of scientific inquiry in university teacher education. If a preservice teacher masters one inquiry method, such as how to conduct proper investigations, particularly well, this may lead to the idealistic (not appropriate) belief that this method is superior to the others. Our conjecture is in line with previous findings that show how naïve NOS beliefs develop with increasing study progress or Ph.D. degrees [53]. The authors explain this with the assumptions of Kuhn [75], who pointed out that during active engagement in research, the epistemological foundation fades into the background.

Furthermore, in university teacher education, science education courses have been shown to emphasize investigations, and particularly experiments, as teaching methods

that may promote preservice teachers' beliefs that there is only one scientific method [76]. Another explanation may be that limiting our SR questionnaire on the SR dimension of conducting investigations for test-economic reasons (see also [30]) may have led to a one-sided focus among the preservice biology teachers. Asking them only about conducting valid investigations probably made them believe that this was the only scientific method when they filled out the questionnaire on NOS beliefs. For future studies, we would recommend investigating such interactions between questionnaires on SR skills and NOS beliefs, and reflecting a greater variety of methods in the questionnaire on SR skills.

4.1. Strengths, Limitations, and Future Research

To the best of our knowledge, this study is the first to investigate the development of both preservice teachers' SR skills and NOS beliefs and their mutual relationship in a longitudinal design. More precisely, our study makes an essential contribution to the understanding of their development by using a longitudinal data set from 25 universities with adequate sample sizes and established instruments. Nevertheless, some limitations of this study should be discussed.

We used established and validated instruments for the assessment of both the SR skills and NOS beliefs (Ko-WADiS instrument: [31]; SUSSI instrument: [42]). Nevertheless, the reliabilities are partly in an unsatisfactory range in that they might have hindered us from detecting more substantial changes by the longitudinal design [50]. In comparison to previous research, the reliabilities determined in the current study are in the range of the typical values for the SUSSI, except for the subscale theories and laws (i.e., Cronbach's $\alpha = 0.44\text{--}0.89$: [42]; $\alpha = 0.16\text{--}0.86$: [77]) and for the Ko-WADiS instrument (i.e., EAP/PV reliability is: 0.54: [12]; 0.55: [34]).

The current study was designed to examine data from preservice teachers in semesters 1 to 11. Because the sample sizes were too small for semesters 9 and 11, these data had to be excluded from our analyses. Thus, we can only make statements for the bachelor's program that precedes the master's program in teacher education. We suggest for future research to examine the development of preservice teachers' SR skills and NOS beliefs during the master's program, because previous research suggests that the explicit learning opportunities that are particularly effective for both SR skills and NOS beliefs tend to occur in the latter part of teacher education [12].

The ScieNo-framework [14] helped us understand the mutual relationship between SR skills and NOS beliefs because it aligns the dimensions of SR skills, such as observing, experimenting, and modeling, with specific NOS beliefs. Unfortunately, we only used the 12 item short version of the Ko-WADiS instrument on the SR dimension when conducting investigations [78]. The full range of items includes another three SR skills of using models [31], so that the SR and NOS scales may be related to each other in a planned manner. In addition to test-economic reasons that led to the use of only one subscale for SR skills, it should be mentioned that the theoretical framework [14] was published after the current study was conducted between 2014 and 2017. However, our results highlight the mutual relationship between SR skills and the NOS beliefs about tentativeness, social and cultural embeddedness, and scientific methods. These mutual relationships are worth further investigation with more closely aligned instruments that are based on the ScieNo-framework.

4.2. Implications

Our results lead to implications, both for further research and for teacher education at university. They show that SR skills and multiple aspects of NOS beliefs develop differently; that is, some are easier to learn than others (e.g., [13]). The longitudinal study approach also suggests that preservice teachers' development of SR skills and NOS beliefs take different trajectories throughout teacher education, that is, the time point of development differs. As a next step, it would be essential to understand why SR skills and NOS beliefs take different trajectories during university teacher education. For this, the consideration of learning

opportunities is essential. Different studies suggest that explicit learning opportunities (i.e., learning opportunities that provide the opportunity for reflection) are particularly effective for developing SR skills and NOS beliefs; development of SR skills and NOS beliefs does not happen on the side. Accordingly, not only is the number of learning opportunities essential, but also their focus (implicit vs. explicit). In order to consider this, one could either refer to module manuals of teacher education or ask preservice teachers to report on the learning opportunities they had between two measurement points. The latter approach appears more promising because it provides information not only about the intended curriculum, but also about the implemented curriculum, that is, what actually took place [79]. We know from previous research that learning opportunities to explicitly reflect on scientific inquiry appear mainly in the master's program of German teacher education at the university [12]. Accordingly, it would be necessary to have an appropriate sample that covers both participants in the bachelor's program and the master's program.

Our results on the relationship between SR skills and NOS beliefs show an inconsistent picture in that there is no mutual relationship between all aspects of NOS beliefs and SR skills. To learn more about their mutual relationship, the ScieNo-framework [14] can be consulted to derive and further investigate hypotheses regarding the relationship between specific aspects of NOS beliefs and SR skills; for example, SR skills for observations are related to views about the role of theory in observations. A longitudinal approach—as in this study—would offer two advantages. First, it subjects the framework to empirical testing in a longitudinal perspective. Second, it would allow us to examine the extent to which learning opportunities for SR skills are also conducive to NOS beliefs, and vice versa. However, careful planning is necessary for such an investigation. For example, when selecting the instruments, it must be ensured that the constructs can be combined at an appropriate level of detail.

Our results also provide suggestions for improving teacher education. Our results on the different trajectories of preservice teachers' SR skills and NOS beliefs are significant in this regard. Our results make clear that preservice teachers' SR skills and NOS beliefs do not simply co-evolve but that their mutual relationship is much more complex across teacher education at university. The positive influence of specific aspects of preservice teachers' NOS beliefs on their SR skills depends on how difficult certain NOS aspects are to learn and when they develop due to learning opportunities in teacher education. Accordingly, a blanket consideration in teacher education is not expedient because inquiry-based learning does not automatically lead to the development of SR skills and NOS beliefs. Instead, learning opportunities must be created that explicitly relate the corresponding SR skills and NOS beliefs to each other (ScieNo-framework; [14]) and ideally provide space for reflection on this interplay.

Furthermore, we found that the more profound SR skills preservice teachers had in semester 3, the less informed were their NOS beliefs about scientific methods in semester 5. We suggest that this might stem from the negative impact of a bias concerning a single scientific method. It is possible that a strong focus on conducting scientific inquiry in undergraduate studies (i.e., doing science: [80]; e.g., [12]), and in our test, leads preservice teachers to idealistic but inadequate beliefs about methods in scientific research. Thus, teacher education should reflect the broad repertoire of methods in scientific research and provide opportunities for reflection on the use of those methods.

5. Conclusions

We investigated the development of SR skills and NOS beliefs—two characteristics of effective science teachers—and their mutual relationship during the undergraduate studies of teacher education at university. Our results add to previous research by taking a longitudinal approach to show how SR skills and NOS beliefs develop throughout teacher education. We present evidence for differing trajectories in the development of SR skills and multiple aspects of NOS beliefs that hints at the importance of further investigation of the learning opportunities. The present findings do not yet account for

learning opportunities in university teacher education, so it would be interesting to see whether the trajectories vary in university teacher education of other countries. If patterns emerge in the country comparison, the different learning opportunities leading to these trajectories can be examined in more detail. Furthermore, we present evidence that the mutual relationship between SR skills and NOS beliefs is stronger for specific aspects of preservice teachers' NOS beliefs and specific time points in their development. Thus, during teacher education at university, preservice teachers' SR skills and NOS beliefs are intertwined in their development, but further research is needed to truly understand their interplay and dependence on learning opportunities.

Author Contributions: Conceptualization, D.M. and T.B.; formal analysis, D.B.; data curation, D.M. and D.B.; writing—original draft preparation, D.M., D.B. and T.B.; writing—review and editing, D.M. and T.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Leibniz Association under grant number SAW-2014-IPN-1.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and no approval of the protocol by the local Ethics Committee was necessary. The reason for this is that the testing was carried out anonymously and proceeded in the familiar surroundings of university lecture halls, therefore causing no distress to the participating preservice teachers.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to [The study is ongoing].

Acknowledgments: We thank Ute Harms for project administration and funding acquisition for the study at hand.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. Outcomes of the linear mixed model and effect sizes (marginal R^2_m for fixed effects, conditional R^2_c for fixed and random effects) for *scientific reasoning* with semester as a fixed effect and participants as a random effect.

Random Effects	Variance	SD			
Participant (Intercept)	0.26	0.51			
Residual	0.50	0.71			
Fixed Effects	B	SE	df	t	p
Intercept	0.00	0.07	293	−0.07	0.943
Semester	0.09	0.02	198	5.43	<0.001
Effect Size	R^2_m	R^2_c			
Semester	0.05	0.37			

Table A2. Estimates, standard errors (SE), t-values, p-values, and effect sizes (d) for planned comparisons of *scientific reasoning*.

Parameter	1 vs. 7	1 vs. 3	3 vs. 5	5 vs. 7
Estimate	0.54	0.15	0.23	0.16
SE	0.11	0.10	0.10	0.10
t	5.00	1.50	2.30	1.57
p	<0.001	0.235	0.067	0.235
d [95% CI]	0.36 [0.21, 0.50]	0.11 [−0.03, 0.25]	0.16 [0.02, 0.31]	0.11 [−0.03, 0.35]

Table A3. Outcomes of the linear mixed model and effect sizes (marginal R^2_m for fixed effects, conditional R^2_c for fixed and random effects) for *observations and inferences* with semester as a fixed effect and participants as a random effect.

Random Effects	Variance	SD			
Participant (Intercept)	0.25	0.50			
Residual	0.18	0.43			
Fixed Effects	B	SE	df	t	p
Intercept	3.61	0.05	295	75.68	<0.001
Semester	0.03	0.01	199	2.70	0.008
Effect Size	R^2_m	R^2_c			
Semester	0.01	0.59			

Table A4. Outcomes of the linear mixed model and effect sizes (marginal R^2_m for fixed effects, conditional R^2_c for fixed and random effects) for *tentativeness* with semester as a fixed effect and participants as a random effect.

Random Effects	Variance	SD			
Participant (Intercept)	0.20	0.45			
Residual	0.15	0.39			
Fixed Effects	B	SE	df	t	p
Intercept	3.82	0.04	294	88.59	<0.001
Semester	0.04	0.01	199	3.81	<0.001
Effect Size	R^2_m	R^2_c			
Semester	0.02	0.58			

Table A5. Outcomes of the linear mixed model and effect sizes (marginal R^2_m for fixed effects, conditional R^2_c for fixed and random effects) for *scientific theories and laws* with semester as a fixed effect and participants as a random effect.

Random Effects	Variance	SD			
Participant (Intercept)	0.17	0.41			
Residual	0.22	0.46			
Fixed Effects	B	SE	df	t	p
Intercept	2.67	0.05	294	58.24	<0.001
Semester	0.03	0.01	199	2.60	0.010
Effect Size	R^2_m	R^2_c			
Semester	0.01	0.44			

Table A6. Outcomes of the linear mixed model and effect sizes (marginal R^2_m for fixed effects, conditional R^2_c for fixed and random effects) for *social and cultural embeddedness* with semester as a fixed effect and participants as a random effect.

Random Effects	Variance	SD			
Participant (Intercept)	0.29	0.53			
Residual	0.27	0.51			
Fixed Effects	B	SE	df	t	p
Intercept	3.48	0.05	294	63.80	<0.001
Semester	0.06	0.01	199	4.30	<0.001
Effect Size	R^2_m	R^2_c			
Semester	0.03	0.53			

Table A7. Outcomes of the linear mixed model and effect sizes (marginal R^2_m for fixed effects, conditional R^2_c for fixed and random effects) for *creativity and imagination* with semester as a fixed effect and participants as a random effect.

Random Effects	Variance	SD			
Participant (Intercept)	0.21	0.46			
Residual	0.33	0.57			
Fixed Effects	B	SE	df	t	p
Intercept	3.30	0.05	294	60.34	<0.001
Semester	0.02	0.01	199	1.40	0.164
Effect Size	R^2_m	R^2_c			
Semester	0.00	0.40			

Table A8. Outcomes of the linear mixed model and effect sizes (marginal R^2_m for fixed effects, conditional R^2_c for fixed and random effects) for *scientific methods* with semester as a fixed effect and participants as a random effect.

Random Effects	Variance	SD			
Participant (Intercept)	0.11	0.33			
Residual	0.16	0.40			
Fixed Effects	B	SE	df	t	p
Intercept	3.66	0.04	294	95.05	<0.001
Semester	0.03	0.01	199	3.07	0.002
Effect Size	R^2_m	R^2_c			
Semester	0.02	0.42			

Table A9. Estimates, standard errors, *t*-values, *p*-values, and effect sizes (*d*) for planned comparisons of *observations and inferences*.

Parameter	1 vs. 7	1 vs. 3	3 vs. 5	5 vs. 7
Estimate	0.17	0.11	0.05	0.01
SE	0.07	0.06	0.06	0.06
<i>t</i>	2.44	1.71	0.85	0.15
<i>p</i>	0.015	0.269	0.796	0.877
<i>d</i> [95% CI]	0.17 [0.03, 0.31]	0.12 [−0.02, 0.26]	0.06 [−0.08, 0.20]	0.01 [−0.13, 0.15]

Table A10. Estimates, standard errors, *t*-values, *p*-values, and effect sizes (*d*) for planned comparisons of *tentativeness*.

Parameter	1 vs. 7	1 vs. 3	3 vs. 5	5 vs. 7
Estimate	0.26	0.05	0.05	0.16
SE	0.06	0.06	0.06	0.06
<i>t</i>	4.00	0.81	0.86	2.76
<i>p</i>	<0.001	0.779	0.779	0.019
<i>d</i> [95% CI]	0.28 [0.14, 0.43]	0.06 [−0.08, 0.20]	0.06 [−0.08, 0.20]	0.20 [0.06, 0.34]

Table A11. Estimates, standard errors, *t*-values, *p*-values, and effect sizes (*d*) for planned comparisons of scientific theories and laws.

Parameter	1 vs. 7	1 vs. 3	3 vs. 5	5 vs. 7
Estimate	0.18	0.07	0.06	0.04
SE	0.07	0.07	0.07	0.07
<i>t</i>	2.41	1.06	0.90	0.65
<i>p</i>	0.017	0.873	0.873	0.873
<i>d</i> [95% CI]	0.17 [0.03, 0.31]	0.08 [−0.06, 0.22]	0.06 [−0.08, 0.20]	0.05 [−0.09, 0.19]

Table A12. Estimates, standard errors, *t*-values, *p*-values, and effect sizes (*d*) for planned comparisons of social and cultural embeddedness.

Parameter	1 vs. 7	1 vs. 3	3 vs. 5	5 vs. 7
Estimate	0.34	0.19	0.07	0.08
SE	0.08	0.08	0.08	0.07
<i>t</i>	4.11	2.53	0.92	1.07
<i>p</i>	<0.001	0.037	0.571	0.571
<i>d</i> [95% CI]	0.29 [0.15, 0.43]	0.18 [0.04, 0.32]	0.07 [−0.07, 0.21]	0.08 [−0.06, 0.22]

Table A13. Estimates, standard errors, *t*-values, *p*-values, and effect sizes (*d*) for planned comparisons of creativity and imagination.

Parameter	1 vs. 7	1 vs. 3	3 vs. 5	5 vs. 7
Estimate	0.14	0.22	−0.10	0.03
SE	0.09	0.08	0.08	0.08
<i>t</i>	1.60	2.68	−1.27	0.31
<i>p</i>	0.112	0.024	0.412	0.755
<i>d</i> [95% CI]	0.11 [−0.03, 0.25]	0.19 [0.05, 0.33]	−0.09 [−0.23, 0.05]	0.02 [−0.12, 0.16]

Table A14. Estimates, standard errors, *t*-values, *p*-values, and effect sizes (*d*) for planned comparisons of scientific methods.

Parameter	1 vs. 7	1 vs. 3	3 vs. 5	5 vs. 7
Estimate	0.20	0.05	0.03	0.12
SE	0.06	0.06	0.06	0.06
<i>t</i>	3.22	0.81	0.52	2.16
<i>p</i>	0.002	0.842	0.842	0.096
<i>d</i> [95% CI]	0.23 [0.09, 0.37]	0.06 [−0.08, 0.20]	0.04 [−0.10, 0.18]	0.15 [0.01, 0.29]

References

1. Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. *Bildungsstandards im Fach Biologie für den Mittleren Schulabschluss [National Educational Standards for the Intermediate School Leaving Certificate in Biology]*; Luchterhand: München, Germany, 2005.
2. NGSS Lead States. *Next Generation Science Standards: For States, by States*; The National Academies Press: Washington, DC, USA, 2013.
3. Capps, D.K.; Crawford, B.A.; Constat, M.A. A review of empirical literature on inquiry professional development: Alignment with best practices and a critique of the findings. *J. Sci. Teacher Educ.* **2012**, *23*, 291–318. [[CrossRef](#)]
4. Schwarz, C. Developing preservice elementary teachers' knowledge and practices through modeling-centered scientific inquiry. *Sci. Educ.* **2009**, *93*, 720–744. [[CrossRef](#)]
5. Lederman, N.G.; Lederman, J.S. Research on Teaching and Learning of Nature of Science. In *Handbook of Research on Science Education*; Lederman, N.G., Abell, S.K., Eds.; Routledge: New York, NY, USA, 2014; ISBN 9780203097267.
6. Capps, D.K.; Crawford, B.A. Inquiry-based professional development: What does it take to support teachers in learning about inquiry and nature of science? *Int. J. Sci. Educ.* **2013**, *35*, 1947–1978. [[CrossRef](#)]

7. Glaze, A. Teaching and Learning Science in the 21st Century: Challenging Critical Assumptions in Post-Secondary Science. *Educ. Sci.* **2018**, *8*, 12. [CrossRef]
8. Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. Ländergemeinsame inhaltliche Anforderungen für die Fachwissenschaften und Fachdidaktiken in der Lehrerbildung (Beschluss der Kultusministerkonferenz vom 16.10.2008 i. d. F. vom 16.05.2019). 2019. Available online: https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2008/2008_10_16-Fachprofile-Lehrerbildung.pdf (accessed on 7 September 2021).
9. National Science Teacher Association. NSTA Standards for Science Teacher Preparation. Available online: <https://www.nsta.org/nsta-standards-science-teacher-preparation> (accessed on 7 September 2021).
10. Kunter, M.; Kleickmann, T.; Klusmann, U.; Richter, D. The Development of Teachers' Professional Competence. In *Cognitive Activation in the Mathematics Classroom and Professional Competence of Teachers: Results from the COACTIV Project*; Kunter, M., Baumert, J., Blum, W., Klusmann, U., Krauss, S., Neubrand, M., Eds.; Springer: New York, NY, USA, 2013; ISBN 978-1-4614-5148-8.
11. Bruckermann, T.; Ochsen, F.; Mahler, D. Learning opportunities in biology teacher education contribute to understanding of nature of science. *Educ. Sci.* **2018**, *8*, 103. [CrossRef]
12. Hartmann, S.; Upmeyer zu Belzen, A.; Krüger, D.; Pant, H.A. Scientific Reasoning in Higher Education. *Z. Psychol.* **2015**, *223*, 47–53. [CrossRef]
13. Cofré, H.; Núñez, P.; Santibáñez, D.; Pavez, J.M.; Valencia, M.; Vergara, C. A Critical Review of Students' and Teachers' Understandings of Nature of Science. *Sci. Educ.* **2019**, *28*, 205–248. [CrossRef]
14. Reith, M.; Nehring, A. Scientific reasoning and views on the nature of scientific inquiry: Testing a new framework to understand and model epistemic cognition in science. *Int. J. Sci. Educ.* **2020**, *42*, 2716–2741. [CrossRef]
15. Kind, P.; Osborne, J. Styles of scientific reasoning: A cultural rationale for science education? *Sci. Educ.* **2017**, *101*, 8–31. [CrossRef]
16. Kremer, K.; Specht, C.; Urhahne, D.; Mayer, J. The relationship in biology between the nature of science and scientific inquiry. *J. Biol. Educ.* **2013**, *48*, 1–8. [CrossRef]
17. Rönnebeck, S.; Bernholt, S.; Ropohl, M. Searching for a common ground—A literature review of empirical research on scientific inquiry activities. *Stud. Sci. Educ.* **2016**, *52*, 161–197. [CrossRef]
18. Baumert, J.; Kunter, M. The COACTIV Model of Teachers' Professional Competence. In *Cognitive Activation in the Mathematics Classroom and Professional Competence of Teachers: Results from the COACTIV Project*; Kunter, M., Baumert, J., Blum, W., Klusmann, U., Krauss, S., Neubrand, M., Eds.; Springer: New York, NY, USA, 2013; ISBN 978-1-4614-5148-8.
19. Baumert, J.; Kunter, M.; Blum, W.; Brunner, M.; Voss, T.; Jordan, A.; Klusmann, U.; Krauss, S.; Neubrand, M.; Tsai, Y.-M. Teachers' Mathematical Knowledge, Cognitive Activation in the Classroom, and Student Progress. *Am. Educ. Res. J.* **2010**, *47*, 133–180. [CrossRef]
20. Brigham, F.J.; Scruggs, T.E.; Mastropieri, M.A. Teacher Enthusiasm in Learning Disabilities Classrooms: Effects on Learning and Behavior. *Learn. Disabil. Res. Pract.* **1992**, *7*, 68–73.
21. Mahler, D.; Großschedl, J.; Harms, U. Using doubly latent multilevel analysis to elucidate relationships between science teachers' professional knowledge and students' performance. *Int. J. Sci. Educ.* **2017**, *39*, 213–237. [CrossRef]
22. Weinert, F.E. Concept of competence: A conceptual clarification. In *Defining and Selecting Key Competencies*; Rychen, D.S., Salganik, L.H., Eds.; Hogrefe & Huber: Seattle, WA, USA, 2001; pp. 45–65.
23. Shulman, L.S. Those Who Understand: Knowledge Growth in Teaching. *Educ. Res.* **1986**, *15*, 4–14. [CrossRef]
24. Gut-Glanzmann, C.; Mayer, J. Experimentelle Kompetenz. In *Theorien in der Naturwissenschaftsdidaktischen Forschung*; Krüger, D., Parchmann, I., Schecker, H., Eds.; Springer: Berlin/Heidelberg, Germany, 2018; pp. 121–140, ISBN 978-3-662-56319-9.
25. Osborne, J. The 21st century challenge for science education: Assessing scientific reasoning. *Think. Ski. Creat.* **2013**, *10*, 265–279. [CrossRef]
26. Lawson, A.E.; Clark, B.; Cramer-Meldrum, E.; Falconer, K.A.; Sequist, J.M.; Kwon, Y.-J. Development of Scientific Reasoning in College Biology: Do Two Levels of General Hypothesis-Testing Skills Exist? *J. Res. Sci. Teach.* **2000**, *37*, 81–101. [CrossRef]
27. Bao, L.; Cai, T.; Koenig, K.; Fang, K.; Han, J.; Wang, J.; Liu, Q.; Ding, L.; Cui, L.; Luo, Y.; et al. Learning and scientific reasoning. *Science* **2009**, *323*, 586–587. [CrossRef] [PubMed]
28. Mayer, D.; Sodian, B.; Koerber, S.; Schwippert, K. Scientific reasoning in elementary school children: Assessment and relations with cognitive abilities. *Learn. Instr.* **2014**, *29*, 43–55. [CrossRef]
29. Krell, M.; Mathesius, S.; van Driel, J.; Vergara, C.; Krüger, D. Assessing scientific reasoning competencies of pre-service science teachers: Translating a German multiple-choice instrument into English and Spanish. *Int. J. Sci. Educ.* **2020**, *42*, 2819–2841. [CrossRef]
30. Mathesius, S.; Krell, M.; Upmeyer zu Belzen, A.; Krüger, D. Überprüfung eines Tests zum wissenschaftlichen Denken unter Berücksichtigung des Validitätskriteriums relations-to-other-variables. *Z. Päd.* **2019**, *65*, 492–511. [CrossRef]
31. Krüger, D.; Hartmann, S.; Nordmeier, V.; Upmeyer zu Belzen, A. Measuring Scientific Reasoning Competencies. In *Student Learning in German Higher Education: Innovative Measurement Approaches and Research Results*, 1st ed.; Zlatkin-Troitschanskaia, O., Pant, H.A., Toepper, M., Lautenbach, C., Eds.; Springer Fachmedien Wiesbaden: Wiesbaden, Germany, 2020; pp. 261–280. ISBN 9783658278854.
32. Opitz, A.; Heene, M.; Fischer, F. Measuring scientific reasoning—A review of test instruments. *Educ. Res. Eval.* **2017**, *23*, 78–101. [CrossRef]

33. Ding, L.; Wei, X.; Mollohan, K. Does Higher Education Improve Student Scientific Reasoning Skills? *Int. J. Sci. Math. Educ.* **2016**, *14*, 619–634. [CrossRef]
34. Krell, M.; Redman, C.; Mathesius, S.; Krüger, D.; van Driel, J. Assessing pre-service science teachers' scientific reasoning competencies. *Res. Sci. Educ.* **2020**, *50*, 2305–2329. [CrossRef]
35. Khan, S.; Krell, M. Scientific Reasoning Competencies: A Case of Preservice Teacher Education. *Can. J. Sci. Math. Technol. Educ.* **2019**, *19*, 446–464. [CrossRef]
36. Voss, T.; Kleickmann, T.; Kunter, M.; Hachfeld, A. Mathematics Teachers' Beliefs. In *Cognitive Activation in the Mathematics Classroom and Professional Competence of Teachers: Results from the COACTIV Project*; Kunter, M., Baumert, J., Blum, W., Klusmann, U., Krauss, S., Neubrand, M., Eds.; Springer: New York, NY, USA, 2013; ISBN 978-1-4614-5148-8.
37. Schoenfeld, A.H. Explorations of Students' Mathematical Beliefs and Behavior. *J. Res. Math. Educ.* **1989**, *20*, 338–355. [CrossRef]
38. Nouri, N.; Saberi, M.; McComas, W.F.; Mohammadi, M. Proposed Teacher Competencies to Support Effective Nature of Science Instruction: A Meta-Synthesis of the Literature. *J. Sci. Teacher Educ.* **2021**, *32*, 601–624. [CrossRef]
39. Kampurakis, K. The “general aspects” conceptualization as a pragmatic and effective means to introducing students to nature of science. *J. Res. Sci. Teach.* **2016**, *53*, 667–682. [CrossRef]
40. Neumann, I.; Kremer, K. Nature of Science und epistemologische Überzeugungen: Ähnlichkeiten und Unterschiede. *Z. Didakt. Naturwiss.* **2013**, *19*, 209–232.
41. Lederman, N.G.; Abd-El-Khalick, F.; Bell, R.L.; Schwartz, R.S. Views of nature of science questionnaire: Toward valid and meaningful assessment of learners' conceptions of nature of science. *J. Res. Sci. Teach.* **2002**, *39*, 497–521. [CrossRef]
42. Liang, L.L.; Chen, S.; Chen, X.; Kaya, O.N.; Adams, A.D.; Macklin, M.; Ebenezer, J. Assessing preservice elementary teachers' views on the nature of scientific knowledge: A dual-response instrument: A Dual-Response instrument. *Asia-Pac. Forum Sci. Learn. Teach.* **2008**, *9*, 1–20.
43. Liang, L.L.; Chen, S.; Chen, X.; Kaya, O.N.; Adams, A.D.; Macklin, M.; Ebenezer, J. Preservice teachers' views about nature of scientific knowledge development: An international collaborative study. *Int. J. Sci. Math. Educ.* **2009**, *7*, 987–1012. [CrossRef]
44. Neumann, I.; Neumann, K.; Nehm, R. Evaluating Instrument Quality in Science Education: Rasch-based analyses of a Nature of Science test. *Int. J. Sci. Educ.* **2011**, *33*, 1373–1405. [CrossRef]
45. Neumann, K.; Härtig, H.; Harms, U.; Parchmann, I. Science teacher preparation in Germany. In *Model Science Teacher Preparation Programs: An International Comparison of What Works Best*; Pedersen, J., Isozaki, T., Hirano, T., Eds.; Information Age Publishing: Charlotte, NC, USA, 2017; pp. 29–52.
46. Kunz, H. Professionswissen von Lehrkräften der Naturwissenschaften im Kompetenzbereich Erkenntnisgewinnung [Professional Knowledge of Science Teachers in the Competence Area of Scientific Inquiry]. Ph.D. Thesis, University of Kassel, Kassel, Germany, 2012.
47. Duschl, R.A.; Grandy, R. Two Views about Explicitly Teaching Nature of Science. *Sci. Educ.* **2013**, *22*, 2109–2139. [CrossRef]
48. Khishfe, R.; Abd-El-Khalick, F. Influence of explicit and reflective versus implicit inquiry-oriented instruction on sixth graders' views of nature of science. *J. Res. Sci. Teach.* **2002**, *39*, 551–578. [CrossRef]
49. Bruckermann, T.; Aschermann, E.; Bresges, A.; Schlüter, K. Metacognitive and multimedia support of experiments in inquiry learning for science teacher preparation. *Int. J. Sci. Educ.* **2017**, *39*, 701–722. [CrossRef]
50. Krell, M.; Koska, J.; Penning, F.; Krüger, D. Fostering pre-service teachers' views about nature of science: Evaluation of a new STEM curriculum. *Res. Sci. Tech. Educ.* **2015**, *33*, 344–365. [CrossRef]
51. McDonald, C.V. The influence of explicit nature of science and argumentation instruction on preservice primary teachers' views of nature of science. *J. Res. Sci. Teach.* **2010**, *47*, 1137–1164. [CrossRef]
52. Mesci, G.; Schwartz, R.S. Changing Preservice Science Teachers' Views of Nature of Science: Why Some Conceptions May be More Easily Altered than Others. *Res. Sci. Educ.* **2017**, *47*, 329–351. [CrossRef]
53. Dogan, N.; Abd-El-Khalick, F. Turkish grade 10 students' and science teachers' conceptions of nature of science: A national study. *J. Res. Sci. Teach.* **2008**, *45*, 1083–1112. [CrossRef]
54. Nehring, A. Naïve and informed views on the nature of scientific inquiry in large-scale assessments: Two sides of the same coin or different currencies? *J. Res. Sci. Teach.* **2019**, *57*, 510–535. [CrossRef]
55. Khishfe, R. Explicit Nature of Science and Argumentation Instruction in the Context of Socioscientific Issues: An effect on student learning and transfer. *Int. J. Sci. Educ.* **2014**, *36*, 974–1016. [CrossRef]
56. Koenig, K.; Schen, M.; Bao, L. Explicitly Targeting Pre-Service Teacher Scientific Reasoning Abilities and Understanding of Nature of Science through an Introductory Science Course. *Sci. Educ.* **2012**, *21*, 1–9.
57. R Core Team. *R: A Language for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2020.
58. Revelle, W. Psych: Procedures for Psychological, Psychometric, and Personality Research. 2021. Available online: <https://cran.r-project.org/web/packages/psych/index.html> (accessed on 1 August 2021).
59. Muthén, L.K.; Muthén, B.O. *Mplus*; Chapman and Hall/CRC: Los Angeles, CA, USA, 2021; Available online: https://www.statmodel.com/download/usersguide/MplusUserGuideVer_7.pdf (accessed on 1 August 2021).
60. Warm, T.A. Weighted likelihood estimation of ability in item response theory. *Psychometrika* **1989**, *54*, 427–450. [CrossRef]
61. Robitzsch, A.; Kiefer, T.; Wu, M. TAM: Test Analysis Modules. 2020. Available online: <https://cran.r-project.org/web/packages/TAM/index.html> (accessed on 1 August 2021).
62. Gajęcki, A.; Burzykowski, T. *Linear Mixed-Effects Models Using R*; Springer: New York, NY, USA, 2013; ISBN 978-1-4614-3899-1.

63. Nakagawa, S.; Schielzeth, H. A general and simple method for obtaining R² from generalized linear mixed-effects models. *Methods Ecol. Evol.* **2013**, *4*, 133–142. [[CrossRef](#)]
64. Bates, D.; Mächler, M.; Bolker, B.; Walker, S. Fitting Linear Mixed-Effects Models Using lme4. *J. Stat. Soft.* **2015**, *67*, 1–48. [[CrossRef](#)]
65. Rosenthal, R. *Meta-Analytic Procedures for Social Research*; SAGE Publications, Inc.: Thousand Oaks, CA, USA, 1991; ISBN 9780803942462.
66. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*; Lawrence Erlbaum: New York, NY, USA, 1988; ISBN 9781134742707.
67. Pinheiro, J.; Bates, D.; R-core. nlme: Linear and Nonlinear Mixed Effects Models. 2021. Available online: <https://cran.r-project.org/web/packages/nlme/index.html> (accessed on 1 August 2021).
68. Barton, K. MuMIn: Multi-Model Inference. 2020. Available online: <https://r-forge.r-project.org/projects/mumin/> (accessed on 1 August 2021).
69. Lenth, R.V. Emmeans: Estimated Marginal Means, aka Least-Squares Means. 2021. Available online: <https://cran.r-project.org/web/packages/emmeans/index.html> (accessed on 1 August 2021).
70. Ben-Shachar, M.S.; Makowski, D.; Lüdtke, D. Effectsize: Indices of Effect Size and Standardized Parameters. *J. Open Source Softw.* **2021**, *5*. [[CrossRef](#)]
71. Rosseel, Y.; Jorgensen, T.D.; Rockwood, N. lavaan: Latent Variable Analysis. 2021. Available online: <https://cran.r-project.org/web/packages/lavaan/lavaan.pdf> (accessed on 1 August 2021).
72. Herman, B.C.; Clough, M.P. Teachers' longitudinal NOS understanding after having completed a science teacher education program. *Int. J. Sci. Math. Educ.* **2016**, *14*, 207–227. [[CrossRef](#)]
73. Kunina-Habenicht, O.; Schulze-Stocker, F.; Kunter, M.; Baumert, J.; Leutner, D.; Förster, D.; Lohse-Bossenz, H.; Terhart, E. Die Bedeutung der Lerngelegenheiten im Lehramtsstudium und deren individuelle Nutzung für den Aufbau des bildungswissenschaftlichen Wissens [The significance of learning opportunities in teacher training courses and their individual use for the development of educational-scientific knowledge]. *Z. Päd.* **2013**, *59*, 1–23.
74. Mathesius, S.; Hartmann, S.; Upmeier zu Belzen, A.; Krüger, D. Scientific Reasoning as an Aspect of Pre-service Biology Teacher Education. Assessing competencies using a paper-pencil test. In *The Future of Biology Education Research*; Tal, T., Yarden, A., Eds.; Technion—Israel Institute of Technology: Haifa, Israel, 2016; pp. 93–110.
75. Kuhn, T.S. *The Structure of Scientific Revolutions*, 2nd ed.; The University of Chicago Press: Chicago, IL, USA, 1970; ISBN 0-226-45803-2.
76. Gyllenpalm, J.; Wickman, P.-O. The Uses of the Term Hypothesis and the Inquiry Emphasis Conflation in Science Teacher Education. *Int. J. Sci. Educ.* **2011**, *33*, 1993–2015. [[CrossRef](#)]
77. Edgerly, H.S.; Kruse, J.W.; Wilcox, J.L. Quantitatively Investigating Inservice Elementary Teachers' Nature of Science Views. *Res. Sci. Educ.* **2021**, 1–14. [[CrossRef](#)]
78. Mathesius, S.; Upmeier zu Belzen, A.; Krüger, D. Eyetracking als Methode zur Untersuchung von Multiple-Choice-Aufgaben zum wissenschaftlichen Denken [Eye tracking as a method for studying multiple-choice scientific reasoning tasks]. In *Lehr- und Lernforschung in der Biologiedidaktik*; Hammann, M., Lindner, M., Eds.; StudienVerlag: Innsbruck, Austria, 2018; pp. 225–244.
79. Marton, F.; Tsui, A.B.M. *Classroom Discourse and the Space of Learning*; Lawrence Erlbaum: Mahwah, NJ, USA, 2004; ISBN 0805840087.
80. Hodson, D. Learning Science, Learning about Science, Doing Science: Different goals demand different learning methods. *Int. J. Sci. Educ.* **2014**, *36*, 2534–2553. [[CrossRef](#)]