Research in Science Education (2023) 53:59–79
https://doi.org/10.1007/s11165-022-10045-x

Pre-Service Science Teachers' Scientific Reasoning Competencies: Analysing the Impact of Contributing Factors

Moritz Krell[1] · Samia Khan[2] · Claudia Vergara[3] · Hernán Cofré[4] · Sabrina Mathesius[5] · Dirk Krüger[5]

Accepted: 16 January 2022 / Published online: 25 January 2022

Abstract

Scientific reasoning competencies (SRC) are one part of science teachers' professional competencies. This study examines the contribution of three factors to the development of pre-service science teachers' SRC: the *amount of science education classes*, the *amount of science classes* and the pre-service science teachers' *age*. The factors *amount of science education classes* and *amount of science classes* have been operationalised in terms of ECTS credit points. $N = 438$ pre-service science teachers from six universities in Germany, Chile and Canada voluntarily and anonymously responded to an established multiple-choice instrument for assessing SRC, which has been developed by the authors and is available in German, Spanish and English. Multiple linear regression analyses show that the included factors explain a proportion of about 9% of the pre-service science teachers' SRC. The factor *amount of science classes* is the only significant predictor and can be seen as an indicator of learning science content knowledge. These findings support the assumption of science content knowledge being a prerequisite for developing pre-service science teachers' SRC.

Keywords Pre-service science teachers · Scientific reasoning competencies · Contributing factors · International cooperation

Introduction

One stated goal of teacher education at university level is to equip future teachers with the competencies needed to plan lessons as well as to teach and reflect upon the teaching–learning processes professionally (Carlson & Daehler, 2019). The knowledge, skills and motivational orientations which pre-service science teachers are asked to develop during teacher education are suggested to encompass professional knowledge, motivational orientations and self-regulation, as well as beliefs, values and goals related to teaching and learning (Baumert & Kunter, 2013). With reference to the seminal work by Shulman (1986), professional knowledge (including practical knowledge and skills) can be further

✉ Moritz Krell
krell@leibniz-ipn.de

Extended author information available on the last page of the article subdivided into content knowledge (CK), pedagogical content knowledge (PCK), pedagogical knowledge (PK) and further knowledge dimensions (e.g. curricular knowledge). For pre-service science teachers, CK includes science content knowledge, procedural knowledge and epistemic knowledge related to science (Bybee, 2014; Großschedl et al., 2015). The knowledge and skills needed for scientific problem-solving and reasoning in science are seen as '[a] necessary element of any competent teacher of science' (Osborne, 2014, 189).

Several studies investigate the development of pre-service science teachers' knowledge, skills or competencies related to scientific reasoning (e.g. Hartmann et al., 2015; Mahler et al., 2021; Stammen et al., 2018). These studies provide suggestions about ways to foster scientific reasoning competencies (SRC) in science teacher education and identify specific factors that significantly contribute to competence development in this domain; however, most of the studies available so far investigate the contribution of such factors within specific contexts, such as single universities with specific study regulation programmes and curricula, or specific interventions (e.g. Khan & Krell, 2019; Stammen et al., 2018). The specificities of programme regulations and the resulting confounding of variables—such as pre-service science teachers' age and cumulative learning opportunities over time—make analyses of the contribution of specific factors challenging to generalise beyond the specific contexts of the studies. For example, university teacher education programmes can be organised concurrently (with disciplinary and pedagogical studies within the same programme) or consecutively (with pedagogical studies following a pure disciplinary study programme) (Cofré et al., 2022; Zuzovsky & Donitsa-Schmidt, 2017).

This study examines the contribution of three of the proposed factors suggested in the literature (Hartmann et al., 2015; Limueco & Prudente, 2018; Schwichow & Nehring, 2018) on the development of pre-service science teachers' SRC: the *amount of science education classes* (as an indicator for explicit reflections on scientific reasoning), the *amount of science classes* (as an indicator for learning science content knowledge) and the pre-service science teachers' *age*. Other factors, such as brain growth (e.g. Kwon & Lawson, 2000) or general cognitive abilities (e.g. Göhner & Krell, 2022; Mayer et al., 2014) are not addressed in this study. Hereby, this study makes use of a sample of pre-service science teachers from six different universities and three different countries to assist in identifying the contribution of the three factors independent from specific study programmes.

Conceptualising Science Teachers' Scientific Reasoning Competencies

It is desirable that science teachers develop their competencies at reasoning scientifically (Bybee, 2014; Osborne, 2014). In line with this contention, science teacher education standard documents in various countries require that pre-service science teachers should have—next to an advanced content knowledge of their subject domain—elaborate procedural and epistemic scientific knowledge in order to reason scientifically (e.g. Chile: Mineduc, 2012; Germany: KMK, 2019). Other science education documents require science teachers to implement scientific inquiry approaches in their lessons, including formulating research questions, developing models, designing experiments and evaluating evidence as tasks for students, in order to develop students' competencies in reasoning well about scientific problems (e.g. BCMOE, 2019; NGSS Lead States, 2013).

In science education research, SRC are defined as the dispositions to be able to solve a scientific problem in a certain situation by applying a set of scientific skills and knowledge. SRC are further understood as a latent, complex construct that also encompasses the capacity to reflect about the problem-solving process at a meta-level (Krell et al., 2018; Lawson, 2004; Morris et al., 2012). This conceptualization of SRC is quite established in science education research. In a review study, Opitz et al. (2017) identified three aspects in which conceptualizations of SRC may differ: (1) the specific skills they include; (2) if scientific reasoning is conceptualised as a general, uniform competence or rather as a more differentiated set of skills and abilities; and (3) if scientific reasoning is assumed to be domain-general or domain-specific.

Different generic, cognitive operations involved in scientific reasoning have been identified, including encoding (process of representing information and its context in memory), retrieval of information and strategy development (e.g. Kuhn, 1989; Kuhn & Pearsall, 2000; Morris et al., 2012). Science content knowledge, procedural and epistemic knowledge related to specific 'styles of scientific reasoning' (e.g. experimentation and modelling) have also been suggested as another important element of SRC (Osborne, 2013). Hence, the dispositions to be able to solve a scientific problem in a certain situation (i.e. SRC) include 'not just a knowledge of its [scientific] domain-specific constructs but also a knowledge of a set of procedural and epistemic constructs' (Kind & Osborne, 2017, 10). As shown above, these dispositions—although not always explicitly named so—are suggested in many science teacher education standard documents as an outcome of a good science teacher education (e.g. KMK, 2019; Mineduc, 2012). Empirical studies suggest that pre-service science teachers' SRC develop over their university studies (e.g. Krüger et al., 2020) or as a result of specifically designed interventions (e.g. Stammen et al., 2018). However, the studies conducted so far have concentrated on the development of pre-service science teachers' SRC within individual study programmes.

For the operationalisation of SRC in this study, a theoretical framework is employed which covers two dimensions named *conducting scientific investigations* and *using scientific models* that are subdivided into seven related skills of scientific reasoning (Table 1). In this framework, the dimension *conducting scientific investigations* covers processes commonly associated with a hypothetic-deductive approach of experimentation, whilst *using scientific models* is related to modelling or the use of models as epistemic tools (Hartmann et al., 2015).

Table 1 Scientific reasoning competencies and associated dimensions of *conducting scientific investigations* and *using scientific models*, suggested by Hartmann et al. (2015)

Scientific reasoning competencies		
Dimensions	Conducting scientific investigations	Using scientific models
Skills	Formulating questions Generating hypotheses Planning investigations Analysing data and drawing conclusions	Judging the purpose of models Testing models Changing models

Factors Contributing to the Development of Science Teachers' SRC

The development of SRC has been conceptualised as a stage-like process, including '[a] growing awareness (i.e., consciousness) of one's reasoning patterns and one's reflectivity as well increases in the contexts to which the patterns can be applied' (Lawson, 2004, 323). Other perspectives emphasise the situated nature of scientific reasoning more strongly. For example, the 'epistemology in practice' perspective assumes that people have various resources (e.g. knowledge, skills), which are combined for sense-making depending on the specific context (Berland et al., 2016). Independent from the theoretical perspective, there is a consensus that both domain-specific knowledge and reasoning skills are needed for successful scientific reasoning (Shavelson, 2018). Hence, content knowledge, procedural knowledge and epistemic knowledge are conceptualised as the three constituent dimensions of SRC (Osborne, 2013).

In most science teacher education programmes, mainly two kinds of classes exist: *discipline-specific science classes* and *science education classes* (the latter also called science methods classes/courses). Discipline-specific science classes typically aim at developing pre-service science teachers' content knowledge and procedural knowledge, for example in biology: knowledge of microbiology, physiology and behavioural biology as well as knowledge of how to use a microscope, how to plan and conduct a laboratory experiment and how to scientifically observe animals' behaviour. Science education classes aim to foster pre-service science teachers' PCK, including, for example, knowledge of students' understanding in science and knowledge of instructional strategies for teaching science (Park & Oliver, 2008). As many science teacher education standard documents include SRC, science education classes typically also address pre-service science teachers' procedural and epistemic knowledge as important parts of CK (Neumann et al., 2017, 2019; Stammen et al., 2018). It is acknowledged that often all three types of knowledge—content knowledge, procedural knowledge and epistemic knowledge—are intertwined and addressed together. It has been suggested, however, that a specific contribution of science classes is the development of pre-service science teachers' content knowledge, whilst epistemic knowledge is emphasised more within science education classes (Hartmann et al., 2015; Neumann et al., 2019). This focus on specific knowledge types in science and science education classes is illustrated in Fig. 1. Based on this, it is clear that both kinds of classes make valuable contributions to the development of pre-service science teachers' SRC. Psychological and science education research has proposed several factors that might specifically contribute to the development of pre-service science teachers' SRC (e.g. Engelmann et al., 2016; Hartmann et al., 2015; Lawson, 2004). In the following sections, three factors that have been suggested in the literature as contributing to the development of pre-service science teachers' SRC will be explained in more detail: (A) the *amount of science education classes*, (B) the *amount of science classes* and (C) the *age* of pre-service science teachers.

A) Science education classes are likely to contribute to the development of SRC

Many authors emphasise that explicit reflections about science significantly contribute to the development of SRC (e.g. Hartmann et al., 2015; Khan & Krell, 2019). Explicit reflection can be provided by, for example, reflecting upon basic features of science (e.g. intersubjectivity and justification in science), discussing basic science terminology (e.g. hypothesis, experiment and theory) or analysing current as well as historical approaches

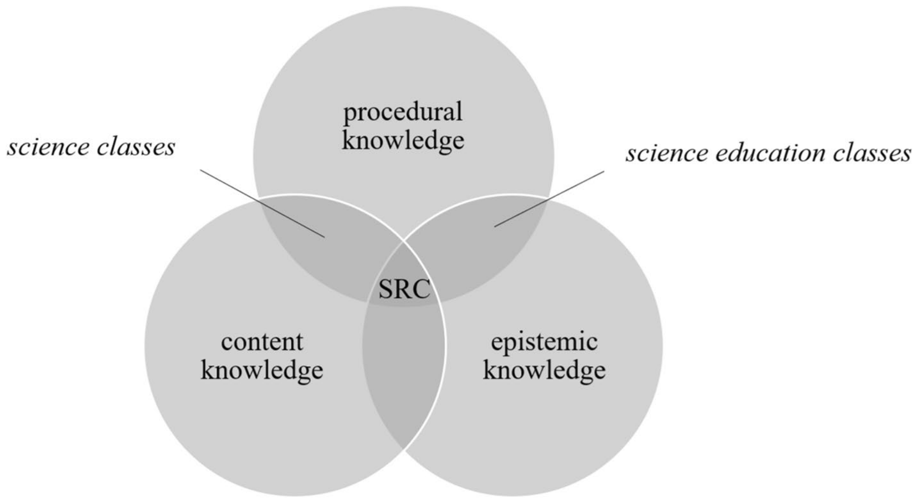

Fig. 1 Scientific reasoning competencies (SRC) are illustrated as an amalgam of content knowledge, procedural knowledge and epistemic knowledge (Osborne, 2018). As emphasised in the text, science classes in teacher education programmes typically aim to develop content and procedural knowledge whilst science education classes target on procedural and epistemic knowledge (and further dimensions of teachers' professional knowledge such as PCK)

of scientists (e.g. Copernicus, Curie and Darwin) (Krell et al., 2015). Explicit reflection is especially discussed in research on the development of epistemic knowledge (e.g. knowledge about nature of science; Duschl & Grandy, 2013), where various studies reveal that explicit-, argumentative- and inquiry-based approaches are effective in promoting participants' epistemic knowledge of science (e.g. Krell et al., 2015). Bruckermann et al. (2018) show that the amount of such opportunities to reflect about science positively contributes to pre-service science teachers' epistemic knowledge.

Other authors further suggest that explicit reflections that are helpful in developing epistemic knowledge about science especially take place within science teacher education classes, where pre-service science teachers focus on both scientific reasoning and how to teach reasoning to students (e.g. Hartmann et al., 2015; Mathesius et al., 2016). For example, Hartmann et al. (2015) used the instrument to assess SRC, which is also applied in this study (see below), and show in a cross-sectional comparative study that pre-service science teachers outperform even science students on this instrument. These findings were hypothetically explained with 'the positive effect of learning opportunities to explicitly discuss and reflect on the inquiry process' (Hartmann et al., 2015, 49), that took place in science teacher education classes. Stammen et al. (2018) propose that science teacher education, combining content (i.e. scientific reasoning) and pedagogy (i.e. teaching SRC), is effective in fostering pre-service science teachers' competencies at scientific reasoning. Relatedly, it can be assumed that the *amount of science education classes* (e.g. quantified in terms of ECTS credits, see below) during teacher education is a potential predictor of science teachers' SRC.

B) Science classes are likely to contribute to the development of SRC

Several scholars also emphasise the importance of science content knowledge as a prerequisite for developing competencies of scientific reasoning (e.g. Fischer et al., 2014). For example, Ruppert et al. (2017) discuss that content knowledge plays a significant role in students' modelling. The authors show that domain-specific content knowledge about a phenomenon can be a prerequisite for the extent to which middle school students' models accounted for the provided evidence. They propose, for example, that students may be better able to apply a domain-general heuristic (i.e. accounting for evidence), if they have the necessary content knowledge to identify the relevant evidence for the problem at hand. Samarapungavan (2018) emphasises that correctly analysing scientific evidence requires a solid understanding of both theoretical and methodological concepts as well as science practices. For the case of argumentation in science education, Sadler and Fowler (2006) propose the 'Threshold Model of Content Knowledge Transfer'. This model assumes a non-linear relationship between argumentation and the level of content knowledge, with specific thresholds allowing people to argue in a more advanced way. Similarly, Schwichow and Nehring (2018) suggest, with a focus on secondary school students' experimental competencies, that science content knowledge might even be a prerequisite for the development of SRC. Based on these propositions, it can be assumed that the *amount of science classes* (e.g. quantified in terms of ECTS credits, see below) during teacher education is a potential predictor of science teachers' SRC.

C) Age and intellectual development are likely to contribute to the development of SRC

Several authors relate SRC development to stages of intellectual development and age. For example, Lawson (2004) proposes with his focus on hypothetic-deductive reasoning that intellectual development involves a growing awareness of reasoning patterns and reflectivity. For this development to occur, an individual needs to be confronted with situations in which the reasoning is not fruitful and, hence, causes the development of new rules or strategies to guide one's future behaviour ('challenging situations'). Depending on the external stimulation, that is, the nature of situations and feedback an individual is confronted with, development of SRC may also be fostered purposefully by designed teaching approaches and interventions (Lawson, 2004). Similarly, limiting the development thesis, Kuhn and Pearsall (2000) state that scientific thinking skills have their origins in young children's abilities to distinguish between theory and evidence, but these skills may not even be fully developed by adulthood (Zimmermann, 2007). Woolley et al. (2018) identified ten common false strategies in solving scientific reasoning problems (e.g. confusing the independent with the dependent variable in an experiment) in a sample of undergraduates. These problems are similar to problems that have been identified for younger students and children (Hammann et al., 2008; Zimmermann, 2007). Kuhn (1989) emphasised that for students from 9th grade up to adults, the level of SRC is strongly influenced by educational level.

In line with these assumptions, several other scholars suggest that university students' SRC increase over their course of studies due to the students' increasing age and related intellectual development (e.g. Limueco & Prudente, 2018); however, other studies do not affirm this finding and have found no effect of formal education on scientific reasoning (e.g. Ding, 2017; Ding et al., 2016). These ambiguous findings suggest that it may depend variably on specific emphases in the curricula (e.g. *amount of science classes* and *amount of science education classes*) or other factors, for tertiary education to positively contribute to the development of students' SRC. One challenge of assessing the contribution of age and intellectual development on SRC is that, typically, other important factors (e.g. *amount of science classes* and *amount of science education classes*) increase with increasing age. Hence, it is often difficult to determine the impact of age due to its correlation with other factors. Alternative perspectives on the development of SRC, such as epistemologies in practice (Berland et al., 2016), might also explain why there is no linear relationship between pre-service science teachers' age or educational level and their SRC.

Aim of the Study and Hypotheses

The prior research on SRC summarised above informs about ways to foster pre-service science teachers' SRC (e.g. focusing on science content and science education/pedagogy subsequently or concurrently in their programme) and provides avenues for further research. The aim of the present study is to evaluate three of the proposed factors (*amount of science education classes*, *amount of sciences classes* and *age*) as predictors of pre-service science teachers' SRC. As suggested in the literature, the *amount of science education classes* is used as an indicator for explicit reflections on scientific reasoning (i.e. fostering epistemic knowledge; Hartmann et al., 2015). The *amount of science classes* is used as an indicator for learning content knowledge (Neumann et al., 2017, 2019). The following hypotheses are tested and discussed within this study:

H1: The *amount of science education classes* during higher education positively predicts pre-service science teachers' SRC (see Bruckermann et al., 2018; Hartmann et al., 2015).
H2: The *amount of science classes* during higher education positively predicts pre-service science teachers' SRC (see Ruppert et al., 2017; Schwichow & Nehring, 2018).
H3: The pre-service science teachers' *age* positively predicts their SRC (see Kuhn & Pearsall, 2000; Limueco & Prudente, 2018; Lawson, 2004).

Methods

The hypotheses are tested within a quantitative methodological framework, applying item response theory models (Bond & Fox, 2001) and classical test theory as well (Field, 2009). Data on pre-service science teachers' SRC are collected by means of a paper–pencil instrument and the proposed factors (H1–H3) are tested as predictor variables in multiple linear regression analyses to explain differences in pre-service science teachers' SRC.

Sample

The sample includes $N=438$ pre-service science teachers from Germany ($n=219$; freshmen to 10th semester; $M_{AGE}=24$; $SD_{AGE}=4.37$; from one university), Chile ($n=118$; 1st to 9th semester; $M_{AGE}=22$; $SD_{AGE}=3.87$; from four different universities) and Canada ($n=101$; 8th and 9th semester; $M_{AGE}=27$; $SD_{AGE}=5.67$; from one university).

The German and Chilean samples stem from concurrent programmes: Bachelor of Science/Arts with a subsequent Master of Education programme (Germany) and Bachelor of

Education programme (Chile), respectively. The Canadian sample is drawn from a 1-year post-graduate Bachelor of Education programme (i.e. consecutive programme).

Only pre-service science teachers without missing responses are considered in the sample. The data partly comes from existing studies (Khan & Krell, 2019; Krell et al., 2018, 2020; Krüger et al., 2020) and is secondarily analysed for this study.

Variables

A multiple-choice instrument to assess SRC, available in German (original version), Spanish and English (translated versions), was applied in this study (Krell et al., 2018, 2020; Mathesius et al., 2016). The instrument is based on a theoretical framework covering seven skills of reasoning in science (Table 1). For each skill, the instrument includes three multiple-choice items (i.e. 21 items in total). Each item is contextualised within an authentic scientific context and the respondents must apply their procedural and epistemic knowledge within the context to identify the attractor (=correct answering option). For example, related to the skill formulating questions, respondents need to know that scientific research questions are related to a phenomenon, empirically testable, intersubjectively comprehensible, unambiguous, principally answerable and internally and externally consistent (Mathesius et al., 2014). For sample items see Krell et al. (2018, 2020); the full instrument is available upon request to the first author.

The original German version of the instrument was extensively evaluated following the recommendations in the *Standards for Educational and Psychological Testing* (AERA et al., 2014), that is, considering various sources of evidence for the valid interpretation of the test scores as measures of SRC. This process of instrument development and evaluation is summarised by Krüger et al. (2020). The quality and equivalence of the three language versions were evaluated by applying the *translation, review, adjudication, pretesting and documentation-* approach, an established approach for questionnaire translation (Harkness, 2003). This systematic translation, including a pretesting of the translated instruments using qualitative and quantitative methods, contributed to test equivalence (Krell et al., 2018, 2020).

To be able to test H1 and H2, the factors *amount of science education classes* (H1) and *amount of science classes* (H2) have been operationalised as workload, in terms of ECTS credits as prescribed in the respective study regulation documents, where 'one credit corresponds to 25 to 30 h of work' (EU, 2015, 10). These credits provide a sound basis that allows a cross-country comparison. Table 2 provides this data for the sample analysed in the present study. Note that the German pre-service science teachers are studying two subjects and that the workload of science education classes and science classes differs depending on these study subjects. The Chilean universities in the sample have different study programmes.

In Table 2, the different study programmes are organised along the distinction between concurrent and subsequent study programmes. In the present sample, five study programmes have been identified, in which pre-service teachers are enrolled in science education classes within the first two years already (i.e. as freshmen or sophomores), and four study programmes, in which science education classes are part of the curriculum not earlier than in the third year (Table 2).

The sampling of different universities is advantageous having a variety of combinations of the respective variables in the dataset, whilst for single universities, the considered factors might be highly confounded.

Table 2 Workload of science (Sc.) and science education (Sc.Ed.) classes for each semester during the course of studies in the analysed samples (in terms of ECTS credits) and the respective group sizes (n) in each semester

	Semester	0	1	2	3	4	5	6	7	8	9	10
	Programmes with science education classes starting within the first two years ($n = 209$)											
Germany; bio, che ($n = 24$)	Sc		27	22	26	27	25	13	25	10		
	Sc.Ed				3.5	3.5	3.5	3.5		10	24	10
	n	2		3		4		2		10		3
Germany; che, bio ($n = 14$)	Sc		27	22	27	29	25	10	25	10		
	Sc.Ed				3.5	3.5	3.5	3.5		10	24	10
	n	3						2		8		1
Germany; bio, noSci ($n = 96$)	Sc		14	14	12	17	15	8	10	5		
	Sc.Ed				3.5	3.5				5	12	5
	n	14		29		6		7		37		3
Chile; university 2 ($n = 63$)	Sc		16.5	24	22.5	15	12	19.5	10.5	10.5	4.5	–-
	Sc.Ed				4.5		4.5			6		–-
	n		1	21	3	9	13	9	2	3	2	–-
Chile; university 4 ($n = 12$)	Sc		4.79	10.79	20.40	15.59	13.79	17.99	17.99	23.99	11.99	–-
	Sc.Ed		4.79	7.80	7.80	2.99	6	6	6	6		–-
	n		1	10		1						–-
	Programmes with science education classes starting not earlier than in the third year ($n = 229$)											
Chile; university 1 ($n = 32$)	Sc		16.67	19.99	29.99	23.33	23.33	21.67	19.99	19.99		
	Sc.Ed						8.33		8.33			
	n			6		19	2	5				
Chile; university 3 ($n = 11$)	Sc		17.99	17.99	17.99	17.99	17.99	17.99			–-	–-
	Sc.Ed							6	17.99	6	–-	–-
	n			2		3	1	3	1	1	–-	–-

Table 2 (continued)

	Semester	0	1	2	3	4	5	6	7	8	9	10
Germany; noSci, bio ($n=85$)	Sc		14	7	10	14	10	5	10	10		
	Sc.Ed						3.5	3.5		5	12	5
	n	7		33		14		4		26		1
Canada; university 1 ($n=101$)	Sc		19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5		
	Sc.Ed										8.23	
	n									48	53	

Two subjects have to be studied in Germany, i.e. either one science or two sciences; Bio: biology; Che: chemistry; noSci: no scientific subject; empty boxes: 0; –-: end of study programme

Data analysis

Person abilities and item difficulties at the latent level have been estimated on the One-Parameter Logistic Model (1PLM) for dichotomous items ('Rasch model') using the software ACER Conquest (Wu et al., 2007). Weighted maximum likelihood estimates (WLE) were used as point estimates for person abilities (Wu et al., 2007). Higher WLE indicate better performance on the test. As done in the related previous studies, a one-dimensional Rasch model was specified, reflecting the view that the described skills are mainly used in both considered processes of scientific reasoning and, hence, the two processes *conducting scientific investigations* and *using scientific models* (Table 1) are not clearly separable.

For the evaluation of model fit, the sum of squared standardised residuals (MNSQ) is proposed (Wu et al., 2007), which has an expected value of 1 with acceptable values ranging from 0.8 to 1.2 (Bond & Fox, 2001). Because the unweighted MNSQ is more sensitive to outliers than the weighted MNSQ, both statistics should be considered. In addition, *t*-standardised fit statistics based on the MNSQ are provided, which should range from −2 to 2 (Wu et al., 2007).

Next to the fit indices, differential item functioning (DIF) for the three language versions of the instrument was evaluated. In this study, the Mantel–Haenszel (MH) statistic was computed for the analysis of DIF, which is appropriate to compare the probability of a correct answer on an item between different groups, depending on their overall test performance (Zwick et al., 1999).

Given that the fit between data and model could be shown as satisfactory and that there is no substantial DIF, the relation between the pre-service science teachers' WLE (as estimates for SRC) and the proposed factors (H1–H3) can be statistically analysed to test the hypotheses. For this reason, correlational and multiple linear regression analyses have been conducted. Furthermore, the kind of study programme (i.e. science education classes starting early or late; Table 2) was also considered as a predictor variable to account for the hierarchical data structure with pre-service science teachers enrolled in different programmes. The more common approach to analyse hierarchical data, which is hierarchical linear modelling, was not applicable in this study since it demands at least five random-effects levels for each factor (Hodges, 2016). Correlational and multiple linear regression analyses were done with IBM SPSS Statistics, Version 26.

Results

Evaluating Model Fit of the Rasch Model

The fit statistics suggest an acceptable fit between the data and the model for most items; however, one item related to the skill *generating hypotheses* shows a weighted *t*-statistic of 3.5. This item has been excluded from the analysis. The DIF analysis reveals a 'moderate to large' DIF for two more items (MH statistic > 1.5), which, therefore, have been excluded as well. Hence, the present study is based on 18 of the instrument's 21 items. For these 18 items, the fit statistics propose a good fit. The reliability measure ($rel._{EAP/PV} = 0.51$) is acceptable (Table 3) and similar to what has been reported in previous studies with the instrument (e.g. $rel._{EAP/PV} = 0.55$, Krell et al., 2018).

Table 3 Reliability coefficients and item fit statistics for the full and the reduced instrument

		Full instrument (21 items)		Reduced instrument (18 items)	
		$rel._{eap/pv} = 0.52$		$rel._{eap/pv} = 0.51$	
	Acceptance range	Unweighted	Weighted	Unweighted	Weighted
MNSQ	0.8–1.2	0.89–1.06	0.95–1.08	0.93–1.05	0.96–1.04
T	$\|t\| < 2$	$\|t\| \leq 1.6$	$\|t\| \leq 3.5$	$\|t\| \leq 1.0$	$\|t\| \leq 1.4$

Basic Data Analysis

The mean estimated person ability of the total sample is $M_{\mathrm{WLE}} = -0.12$ ($SD_{\mathrm{WLE}} = 0.74$), with the Chilean pre-service science teachers' scoring significantly lower ($M_{\mathrm{WLE}} = -0.34$; $SD_{\mathrm{WLE}} = 0.66$) than the Canadian ($M_{\mathrm{WLE}} = 0.18$; $SD_{\mathrm{WLE}} = 0.64$; $p < 0.001$) and the German ($M_{\mathrm{WLE}} = -0.13$; $SD_{\mathrm{WLE}} = 0.77$; $p < 0.01$) pre-service science teachers. There is a significant ($p = 0.001$) difference between the German and the Canadian pre-service science teachers' WLE as well (ANOVA with Bonferroni post-hoc test). The relatively high WLE of the Canadian pre-service science teachers might be a result of their rather high study semesters (> 7 semesters) and related opportunities to reflect and learn the respective content (Table 2). There is no significant difference in the WLE of pre-service science teachers from study programmes with science education classes starting early ($M_{\mathrm{WLE}} = -0.17$; $SD_{\mathrm{WLE}} = 0.76$) or late in the curriculum ($M_{\mathrm{WLE}} = -0.07$; $SD_{\mathrm{WLE}} = 0.72$; $p = 0.12$; independent *t*-test).

For the total sample, there is a significant bivariate (Pearson) correlation between the pre-service science teachers' semester and the estimated WLE ($r = 0.31$; $p < 0.001$); this was also found for the German subsample ($r = 0.28$; $p < 0.001$) but not for the Canadian and Chilean subsamples. This indicates, for the total sample and the German subsample, a significant linear progression of SRC over the course of the pre-service science teachers' studies. The Chilean pre-service science teachers show very similar WLE in the different semesters with a peak in the 7th semester, which is, however, only based on $n = 3$. The Canadian pre-service science teachers' SRC slightly decrease (though not significantly) from semester 8 to 9. This non-significant decrease in the tail-end can also be found for the German and Chilean pre-service science teachers (Table 4; Fig. 2).

Testing the Hypotheses

For the total sample, there are significant positive (bivariate) correlations between the estimated WLE and the *amount of science education classes*, the *amount of science classes* and the pre-service science teachers' *age*, and very similar values could be found when distinguishing between the two kinds of study programmes as well (Table 5).

These findings suggest a positive relation between each of the three factors and the science teachers' SRC. This interpretation is limited by the fact that there are significant correlations between these variables as well (Table 6), due to the respective study programme regulations (Table 2). Consequently, for the total sample, the partial correlations between the WLE and the *amount of science education classes* ($r_p = 0.04$; $p = 0.419$) and the *amount of science classes* ($r_p = 0.27$; $p < 0.001$) are smaller than the bivariate correlations shown in

Table 4 Estimated WLE ($M_{WLE} \pm 2 \cdot SE$) of the pre-service science teachers in the single semesters, for the German (GER), Chilean (CHI) and Canadian (CAN) subsamples. Note that the data is cross-sectional and does not provide information about individual pre-service science teachers' development of SRC. See Table 2 for further information about the semester groups. SEM: semester

SEM	0	1	2	3	4	5	6	7	8	9	10
GER	−0.20 ± 0.27		−0.42 ± 0.18		−0.34 ± 0.23		0.14 ± 0.42		0.14 ± 0.17		0.03 ± 0.62
	n = 26		*n* = 65		*n* = 24		*n* = 15		*n* = 81		*n* = 8*
CHI		−1.02 ± 0.58	−0.38 ± 0.17	−0.77 ± 1.50	−0.33 ± 0.26	−0.24 ± 0.36	−0.40 ± 35	0.59 ± 0.45	−0.29 ± 0.36	−0.35 ± 0.25	
		n = 2*	*n* = 39	*n* = 3*	*n* = 32	*n* = 16	*n* = 17	*n* = 3*	*n* = 4*	*n* = 2*	
CAN									0.24 ± 0.18	0.12 ± 0.18	
									n = 48	*n* = 53	

Values marked with an asterisk (*) are based on small numbers ($n < 15$) and should be interpreted cautiously

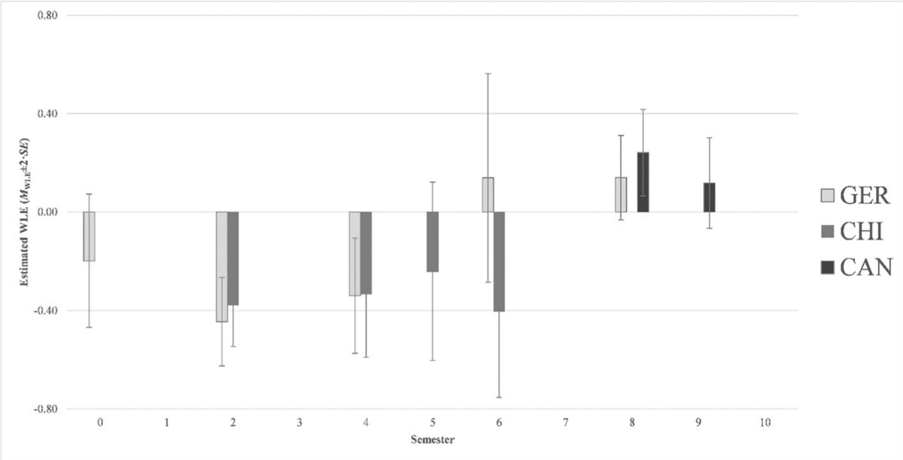

Fig. 2 Estimated WLE ($M_{WLE} \pm 2 \bullet SE$) of the pre-service science teachers in the single semesters, for the German, Chilean and Canadian sub-samples. Subgroups with $n < 15$ are not shown (see Table 4). Note that the data is cross-sectional and does not provide information about individual pre-service science teachers' development of SRC. See Table 2 for further information about the semester groups

Table 5. These findings suggest only the factor *amount of science classes* (H2) as positively contributing to the pre-service science teachers' WLE.

A multiple linear regression analysis with the pre-service science teachers' WLE and the proposed factors (H1–H3) shows no violations of assumptions; the Durban-Watson-statistic is 1.89, indicating that there is no considerable autocorrelation, and VIF is < 10 for all items, indicating no serious multicollinearity (Field, 2009). The adjusted R^2 illustrates that the three factors explain a proportion of about 9–10% of the pre-service science teachers' SRC. Only the *amount of science classes* was found to be a statistically significant predictor, with a standardised $\beta = 0.277$ and a partial $r^2 = 0.057$ (Table 7; model 1).

If the kind of study programme (Table 2) is included as an additional predictor variable in the regression model (Durban-Watson-statistic = 1.88), the adjusted R^2 stays almost the same (Table 7; model 2). In this analysis, the significant positive effect of the *amount of science classes* can still be found with almost the same size of effect (i.e. standardised β) and a slightly smaller r^2_p (Table 7; model 2).

Summarising, the findings of the multiple linear regression analyses supported the conclusion that the *amount of science classes* positively contributes to the pre-service science teachers' SRC, even if the kind of study programme is considered (Table 7).

Discussion

This study investigated the *amount of science education classes* during higher education (see Bruckermann et al., 2018; Hartmann et al., 2015), the *amount of science classes* during higher education (see Ruppert et al., 2017; Schwichow & Nehring, 2018) and the pre-service science teachers' *age* (see Kuhn & Pearsall, 2000; Lawson, 2004; Limueco & Prudente, 2018) as predictors of their SRC.

Due to the theoretical framework applied for item development in this study (Table 1), the SRC items are related to the dimensions *conducting scientific investigations* and *using*

Table 5 Pearson correlations between the estimated WLE and workload in science classes and science education classes and the pre-service science teachers' age, specific for the two kinds of study programmes and the total sample

	Science	*Science education*	*Age*
Science education starting early ($n=209$)	0.28 (***)	0.20 (**)	0.15 (*)
Science education starting late ($n=229$)	0.33 (***)	0.21 (**)	0.17 (*)
Total sample ($N=438$)	0.31 (***)	0.17 (***)	0.17 (**)

Pearson correlation coefficient and *p*-value (*: $p<0.05$; **: $p<0.01$; ***: $p<0.001$)

Table 6 Pearson correlations between the workload in *science classes* and *science education classes* and the pre-service science teachers' age , specific for the two kinds of study programmes and the total sample

		Science education	*Age*
Science education starting early ($n=209$)	*Science*	0.77 (***)	0.39 (***)
	Science education	—	0.33 (***)
Science education starting late ($n=229$)	*Science*	0.26 (***)	0.38 (***)
	Science education	—	0.18 (**)
Total sample ($N=438$)	*Science*	0.43 (***)	0.40 (***)
	Science education	—	0.19 (***)

Pearson correlation coefficient and *p*-value (*: $p<0.05$; **: $p<0.01$; ***: $p<0.001$)

Table 7 Coefficients of the multiple linear regression analyses

Coefficient	*B*	*SE*(*B*)	β	*p*	r^2_p	VIF
Model 1: including the three proposed factors (adjusted $R^2=0.095$)						
(Constant)	−0.638	0.167	—	0.000	—	—
Science education	0.003	0.004	0.040	0.434	0.001	1.22
Science	0.004	0.001	0.277	0.000	0.057	1.41
Age	0.007	0.008	0.047	0.347	0.002	1.19
Model 2: model 1 plus dummy-coded country variables (adjusted $R^2=0.093$)						
(Constant)	−0.617	0.178	—	0.001	—	—
Science education	0.004	0.005	0.049	0.394	0.002	1.545
Science	0.004	0.001	0.270	0.000	0.047	1.658
Age	0.007	0.008	0.046	0.359	0.002	1.198
Science education starting early (1 = yes)	−0.027	0.078	−0.018	0.732	0.000	1.345

scientific models and their associated skills (Hartmann et al., 2015). The assessment instrument includes multiple-choice items requiring respondents to identify the attractor out of four answering options and various sources of validity evidence have been considered during test development that support the proposed test score interpretation (Krüger et al., 2020). In addition, this study employed a testing instrument which was systematically translated and evaluated and is available in three languages (Krell et al., 2018, 2020).

In this study, the *amount of science education classes* was used as an indicator for explicit reflections on scientific reasoning (specifically emphasising epistemic knowledge), the *amount of science classes* was used as an indicator for learning science content knowledge and the workload in both kinds of classes was measured in terms of ECTS credits (EU, 2015). These credits provided a sound basis that allowed a cross-country comparison, although not all countries follow this approach. On the other hand, ECTS credits are proposed to reflect typical workload necessary to achieve the curriculum goals, but not the exact workload of each individual student, and no information about the students' individual workload, the teaching methods or specific course content could be considered.

For the first time, a statistical analysis of SRC at six universities in three different countries was undertaken. The findings propose specific relations between the considered factors and the pre-service science teachers' SRC for the different subsamples. The basic data analysis revealed that the Canadian pre-service science teachers scored higher on the instrument than the Chilean and German ones. Whilst all three subsamples scored similarly on the items related to conducting scientific investigations (correct responses: GER: 53%, CHI: 46% and CAN: 53%), substantial differences exist for the items related to using scientific models (GER: 39%, CHI: 35% and CAN: 48%). These findings indicate rather basic competencies for all three subsamples, with the Canadian pre-service science teachers possessing superior competencies related to scientific modelling compared to the German and Chilean participants. This finding might be explained with the Canadian consecutive programme, which–on average–recruited older, more educated candidates than the Chilean and German programmes (Zuzovsky & Donitsa-Schmidt, 2017). We also found, as previous literature suggests (Hartmann et al., 2015; Limueco & Prudente, 2018; Schwichow & Nehring, 2018), a significant positive correlation between the pre-service science teachers' WLE and the *amount of science education classes*, the *amount science classes* and their *age* (with small to medium effect sizes; Fritz et al., 2012); however, only the correlation between the *amount science classes* and the WLE could be observed anymore if the other variables were controlled (i.e. partial correlation).

Generally, it is plausible that the identified relations are the result of different study regulations in the three countries, resulting, for example, in high correlations between the *amount of science education* and the *amount of science classes* in concurrent programmes (e.g. Germany) but not in consecutive programmes (e.g. Canada). The specificities of study regulations and the resulting confounding or constancy of variables (Table 6) make analyses of the impact of contributing factors on the development of SRC, focusing on single countries or universities (e.g. Khan & Krell, 2019), difficult and potentially limit their generalizability.

The first multiple linear regression analysis reveals that the inclusion of the proposed factors explained about 9–10% of the total variance in the pre-service science teachers' SRC (medium effect size; Fritz et al., 2012). The only significant factor in the regression analysis was *amount of science classes*, with standardised $\beta = 0.277$ (medium effect size; Acock, 2014), that is, pre-service science teachers with more ECTS credits in science classes significantly outperformed their peers on the SRC instrument. This finding remains if the two kinds of study programmes are considered as well (Table 7).

Based on these findings, H2 can be confirmed but H1 and H3 must be rejected. Hence, the present study supports the assumption of science content knowledge being a specific prerequisite for SRC (e.g. Fischer et al., 2014). The participants in the present study may have benefited from higher levels of scientific content knowledge to better understand the authentic scientific contexts in the items and to apply their procedural and epistemic knowledge within the respective context to identify the attractor. Related to H1, the findings indicate that the *amount of science education classes* did not necessarily impact pre-service science teachers' SRC. This contradicts the assumption the *amount of science education classes* (as indicator for explicit reflections on scientific reasoning) positively contributes to pre-service science teachers' SRC (e.g. Bruckermann et al., 2018; Hartmann et al., 2015). The present findings could be explained with insufficient science education classes (Cofré et al., 2015). Another explanation—at least partly—might be the present sample groups, which do not fully cover all semesters of teacher education at university level in all six universities (Table 2). H3 has also to be rejected based on the present findings. Hence, the pre-service science teachers' *age* did not significantly predict their SRC. This is less surprising as there are ambiguous findings related to age and formal education as predictors of SRC (Ding, 2017; Ding et al., 2016; Limueco & Prudente, 2018). Most studies emphasising the importance of age and intellectual development as positively contributing to scientific reasoning focus on younger children (Kuhn, 1989; Zimmermann, 2007). Hence, the impact of age might be less significant for pre-service science teachers, as processes of cognitive development already occurred in these age groups. Also, the present findings support perspectives more strongly emphasising the situated nature of scientific reasoning, such as the 'epistemology in practice' perspective (Berland et al., 2016).

Conclusion

SRC are defined as a latent, complex construct that encompasses the skills and knowledge needed for scientific problem solving and the capacity to seriously reflect about the problem-solving process on a meta-level (Lawson, 2004; Morris et al., 2012). It is proposed, in general, that competence development in teacher education might be best supported with a sequence of authentic and complex contexts for problem solving, accompanied by explicit reflections about the problem-solving processes carried out (Khan & Krell, 2019; Max, 1999). In terms of the development of SRC, pre-service science teachers might benefit when challenged with multiple situations, in which their recent reasoning is not fruitful and, hence, new rules or strategies have to be applied (Lawson, 2004). Based on the present findings, it can be concluded, unlike other studies focusing on single countries or universities (e.g. Hartmann et al., 2015), that pre-service science teachers benefit from strong science content knowledge which supports them to competently reason scientifically. For science teacher education programmes to be more efficient in fostering pre-service science teachers' SRC, this would suggest to create a stronger coherence between what is taught in different kind of classes and within each kind of class. For example, pre-service science teachers could benefit if procedural and epistemic knowledge is fostered and discussed in science education classes within contexts they already learned about in their science classes and, hence, possess the relevant content knowledge (i.e. coherence between different kinds of classes). Furthermore, and as suggested for competence development in general (Max, 1999), a coherent instructional sequence within science education classes should provide pre-service teachers with learning opportunities with increasing complexity to support the transferability of procedural and epistemic knowledge to various contexts and the development of new strategies.

In general, it is proposed that a wide range of high impact course activities are needed for the development of pre-service science teachers' SRC and that the interplay of various learning opportunities, including appropriately designed interventions, might positively contribute to the development of SRC (Krüger et al., 2020; Lawson, 2004; Stammen et al., 2018). Hence, as the present study did not evaluate interaction effects and delayed effects, it might still be the case that the combination of the *amount of science classes* and the *amount of science education classes* in all study programmes considered in this study positively contributed to the development of the pre-service science teachers' SRC. For example, pre-service science teachers visiting science classes may have benefited from having visited science education classes earlier—or vice versa.

The inclusion of pre-service science teachers from different universities in this study, a first of its kind, makes the present findings more robust. Future studies can build on the findings and replicate them based on larger samples, including additional countries, and conduct hierarchical linear modelling, a more common approach to analysing hierarchical data (Hodges, 2016). Future studies can also build on the present findings with a more in-depth analysis of the two factors: explicit reflections on scientific reasoning and science content knowledge and their contribution to the development of pre-service science teachers' SRC. For example, by considering the students' individual workload, the teaching methods and specific course content and teacher education activities, we might be able to ascertain further how and where science teacher education can best promote scientific reasoning. This type of analysis might contribute to learning more about the complex nature of competence development related to scientific reasoning, for example by testing different assumptions about the way content knowledge is used for scientific reasoning and which quantity and quality of content knowledge are needed to allow scientific reasoning on a specific level (Ruppert et al., 2017; Sadler & Fowler, 2006). Finally, future studies should extend their focus, consider further dimensions of teachers' professional competencies (e.g. PCK; Baumert & Kunter, 2013) and investigate how the level of science teachers' SRC impacts their teaching practices and—ultimately—the development of their students' science competencies.

Funding Open Access funding enabled and organized by Projekt DEAL. This work was supported by the German Federal Ministry of Education and Research (grant numbers 01PK11004A, 01PK15004A to DK), the International Cooperation programme for attraction of advanced human capital (grant number MEC 80180066 to HC) and the 2018 UBC-FUB Joint Funding Scheme (to MK and SK).

References

Acock, A. (2014). *A gentle introduction to Stata*. Stata Press.

AERA, APA, & NCME [American Educational Research Association, American Psychological Association, & National Council on Measurement in Education.] (2014). *Standards for educational and psychological testing*. Washington, DC: American educational research association.

Baumert, J., & Kunter, M. (2013). The COACTIV model of teachers' professional competence. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), *Cognitive activation in the mathematics classroom and professional competence of teachers* (pp. 25–48). Springer, US.

BCMOE [British Columbia Ministry of Education.] (2019). BC's new curriculum: Science. Retrieved from: https://curriculum.gov.bc.ca/curriculum/science

Berland, L., Schwarz, C., Krist, C., Kenyon, L., Lo, A. S., & Reiser, B. (2016). Epistemologies in practice: Making scientific practices meaningful for students. *Journal of Research in Science Teaching, 53*, 1082–1112.

Bond, T., & Fox, C. (2001). *Applying the Rasch model*. Erlbaum.

Bruckermann, T., Ochsen, F., & Mahler, D. (2018). Learning opportunities in biology teacher education contribute to understanding of nature of science. *Education Sciences, 8*, 103. https://doi.org/10.3390/educsci8030103

Bybee, R. (2014). NGSS and the next generation of science teachers. *Journal of Science Teacher Education, 25*, 211–221. https://doi.org/10.1007/s10972-014-9381-4

Carlson, J., & Daehler, K. (2019). The refined consensus model of pedagogical content knowledge in science education. In A. Hume, R. Cooper, & A. Borowski (Eds.), *Repositioning pedagogical content knowledge in teachers' knowledge for teaching science* (pp. 77–92). Springer.

Cofré, H. L., Vergara, C., Gonzalez–Weil, C., Santibáñez, D., Ahumada, G., Furman, M., Podesta, M. E., Camacho, J., Gallego, R., & Pérez, R. (2015). Science teacher education in South America. *Journal of Science Teacher Education, 26,* 45–63.

Cofré, H. L., Vergara, C., Santibáñez, D., & Pavez, J. (2022). Preservice science teacher education around the globe: trends, challenges, and future directions. In J. Luft & G. Jones (Eds.), *Handbook of Research on Science Teacher Education* (pp. 157–178). Taylor & Francis.

Ding, L. (2017). Progression trend of scientific reasoning from elementary school to university. *International Journal of Science and Mathematics Education, 28*, 1161. https://doi.org/10.1007/s10763-017-9844-0

Ding, L., Wei, X., & Mollohan, K. (2016). Does higher education improve student scientific reasoning skills? *International Journal of Science and Mathematics Education, 14*, 619–634. https://doi.org/10.1007/s10763-014-9597-y

Duschl, R. A., & Grandy, R. (2013). Two views about explicitly teaching nature of science. *Science & Education, 22*, 2109–2139. https://doi.org/10.1007/s11191-012-9539-4

Engelmann, K., Neuhaus, B., & Fischer, F. (2016). Fostering scientific reasoning in education. *Educational Research and Evaluation, 22*, 333–349.

EU. (2015). *ECTS users' guide 2015*. Luxembourg: Publication Office of the European Union.

Field, A. (2009). *Discovering statistics using SPSS*. Sage.

Fischer, F., Kollar, I., Ufer, S., Sodian, B., Hussmann, H., Pekrun, R., Neuhaus, B., Dorner, B., Pankofer, S., Fischer, M., Strijbos, J.-W., Heene, M., & Eberle, J. (2014). Scientific reasoning and argumentation. *Frontline Learning Research, 5*, 28–45.

Fritz, C., Morris, P., & Richler, J. (2012). Effect size estimates. *Journal of Experimental Psychology: General, 141*, 2–18.

Göhner, M., & Krell, M. (2022). Analyzing the relationship between pre-service biology teachers' modeling processes, scientific reasoning competencies, and general cognitive abilities. In K. Korfiatis & M. Grace (Eds.), *Current research in biology education*. Springer (in press).

Großschedl, J., Harms, U., Kleickmann, T., & Glowinski, I. (2015). Preservice biology teachers' professional knowledge. *Journal of Science Teacher Education, 26*, 291–318. https://doi.org/10.1007/s10972-015-9423-6

Hammann, M., Phan, T., Ehmer, M., & Grimm, T. (2008). Assessing pupils' skills in experimentation. *Journal of Biological Education, 42*, 66–72.

Harkness, J. (2003). Questionnaire translation. In J. Harkness, van de Vijver, Fons J. R., & P. Mohler (Eds.), *Cross-cultural survey methods* (pp. 35–56). Hoboken, NJ: Wiley.

Hartmann, S., Upmeier, A., zu Belzen, A. U., Krüger, D., & Pant, H. A. (2015). Scientific Reasoning in Higher Education. *Zeitschrift für Psychologie, 223*(1), 47–53. https://doi.org/10.1027/2151-2604/a000199

Hodges, J. (2016). *Richly parameterized linear models*. Chapman; Hall/CRC.

Khan, S., & Krell, M. (2019). Scientific reasoning competencies. *Canadian Journal of Science, Mathematics and Technology Education.* https://doi.org/10.1007/s42330-019-00063-9

Kind, P., & Osborne, J. (2017). Styles of scientific reasoning? *Science Education, 101*, 8–31. https://doi.org/10.1002/sce.21251

KMK (2019). Ländergemeinsame inhaltliche Anforderungen für die Fachwissenschaften und Fachdidaktiken in der Lehrerbildung [Common federal requirements for science and science education in teacher education]. Retrieved from https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2008/2008_10_16- Fachprofile-Lehrerbildung.pdf. Accessed 21 Jan 2022.

Krell, M., Koska, J., Penning, F., & Krüger, D. (2015). Fostering pre-service teachers' views about nature of science. *Research in Science & Technological Education, 33*(3), 344–365. https://doi.org/10.1080/02635143.2015.1060411

Krell, M., Redman, C., Mathesius, S., Krüger, D., & van Driel, J. (2018). Assessing pre-service science teachers' scientific reasoning competencies. *Research in Science Education*. https://doi.org/10.1007/s11165-018-9780-1

Krell, M., Mathesius, S., van Driel, J., Vergara, C., & Krüger, D. (2020). Assessing scientific reasoning competencies of pre-service science teachers. *International Journal of Science Education*. https://doi.org/10.1080/09500693.2020.1837989

Krüger, D., Hartmann, S., Nordmeier, V., & Upmeier zu Belzen, A. (2020). Measuring scientific reasoning competencies. In O. Zlatkin-Troitschanskaia, H. A. Pant, M. Toepper, & C. Lautenbach (Eds.), *Student Learning in German Higher Education* (pp. 261–280). Wiesbaden: Springer VS.

Kuhn, D. (1989). Children and adults as intuitive scientists. *Psychological Review, 96*, 674–689.

Kuhn, D., & Pearsall, S. (2000). Developmental origins of scientific thinking. *Journal of Cognition and Development, 1*, 113–129. https://doi.org/10.1207/S15327647JCD0101N_11

Kwon, Y., & Lawson, A. (2000). Linking brain growth with the development of scientific reasoning ability and conceptual change during adolescence. *Journal of Research in Science Teaching, 37*, 44–62.

Lawson, A. (2004). The nature and development of scientific reasoning. *International Journal of Science and Mathematics Education, 2*, 307–338. https://doi.org/10.1007/s10763-004-3224-2

Lead States, N. G. S. S. (Ed.). (2013). *Next Generation Science Standards*. The National Academies Press.

Limueco, J., & Prudente, M. (2018). Predicting progression trends of scientific reasoning skills and meta-cognitive awareness among secondary level students. *DLSU Research Congress Proceedings, 6*, LLI-14. Retrieved from https://www.dlsu.edu.ph/wp-content/uploads/pdf/conferences/research-congress-proceedings/2018/lli-14.pdf. Accessed 21 Jan 2022.

Mahler, D., Bock, D., & Bruckermann, T. (2021). 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? *Education Sciences, 11*, 558. *Retrieved from*. https://doi.org/10.3390/educsci11090558

Mathesius, S., Upmeier zu Belzen, A., & Krüger, D. (2014). Kompetenzen von Biologiestudierenden im Bereich der naturwissenschaftlichen Erkenntnisgewinnung: Entwicklung eines Testinstruments. *[Scientific inquiry competencies of pre-service biology teachers: Developing a measurement instrument] Erkenntnisweg Biologiedidaktik, 13*, 73–88.

Mathesius, S., Hartmann, S., Upmeier zu Belzen, A., & Krüger, D. (2016). Scientific reasoning as an aspect of pre-service biology teacher education. In T. Tal & A. Yarden (Eds.), *The future of biology education research. Proceedings of the 10th conference of European Researchers in Didactics of Biology (ERIDOB) (pp. 93*–110). Haifa, Israel.

Max, C. (1999). *Entwicklung von Kompetenz*. Peter Lang.

Mayer, D., Sodian, B., Koerber, S., & Schwippert, K. (2014). Scientific reasoning in elementary school children: Assessment and relations with cognitive abilities. *Learning and Instruction, 29*, 43–55.

Mineduc (2012). *Estándares Orientadores para Carreras de Pedagogía en Educación Media* [Guiding standards for teaching programs in high school education]. Ministry of Education.

Morris, B., Croker, S., Masnick, A., & Zimmerman, C. (2012). The emergence of scientific reasoning. In H. Kloos, B. Morris, & J. Amaral (Eds.), *Current topics in children's learning and cognition* (pp. 61–82). InTech.

Neumann, K., Kind, V., & Harms, U. (2019). Probing the amalgam. *International Journal of Science Education, 41*, 847–861.

Neumann, K., Härtig, H., Harms, U., & Parchmann, I. (2017). Science teacher preparation in Germany. In *Model science teacher preparation programs*, edited by J. Pedersen, T. Isozaki, and T. Hirano, 29–52. Information Age.

Opitz, A., Heene, M., & Fischer, F. (2017). Measuring scientific reasoning: A review of test instruments. *Educational Research and Evaluation, 23*, 78–101.

Osborne, J. (2013). The 21st century challenge for science education. *Thinking Skills and Creativity, 10*, 265–279. https://doi.org/10.1016/j.tsc.2013.07.006

Osborne, J. (2014). Teaching scientific practices. *Journal of Science Teacher Education, 25*, 177–196. https://doi.org/10.1007/s10972-014-9384-1

Park, S., & Oliver, J. S. (2008). Revisiting the conceptualization of pedagogical content knowledge (PCK): PCK as a conceptual tool to understand teachers as professionals. *Research in Science Education, 38*, 261–284.

Ruppert, J., Duncan, R., & Chinn, C. (2017). Disentangling the role of domain-specific knowledge in student modeling. *Research in Science Education, 27*, 33. https://doi.org/10.1007/s11165-017-9656-9

Sadler, T., & Fowler, S. (2006). A threshold model of content knowledge transfer for socioscientific argumentation. *Science Education, 90*, 986–1004. https://doi.org/10.1002/sce.20165
Samarapungavan, A. (2018). Construing scientific evidence. In F. Fischer, C. Chinn, K. Engelmann, & J. Osborne (Eds.), *Scientific reasoning and argumentation* (pp. 56–76). Routledge.
Schwichow, M., & Nehring, A. (2018). Variablenkontrolle beim Experimentieren in Biologie, Chemie und Physik [Controlling variables in biology, chemistry and physics]. *Zeitschrift Für Didaktik Der Naturwissenschaften, 24*, 217–233. https://doi.org/10.1007/s40573-018-0085-8
Shavelson, R. (2018). Discussion of papers and reflections on "exploring the limits of domain-generality". In F. Fischer, C. A. Chinn, K. F. Engelmann & J. Osborne (Eds.), *Scientific reasoning and argumentation. The roles of domain-specific and domain-general knowledge* (pp. 102–117). New York: Routledge.
Shulman, L. (1986). Those who understand. *Educational Researcher, 15*, 4–14.
Stammen, A., Malone, K., & Irving, K. (2018). Effects of modeling instruction professional development on biology teachers' scientific reasoning skills. *Education Sciences, 8*, 119. https://doi.org/10.3390/educsci8030119
Woolley, J., Deal, A., Green, J., Hathenbruck, F., Kurtz, S., Park, T., Pollock, S., Transtrum, M., & Jensen, J. (2018). Undergraduate students demonstrate common false scientific reasoning strategies. *Thinking Skills and Creativity, 27*, 101–113.
Wu, M. L., Adams, R., Wilson, M., & Haldane, S. (2007). *ACER ConQuest*. ACER Press.
Zimmermann, C. (2007). The development of scientific thinking skills in elementary and middle school. *Developmental Review, 27*, 172–223.
Zuzovsky, R., & Donitsa-Schmidt, S. (2017). Comparing the effectiveness of two models of initial teacher education programmes in Israel. *European Journal of Teacher Education, 40*, 413–431. https://doi.org/10.1080/02619768.2017.1318377
Zwick, R., Thayer, D., & Lewis, C. (1999). An empirical bayes approach to Mantel-Haenszel DIF analysis. *Journal of Educational Measurement, 36*, 1–28.

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

Authors and Affiliations

Moritz Krell[1] · Samia Khan[2] · Claudia Vergara[3] · Hernán Cofré[4] · Sabrina Mathesius[5] · Dirk Krüger[5]

Samia Khan
samia.khan@ubc.ca

Claudia Vergara
claudia.vergara12@gmail.com

Hernán Cofré
hernan.cofre@pucv.cl

Sabrina Mathesius
sabrina.mathesius@fu-berlin.de

Dirk Krüger
dirk.krueger@fu-berlin.de

[1] IPN - Leibniz Institute for Science and Mathematics Education, Kiel, Germany

[2] University of British Columbia, Vancouver, Canada

[3] Alberto Hurtado University Santiago, Santiago, Chile

[4] Pontificia Universidad Católica de Valparaíso, Valparaíso, Chile

[5] Freie Universität Berlin, Berlin, Germany