Progress in Science Education
2023, Vol. 6, No.1, 69-79
ISSN 2405-6057
DOI.10.25321/prise.2023.1390

Special Issue

Proceedings of the 12th DiNat Forum 2022

Capturing prospective teachers' beliefs about experimentation in geography classrooms: Validation of a new instrument

Hanna Velling[1], Jan Christoph Schubert[1]

Received: September 2022 / Accepted: November 2022

Structured Abstract

Background: Experimentation in geography teaching is considered to have great didactic potential, especially in the light of scientific literacy, but this potential is countered by the fact that the method is rarely used in the classroom. Geography teachers, with their subjective *beliefs* as part of their professional competence, represent a decisive factor for the use, design, and quality of experimentation in the classroom. Up to now, however, there is a lack of knowledge about the characteristics of (prospective) teachers' *beliefs* about experimentation in geography education as well as of corresponding measurement instruments.
Purpose: Therefore, the aim of the present study is the confirmatory validation of a newly developed measurement instrument in the form of a questionnaire on prospective teachers' *beliefs* about experimentation in geography classrooms.
Design and Methods: A measurement instrument on prospective teachers' *beliefs* about experimentation in geography classrooms was previously developed based on re-analyses of qualitative interview studies and on literature research and the dimensionality of the construct was examined using exploratory factor analysis. Within the scope of the present study, the factor structure obtained was then specified by means of confirmatory factor analysis with an independent sample of prospective geography teachers. We further examined the reliability and validity of the five *belief* scales about potentials, learning preconditions and the implementation of experimentation in geography teaching by employing measurement invariance evaluation and correlation analysis.
Sample: The confirmatory factor analysis was conducted with a sample of $N = 344$ student teachers of all semesters (1–13) and all German school types training to become geography teachers. The study was conducted as an online survey between June and November 2021.
Results: Our analyses support the construct validity ($\chi^2 = 428.255$, $p = .001$, $df = 339$; CFI: .962; TLI: .958; RMSEA = .028, 90% C.I. [.019, .035]; SRMR = .051; $\chi^2/df = 1.26$) and reliability ($.64 \leq \alpha \leq .88$; $65 \leq \varrho_c \leq .87$) of the modified instrument. The results indicate strong measurement invariance across gender, school type (primary vs. secondary school), practical experience, and further studied STEM subject, suggesting that the instrument can be used for the designated subgroups and is suitable for mean value comparisons. Furthermore, the analysis provides first evidence for convergent and discriminant validity for the five scales.
Conclusions: With the questionnaire, for the first time an exploratively and confirmatory validated measurement instrument is available to survey the *beliefs* of prospective teachers about experimentation in geography teaching. Planned further steps in our research project as well as possible applications in practice are discussed.
Keywords: *teacher beliefs, experiments, geography class, science teaching, questionnaire, scale development*

[1]Friedrich-Alexander-Universität Erlangen-Nürnberg, Lehrstuhl für Didaktik der Geographie
✉ hanna.velling@fau.de

1 Introduction

As a method of scientific knowledge, experiments offer a wide range of didactic potential: they can contribute to the development of methodological skills as well as of flexibly applicable knowledge, arouse curiosity and interest in the natural sciences, and can trigger important cognitive processes in students. In addition, their use offers the possibility to realise inquiry-based and problem-oriented learning in the classroom (Kremer et al., 2019; Lethmate, 2006; Otto, 2009; Schubert, 2016; Wilhelmi, 2012), so they are particularly suitable for geography teaching for the implementation of scientific literacy (Labudde & Möller, 2012; Lethmate, 2006; Otto & Mönter, 2015). At the same time, the implementation of experiments in geography lessons puts high requirements on teachers, both in terms of preparation and in terms of conducting and designing these lessons (Fögele et al., 2020; Otto & Mönter, 2015). In reality, geography teachers rarely apply experiments (Hemmer & Hemmer, 2010), which is in clear contradiction to students' decades-long high interest in experimenting (Hemmer & Hemmer, 2002, 2010, 2021) as well as its great didactic potential. This observed discrepancy highlights the need to investigate factors among teachers that are relevant to the implementation of experiments in geography classrooms.

While first evidence suggests that the didactic-methodological qualification of a teacher as well as, on the external level, organisational-formal framework conditions might act as barriers (Schubert & Höhnle, 2016), little is known about other (internal) factors influencing the professional action of geography teachers when conducting experiments in the classroom. In the meantime, the subjective *beliefs* of teachers are, among others, considered to be a particularly important predictor of the design of lessons as well as the teachers' actions during these lessons (Fives & Buehl, 2012; Nespor, 1987; Pajares, 1992; Reusser & Pauli, 2014) — and thus possibly also of the use of experiments in geography teaching. Nevertheless, little is known about geography teachers' *beliefs* in general and even less about their *beliefs* on experimentation in geography lessons, and there is as yet no instrument to measure these. However, in order to examine the connection between teachers'*beliefs* and their actual teaching activities when applying experiments in geography lessons, as well as the interaction with other influencing factors, it is a first and at the same time necessary step to make their *beliefs* ascertainable on the basis of a valid measurement instrument. Therefore, the focus of the present study is the confirmatory validation of a newly developed measurement instrument to capture the *beliefs* of prospective teachers about experimentation in geography teaching in form of a questionnaire.

2 Research Background

Teachers' professional competencies are considered important predictors of their classroom actions and thus of the quality of their teaching (Baumert & Kunter, 2006; Steinmann & Oser, 2012). Among additional components of this professional competence, such as their pedagogical content knowledge, teachers' subjective *beliefs* play a significant role in shaping classroom practice through filtering, framing, and guiding (Fives & Buehl, 2012). They influence a teacher's perception and interpretation of teaching situations, the selection of teaching goals, and thus also the planning and designing of lessons as well as instructional decisions (Fives & Buehl, 2012; Kunter et al., 2013; Nespor, 1987; Pajares, 1992; Reusser & Pauli, 2014).

For a definition of teacher *beliefs*, Richardson is widely referenced, who describes *beliefs* as "psychologically held understandings, premises or propositions about the world that are felt to be true" (1996, p. 103). Consequently, unlike knowledge, for their effectiveness in action, *beliefs* require neither discursive validation nor freedom from contradiction nor argumentation justification (Baumert & Kunter, 2006). In addition, *beliefs* are described as implicit or explicit object-related conceptions with both affective and cognitive components, which partly consciously and partly unconsciously guide one's actions (Kuhl et al., 2013).

Teacher *beliefs* about learning content and instructional practice relate to different contexts and contents (Fives & Buehl, 2012; Mansour, 2009), such as which teaching methods are effective, the way children learn, or what role student activity plays (Oser & Blömeke, 2012). Consequently, they vary according to the level of specificity studied (Gimbel et al., 2018). Compared to epistemological *beliefs*, for example, *beliefs* about learning content and instructional practice are considered to guide instruction more directly (Leuchter et al., 2006). Depending on the requirements of a certain situation, certain *beliefs* are activated in a strongly contextualised way, so that teacher *beliefs* are of high importance, especially for concrete didactic and methodological decisions. This illustrates the necessity to operationalise teacher *beliefs* subject-specific on the one hand and as requirement-specific as possible on the other (Blömeke et al., 2008; Fives & Buehl, 2012; Petermann & Vorholzer, 2022).

For science teachers, too, several studies suggest that their *beliefs* about inquiry-based learning and teaching and about laboratory application in the science classroom are important predictors of the actual implementation of these (Cheung, 2011; Crawford, 2007; Lucero et al., 2013; Mansour, 2009; Ramnarain, 2014; Wallace & Kang, 2004; Wilkins, 2008). Prospective and in-service science teachers seem to predominantly hold positive *beliefs* towards the application of laboratory (e.g., Aka, 2016; Cheung, 2011; Siorenta & Jimoyiannis, 2008; Wallace & Kang, 2004) as well as inquiry-based learning and teaching (e.g., Bevins et al., 2019; DiBiase & McDonald, 2015; Gavora & Wiegerová, 2019; Ramnarain & Hlatswayo, 2018; Wilkins, 2008) and seem to mainly believe in the value of these approaches for science education.

However, also for *beliefs* regarding these domains, the evidence on the practical effectiveness of teacher *beliefs* is partly inconsistent, and some studies fail to demonstrate a clear relationship (Mellado et al., 2007; Saad & BouJaoude, 2012). Fives and Buehl (2012) suggest that this inconsistent picture may partly be attributed to the level of specificity with which the *beliefs* are examined. Furthermore, the impact of teacher *beliefs* on instructional practice is influenced by numerous factors, which may oppose or support the congruence between teachers' *beliefs* and their actual classroom actions (Buehl & Beck, 2015; Fives & Buehl, 2012; Mansour, 2009). With respect to science teachers' *beliefs* about experimentation-related fields such as inquiry and laboratory application, studies report about, on the one hand, internal factors such as conflicting and competing *beliefs* of one's own *belief* system, professional knowledge, or self-efficacy *beliefs*. On the other hand, external factors such as contextual settings of teaching, like inadequate equipment, time constraints, and the curriculum are referred to (Crawford, 2007; DiBiase & McDonald, 2015; Fang, 2021; Mansour, 2009; Ramnarain, 2016; Siorenta & Jimoyiannis, 2008; Wallace & Kang, 2004). Even though experimentation is closely related to the requirements of inquiry-based learning and laboratory application, it is unclear to what extent the findings listed are transferable to experimentation in geography classes. However, for the application of scientific methods in geography classrooms, too, it has been shown that prospective teachers perceive organisational-formal framework conditions and the teacher's didactic-methodological qualification as obstacles to the use of experiments in geography class (Schubert & Höhnle, 2016). In addition, in-service geography teachers name the lack of interest and insufficient preknowledge of students as further obstacles to experimentation. Nevertheless, they perceive experimentation as an opportunity to motivate students, facilitate their understanding of subject content, and promote competencies related to gaining scientific knowledge (Miener & Köhler, 2013).

Other findings, especially on competency aspects regarding the specific case of experimentation in the geography classroom, are largely missing for (prospective) geography teachers (Rosendahl et al., 2020), including their *beliefs*. The nature of these *beliefs*, how and to what extent they affect the use of experiments in geography classrooms, and the influence of, for example, perceived barriers or self-efficacy *beliefs* have hardly been researched so far. Overall, there is still little research on the *beliefs* of geography teachers, especially with regard to specific teaching approaches and methods in geography lessons. However, it is precisely such a subject- and requirement-specific approach that is of particular interest (Fives & Buehl, 2012; Petermann & Vorholzer, 2022). While there exist measurement instruments for *beliefs* in the aforementioned areas of science education (inquiry-based learning and teaching: e.g., Gavora & Wiegerová, 2019; Marshall et al., 2009; laboratory application: e.g., Aka, 2016; Cheung, 2011), there is a lack of measurement instruments for *beliefs* specifically about experimentation and specifically for (prospective) geography teachers. Furthermore, extensive previous research often focuses on qualitative methods and smaller samples. The development of such a tool is therefore a first step towards understanding the described research gaps, with its careful design and validation being of high importance to justify its use for further research and practice.

From the above, the following objective arises: A previously newly developed quantitative instrument for the survey of the *beliefs* of prospective geography teachers on experimentation in the classroom is to be validated, on the basis of which various aspects of these *beliefs* can be measured. Specifically, these *beliefs* are to be made empirically ascertainable by means of a time-economically manageable measurement instrument in the form of a standardised questionnaire.

This leads to the central research question of the present article: Can the *beliefs* of prospective teachers about various aspects of experimentation in geography classrooms be validly and reliably assessed by means of the developed questionnaire?

> Hypothesis 1: We expect substantial loadings of scale items on the respective latent factors and satisfactory reliability of the scales.

With the goal of administering the measurement instrument in a heterogeneous population of prospective geography teachers and drawing comparisons across the different groups, generalisability of the construct across these groups (i.e. measurement invariance) is also to be tested.

> Hypothesis 2: We expect invariance of item intercepts regarding gender, school form (primary vs. secondary school), existing practical experience during the studies, and further studied STEM subject.

Furthermore, in the course of construct validation, first evidence of convergent and discriminant validity is to be examined.

> Hypothesis 3: We expect modest correlations between factors describing theoretically overlapping constructs and statistically nonsignificant or small correlations between factors relating to theoretically distinct constructs.

3 Methods

3.1 Scale Development

The development of the *belief* scales involved several steps, of which the present study focuses on the confirmatory validation. The previous steps are briefly outlined below; a detailed description of the scale development can be found in Velling et al. (2022).

We started with an open approach to develop the measurement instrument by first re-analysing two qualitative interview studies with geography teachers on potentials and barriers to experimentation in geography classrooms by Miener and Köhler (2013) and Bussfeld (2013). Using the categories developed in the summarising content analyses of these studies in combination with literature research, several possible content areas of *beliefs*, including goals and prerequisites for experimentation in the geography classroom, were identified. Based on these categories, a comprehensive item pool was generated. A first pre-test on a smaller sample of teacher education students with geography as a school subject ($N = 46$) was used as the basis for revising this first draft of the measurement instrument, and new items were added after a further literature review. Additionally, a team of geography didactics experts and geography teachers reviewed the developed item pool for clarity and comprehensibility as well as content breadth to ensure good content validity, resulting in a modified second draft with 50 items.

To examine the underlying factor structure of the *beliefs* about experimentation in the geography classroom, an exploratory factor analysis (EFA) was then conducted using an independent sample of teacher education students with geography as a school subject ($N = 198$). The factor analysis identified five factors with 30 items describing different content-related aspects of *beliefs* about experimentation in the geography classroom related to potentials, learning preconditions and the implementation of experiments. The first factor, *Potentials for Students*, describes the potential of experimentation to promote various competencies in students at the motivational, cognitive, and motoric level. In contrast, another factor, *Potentials for Structuring Lessons*, depicts the potential of experimentation for structuring lessons and shaping the lesson process through experimentation. Two other factors represent the learning preconditions of students regarding experimentation, firstly the *Cognitive and Motoric Learning Preconditions* and secondly the *Motivational Learning Preconditions*. The last factor highlights the importance of openness regarding the process and the results of experimentation.

For the validation study presented here, a total of five originally inversely formulated items were reformulated, since method effects in the form of cross-loadings were indicated for individual items by the inverse formulation. The response format used was a 5-point Likert scale with the anchors 1 = "do not agree at all" and 5 = "completely agree", and the items were arranged in random order to avoid unwanted sequence effects.

In addition to the 30 items, the questionnaire contained general information on socio-demographic data (age, gender, semester, studied school type, practical teaching experience, subjects studied). In order to get a common understanding of experimentation as opposed to other scientific working methods in geography classrooms, a brief information block with corresponding definitions according to Lethmate (2006) was added in German. This included additional advice to focus on experimentation in geography classrooms rather than in science classrooms in general ("Please remember the definition of experiments given below and refer only to experimentation in geography classrooms when answering the statements").

3.2 Sample

We included teacher education students with geography as a school subject from different Bavarian universities. They were invited to voluntarily participate in the online survey with the software SoSciSurvey via a questionnaire. After removing 31 cases from the data set due to ≥ 50% missing values, a sample of $N = 344$ remained. Three-quarters of the students identified as female ($n = 258$), 24.4% identified as male ($n = 84$), and 0.6% identified as diverse ($n = 2$). The mean age of the students was 22 ($SD = 3.5$). Students between their first and 13th semesters ($M = 4.56$, $SD = 3.02$) and of all Bavarian school types participated in the survey. A total of 37% of the respondents studied elementary school teaching ($n = 128$), while 63% of the students studied secondary school teaching (36% "Gymnasiallehramt" ($n = 123$), 14% "Mittelschullehramt" ($n = 47$), and 13% ($n = 46$) "Realschullehramt"). Slightly more than half of the students ($n = 196$) had no practical teaching experience, and 45% of the students studied another STEM subject ($n = 155$).

3.3 Procedures

Data gaps were found for only one additional subject. These occurred randomly according to Little's MCAR test ($\chi^2 = 25.860$; $df = 27$; $p > 0.05$), so missing at random can be assumed. The three missing values were therefore imputed using an expectation-maximisation algorithm. This procedure tolerates the violation of the normal distribution assumption and is also well suited for factor and reliability analyses (Graham, 2009; Lüdtke et al., 2007).

Having previously explored and refined the latent structure of the measurement instrument by EFA, following the recommendations for scale development by Brown (2015), we now conducted confirmatory factor analysis (CFA) with an independent larger sample. To evaluate the model fit, we applied Satorra–Bentler χ^2-tests as an absolute fit index, and inspected the comparative fit index (CFI), Tucker–Lewis index (TLI), root mean square error of approximation (RMSEA), and standardised root mean square residual (SRMR) (Brown, 2015). For the fit indices, the cut-off values suggested by Hu and Bentler (1999; see also Brown, 2015) were used: CFI/TLI ≥ .95, RMSEA ≤ .06, and SRMR ≤ .08. For factor identification, the first indicator of each factor was fixed to 1, and all other loadings of the other factors were freely estimated. For the measurement of internal consistency, Cronbach's alpha and congeneric reliability (ϱ_c, *composite reliability*, measured by $\varrho_c \geq .60$; Hair et al., 2014) were calculated.

In addition, we performed correlation analysis as well as measurement invariance evaluation (invariance of indicator intercept) for different person-related variables for construct validation of the questionnaire. For the latter, multiple-groups CFA was conducted and stepwise increasing equality restrictions (equal form, equal factor loadings, equal intercepts) were implemented to determine the degree of measurement invariance (Brown, 2015). The respective model comparisons were assessed using the Satorra–Bentler scaled χ^2-difference test as well as, due to the sensitivity of this goodness-of-fit test to sample size, based on Chen's rule of thumb (Chen, 2007), where as long as the RMSEA does not increase by .015 units and the CFI does not decrease by .010 units, measurement invariance is present. Since a large proportion of the items deviated from the standard normal distribution, all calculations were performed using the MLR estimator (*robust maximum likelihood estimation*) due to the robustness of this estimator towards deviation from the normal distribution (Brown, 2015). All analyses were conducted in R (version 4.1.0) using the lavaan (Rosseel, 2012) and the semTools (Jorgensen et al., 2018) packages.

4 Results

4.1 Confirmatory Factor Analysis and Reliability (Hypothesis 1)

To test Hypothesis 1, we conducted confirmatory factor analysis (CFA). The initial five-factorial CFA showed only reasonable fit for the indices CFI and TLI (χ^2 = 607.096, p < .001, df = 395; CFI: .920; TLI: .912; RMSEA = .040, 90% C.I. [.033, .045]; SRMR = .062). Therefore, we inspected the modification indices for plausible changes in the model. With the highest modification indices, this suggested cross loadings for two items on a non-targeted factor. For one item of the factor *Cognitive and Motoric Learning Preconditions* ("From my point of view, experiments are appropriate toward cognitive performance"), an additional salient factor loading on the factor *Potentials for Students* was suggested (M.I. = 57.154, EPC = 1.153). We assume that due to its wording ("appropriate"), this item achieved a higher mean value (M = 4.14; SD = .75) than other items of the factor (M = 3.35; SD = .73), so we attribute the factor loading to a method effect. For an item of the factor *Potentials for Students* ("I think experiments have a motivating effect when there are practical experimentation phases for the students"), a cross loading on the factor *Motivational Learning Preconditions* seemed plausible given the common motivational aspect (M.I. = 51.491, EPC = .550). Based on these considerations, we decided to exclude both items from the questionnaire, also to ensure good discriminant validity as "individual measured items should represent only one latent construct" (Hair et al., 2014, p. 620). Furthermore, a residual correlation between two items ("I believe that experimentation trains students' practical skills" and "I think experimentation promotes students' hands-on activity") was suggested (M.I. = 21.836, EPC = .084), amounting to r = .27 (.03), which seemed plausible given the similarity of the items in their German wordings. Further proposed changes in the model did not seem to be clearly theoretically justifiable. By dropping both items and admitting this residual correlation, the model fit could be improved and turned out to be good (χ^2 = 428.255, p = .001, df = 339; CFI: .962; TLI: .958; RMSEA = .028, 90% C.I. [.019, .035]; SRMR = .051). Even though the χ^2-test turned out to be significant again, the ratio χ^2/df = 1.26 indicated a very good model–data fit (Schermelleh-Engel et al., 2003). All standardised factor loadings were substantial and proceeded the cut-off of $\lambda \geq .45$ by Tabachnick and Fidell (2007) ($.45 \leq \lambda \leq .81$, median = .61).

All items, item means, standard deviations, and standardised factor loadings of the final version are shown in table 1. The English translations given are for communicative understanding only; the items used were all validated in their original German versions.

The internal consistency for all scales turned out to be satisfactory to good, as indicated by Cronbach's alpha ($.64 \leq \alpha \leq .88$) as well as by congeneric reliability ($.65 \leq \varrho_c \leq .87$), where all scales exceeded the value of $\varrho_c \geq .60$ (Hair et al., 2014) (Tab. 2).

Tab. 1. Item pool, means and standard deviations (in parentheses), and standardised factor loadings (λ) with standard errors (in parentheses).

	Factors and Items	*M* (*SD*)	λ (*SE*)
	F1 Potentials for Students ***F1 Potenziale für Schüler*innen***	**4.23 (.49)**	
(1)	I am convinced that students understand subject content better through experimentation. *Ich bin der Überzeugung, dass Schüler*innen durch das Experimentieren fachliche Inhalte besser verstehen.*	4.39 (.74)	.63 (.06)
(2)	From my point of view, experiments are suitable for the cognitive activation of students. *Aus meiner Sicht eignen sich Experimente zur kognitiven Aktivierung der Schüler*innen.*	4.35 (.78)	.61 (.06)
(3)	I think that experimentation teaches students to conduct simple geographical-scientific methods. *Ich denke, dass die Schüler*innen beim Experimentieren lernen, einfache geographisch-naturwissenschaftliche Arbeitstechniken auszuführen.*	4.21 (.75)	.56 (.05)
(4)	I believe that students learn to recognise cause-and-effect relationships through experimentation. *Ich glaube, Schüler*innen lernen beim Experimentieren Ursache-Wirkungs-Zusammenhänge zu erkennen.*	4.38 (.72)	.61 (.06)
(5)	In my opinion, experimentation trains the ability to develop explanations for geographical phenomena. *Meiner Meinung nach schult das Experimentieren die Fähigkeit, Erklärungen für geographische Phänomene zu entwickeln.*	4.24 (.77)	.63 (.05)
(6)	In my opinion, experimentation teaches students the ability to interpret data and information.	4.21 (.82)	.60 (.06)

	*Meiner Meinung nach erlernen Schüler*innen beim Experimentieren die Fähigkeit, Daten und Informationen zu interpretieren.*		
(7)	I think that experimentation trains the ability to pose geographical questions about geographical phenomena. *Ich meine, dass das Experimentieren die Fähigkeit schult, geographische Fragestellungen zu geographischen Phänomenen zu stellen.*	3.95 (.87)	.59 (.04)
(8)	I think experimentation teaches students to make predictions about geographical phenomena. *Ich meine, beim Experimentieren lernen die Schüler*innen Vorhersagen über geographische Phänomene zu treffen.*	3.75 (.87)	.47 (.04)
(9)	From my point of view, experiments encourage people to get to the bottom of things on their own. *Aus meiner Sicht ermutigen Experimente, selbst den Dingen auf den Grund zu gehen.*	4.44 (.72)	.61 (.06)
(10)	I believe that experimentation trains students' practical skills. *Ich bin der Meinung, dass das Experimentieren die praktischen Fähigkeiten der Schüler*innen schult.*	4.48 (.72)	.60 (.07)
(11)	I think that the students can learn a lot about the purposeful use of measurement tools and laboratory materials while experimenting. *Ich meine, dass die Schüler*innen beim Experimentieren viel über die zielgerichtete Nutzung von Messgeräten und Labormaterialien lernen können.*	4.14 (.87)	.45 (.05)
(12)	I think experimentation promotes students' hands-on activity. *Ich denke, das Experimentieren fördert die praktisch-handelnde Aktivität der Schüler*innen.*	4.51 (.70)	.61 (.06)
(13)	In my opinion, students learn to work in a focused way when they experiment. *Meiner Meinung nach erlernen die Schüler*innen beim Experimentieren fokussiertes Arbeiten.*	3.80 (.94)	.56 (.04)
(14)	In my opinion, experimentation trains the ability to observe attentively and in detail. *Meiner Meinung nach schult das Experimentieren die Fähigkeit, aufmerksam und detailliert zu beobachten.*	4.36 (.72)	.66 (.05)
F2 Potentials for Structuring Lessons ***F2 Potenziale für die Unterrichtsstrukturierung***		**3.31 (.79)**	
(15)	From my point of view, experiments facilitate the design and implementation of lessons. *Aus meiner Sicht erleichtern Experimente die Gestaltung und Durchführung von Unterricht.*	3.32 (1.11)	.51 (.07)
(16)	In my opinion, the use of experiments has a structuring effect on lessons. *Meiner Meinung nach wirkt der Einsatz von Experimenten strukturierend auf den Unterricht.*	3.37 (.97)	.71 (.06)
(17)	I think experiments help determine the class schedule. *Ich denke, Experimente helfen den Ablauf des Unterrichts festzulegen.*	3.24 (1.00)	.68 (.07)
F3 Cognitive & Motoric Learning Preconditions ***F3 kognitive & motorische Lernvoraussetzungen***		**3.35 (.73)**	
(18)	In my opinion, many students have the prior knowledge for experimentation. *Meiner Meinung nach verfügen viele Schüler*innen über das Vorwissen für das Experimentieren.*	2.69 (.97)	.54 (.06)
(19)	I think that most students are able to plan the design of an experiment. *Ich meine, dass die meisten Schüler*innen den Aufbau eines Experiments planen können.*	3.12 (.96)	.54 (.06)
(20)	From my point of view, many students have the manual skills to experiment. *Aus meiner Sicht besitzen viele Schüler*innen das handwerkliche Geschick zum Experimentieren.*	3.54 (.95)	.63 (.06)
(21)	I believe that the students are able to use the equipment necessary for experimentation. *Ich bin der Meinung, dass die Schüler*innen mit den für das Experimentieren notwendigen Geräten umgehen können.*	3.38 (.92)	.72 (.05)
F4 Motivational Learning Preconditions ***F4 motivationale Lernvoraussetzungen***		**4.51 (.61)**	
(22)	From my point of view, students are interested in experiments. *Aus meiner Sicht bringen Schüler*innen Experimenten Interesse entgegen.*	4.50 (.71)	.81 (.06)
(23)	I think that students are motivated for experimentation. *Ich denke, dass Schüler*innen für das Experimentieren motiviert sind.*	4.53 (.72)	.80 (.05)
(24)	In my opinion, students are open-minded and positive about experiments. *Meiner Meinung nach stehen Schüler*innen Experimenten aufgeschlossen und positiv gegenüber.*	4.51 (.69)	.73 (.06)
F5 Importance of Openness ***F5 Wichtigkeit von Offenheit***		**3.41 (.72)**	
(25)	In my opinion, it is important that all groups of students choose the same experimental design. (-) *Meiner Meinung nach ist es wichtig, dass alle Schülergruppen den gleichen Experimentaufbau wählen. (-)*	3.57 (1.07)	.64 (.07)
(26)	I think it is important that all groups of students arrive at identical results when experimenting. (-) *Ich finde es wichtig, dass alle Schülergruppen beim Experimentieren zu identischen Ergebnissen kommen. (-)*	3.96 (.96)	.55 (.07)
(27)	For me, clearly interpretable results are a condition for the use of experiments. (-) *Für mich sind eindeutig interpretierbare Ergebnisse Voraussetzung für den Einsatz von Experimenten. (-)*	3.01 (1.10)	.46 (.08)
(28)	I think it is important that the course of experimentation in class is exactly according to plan. (-) *Ich halte es für wichtig, dass der Experimentablauf im Unterricht genau nach Plan verläuft. (-)*	3.12 (1.04)	.59 (.07)

Items in English translation for communicative understanding only and original German wording (italics) with 5-point Likert scale (1 = "do not agree at all"/"stimme gar nicht zu", 5 = "completely agree"/"stimme vollkommen zu"); N = 344. (-) reverse-coded items; p <.001 for all factor loadings.

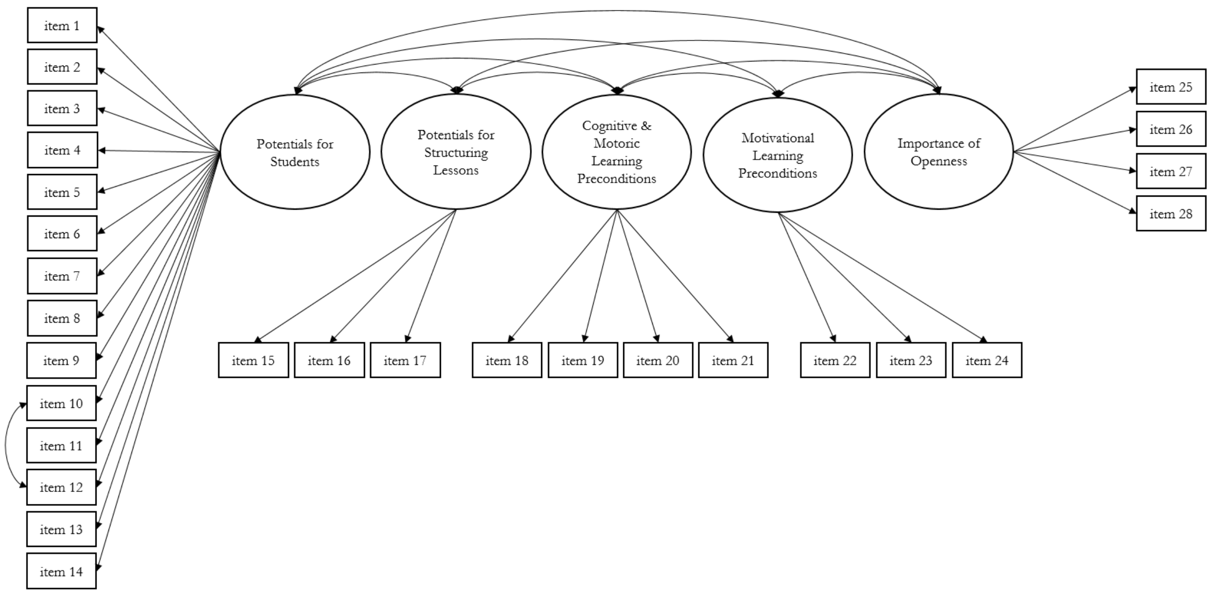

Fig. 1. Confirmatory Factor Analysis with Residual Correlation.

4.2 Measurement Invariance (Hypothesis 2)

To test Hypothesis 2, we conducted multiple-groups CFA (Brown, 2015) in order to check the measurement invariance of the indicator intercepts regarding gender (female vs. male), school type (primary vs. secondary school), and students with and without a further studied STEM subject as well as with or without practical teaching experience. For testing gender invariance, the gender data of the two persons who indicated "diverse" were coded as missing. The multi-group analyses were able to demonstrate strong measurement invariance regarding indicator intercepts across school type, practical experience, and further studied STEM subject as the inspection of the modification indices indicated that no substantial differences between the models existed. For gender invariance, the χ^2-test comparing the metric model and the less restrictive configural invariance model was significant ($\Delta\chi^2 = 38.824$, $\Delta df = 23$, $p < .05$). However, the more robust indices show that no substantial difference between both models existed (ΔCFI = .006, ΔRMSEA = .001), so strong measurement invariance can be assumed for all variables. This result suggests that the factor structure, the factor loadings, and the thresholds are comparable, and the construct is structurally consistent across the subgroups (Brown, 2015).

4.3 First Evidence for Discriminant and Convergent Validity (Hypothesis 3)

In order to test Hypothesis 3, we inspected the latent correlations between the factors (Tab. 2). The correlations found were modest to high but did not exceed the threshold value of $r = .80$, suggesting that they represent distinct constructs and thus indicating favourable discriminant validity (Brown, 2015).

Furthermore, between factors of theoretically similar constructs, modest correlations were found: the two factors relating to the potentials of experimentation in geography teaching (*Potentials for Students* and *Potentials for Structuring Lessons,* $r = .44$ [.02]) and the two factors relating to the students' learning preconditions (*Cognitive and Motoric Learning Preconditions* and *Motivational Learning Preconditions,* $r = .38$ [.03]) were substantially correlated, respectively. In addition, all factors directly related to the students were substantially correlated (*Potentials for Students* and *Motivational Learning Preconditions,* $r = .63$ [.05]; *Potentials for Students* and *Cognitive and Motoric Learning Preconditions,* $r = .49$ [.02]). In combination with all indicators loading significantly on the corresponding factors ($\lambda \geq .45$), this can be seen as first evidence for convergent validity (Brown, 2015).

Statistically nonsignificant or small correlations were found between factors related to theoretically distinct constructs, again indicating favourable discriminant validity (Brown, 2015).

Tab. 2. Reliability and latent bivariate standardised correlations (standard errors in parentheses) for the five factors.

	α	ϱ_c	(1)	(2)	(3)	(4)
(1) Potentials for Students	.88	.87	-			
(2) Potentials for Structuring Lessons	.66	.66	.44* (.02)	-		
(3) Cognitive & Motoric Learning Preconditions	.70	.70	.49* (.02)	.29* (.03)	-	
(4) Motivational Learning Preconditions	.83	.83	.63* (.05)	.14 (.02)	.38* (.03)	-
(5) Importance of Openness	.64	.65	.02 (.03)	-.27* (.04)	-.16 (.03)	.14 (.04)

N = 344, 5-point Likert-scale (1 = "do not agree at all", 5 = "completely agree"), *$p < .001$.

5 Discussion and Conclusions

The aim of the present study was to validate a newly developed measurement instrument in the form of a questionnaire on prospective teachers' *beliefs* about experimentation in geography classrooms. With the questionnaire, for the first time, an instrument is available to survey prospective teachers' *beliefs* about experimentation in geography teaching with robust evidence of its validity and reliability. The measuring instrument can be used to examine different aspects of prospective geography teachers' *beliefs* about experimentation and to compare them across different subgroups. With its focus on the specific requirement of experimentation and on geography teachers in particular, the developed questionnaire can serve as an example for the subject- and requirement-specific assessment of the *beliefs* of (prospective) teachers.

To justify its use for further research and practice, we included several sources of evidence as recommended by Brown (2015) to support the construct validity of the developed measurement instrument on *beliefs* about experimentation. Since the dimensionality of the measurement instrument was previously unknown, in a previous step of scale development, EFA was used to empirically uncover the underlying factor structure. In the process, five factors related to potentials, learning preconditions, and the implementation of experimentation in geography teaching were found. Within the scope of the present study, the hypothesised factor structure was replicated in a new independent sample using CFA to support the validity of the measurement instrument (Brown, 2015). We found a good fit of the model to the data in the new sample, which indicates good construct validity of the measurement instrument. The final instrument consisted of five different scales and 28 items with acceptable to good reliability.

It should be noted that, because our model initially only reasonably fit the data, we modified our model in a data-driven manner using modification indices to determine the nature of the misfit of the initial hypothesised model. To avoid a discriminant validity problem, two items with cross-loadings had to be removed from the measurement instrument (Hair et al., 2014) after additional conceptual consideration (Brown, 2015). Thus, we have moved away from our confirmatory approach, which is nevertheless "of less consequence than not knowing the reason behind poor model fit" (Worthington & Whittaker, 2006, p. 829). However, the deletion of only two of the 30 variables can be considered a minor modification (Hair et al., 2014). Ideally, the revised model should be retested in a confirmatory approach using an independent sample (Brown, 2015).

The measurement instrument is intended to be administered within different subgroups and to be used for comparisons across them. Therefore, we evaluated measurement invariance as "an important aspect of test development" (Brown, 2015, p. 4) for the final model to ensure the generalisability of the construct across those groups. The results indicate strong measurement invariance across gender, school type (primary vs. secondary school), practical experience and further studied STEM subject, suggesting equivalent measurement properties in the subgroups. Thus, it can be assumed that group differences in the mean values can be attributed to actual differences in construct expression. The instrument is therefore suitable for mean value comparisons between these subgroups (Brown, 2015).

To further administer construct validity, we conducted correlation analysis to obtain first evidence of convergent and discriminant validity. The correlations found indicate the convergent validity of the instrument, as factors of theoretically overlapping constructs (resp., relating to potentials or learning preconditions, addressing the students directly) were moderately correlated, as hypothesised. Since "reliability is also an indicator of convergent validity … meaning that the measures all consistently represent the same latent construct" (Hair et al., 2014, p. 619), and all indicators related to theoretically overlapping constructs significantly loaded on the respected factors (Brown, 2015), this, too, can be seen as first evidence of the instrument's convergent validity. In addition, the correlations between all scales were modest ($r < .80$), indicating that the scales measure distinct constructs. Furthermore, only statistically nonsignificant or small correlations were found between factors related to theoretically distinct constructs, suggesting the overall favourable discriminant validity of the measurement instrument (Brown, 2015).

A central limitation of our procedure for capturing the prospective teachers' *beliefs* is the format of the developed instrument, namely the closed answer format and the categories that were determined in advance based on qualitative interview studies and literature research. As little knowledge is available to date on *beliefs* about experimentation in geography education, this open approach was necessary for the development of the items. Despite a closed answer format, to emphasise the subjective and subconscious character of the *beliefs*, we chose a suitable item stem for the survey of *beliefs* according to Rokeach (1969). In addition, we requested quick and intuitive responses to the items in the instructions. By choosing this answer format, it must be accepted that only excerpts of *beliefs* about experimentation can be captured by the questionnaire.

At the same time, the chosen closed questionnaire format makes it possible to conduct larger-scale studies as well as to capture the *beliefs* about experimentation in a time-efficient way. Thus, the measurement instrument can possibly serve in future for the diagnosis of the *beliefs* of geography student teachers as learning preconditions for university teaching and training programmes. This is of particular importance, as the student teachers' pre-existing *beliefs* act as a kind of filter, influencing the interpretation and implementation of what is learned in academic education (Maier et al., 2013). Since the measurement instrument can additionally enable an introspective access to one's own *beliefs*, it could furthermore be used for didactic (self-)reflection. This is especially relevant since just such methods that include an active involvement with one's own *beliefs* seem to be most profitable for the change of *beliefs* (Blömeke, 2003; Fives & Buehl, 2012; Steinmann & Oser, 2012). We also consider it possible to transfer an adapted questionnaire to other prospective science teachers, so that the questionnaire could also be suitable for teacher training in these areas. Whereas

experimentation in geography lessons differs from other science subjects in terms of context, in essence it is equivalent to the conceptual understanding in these, such as the goal of identifying cause-and-effect relationships.
In addition to only capturing excerpts of the *beliefs* about experimentation in geography classrooms, the developed measurement instrument for the *beliefs* initially focuses on only one sub-component of teachers' professional competence, even though our long-term goal is to examine further components. Against this background, it is not possible to make any reliable predictions about the (future) teaching activities of the student teachers on the basis of the results presented. Especially since, as outlined in the beginning, further factors would have to be taken into account for this. Finally, the high level of agreement with individual items and a possible ceiling effect should be viewed critically, especially with regard to the possible use of the instrument to examine changes in *beliefs*. In order to assess this, a comparison with other groups of in-service geography teachers will be necessary in future. It is therefore planned to replicate our study and validate the questionnaire with a sample of in-service teachers. This will allow us, on the one hand, to examine the generalisability of our findings beyond student teachers, and, on the other, to gain insights into geography teachers' professional actions when experimenting. The development and validation of the questionnaire on the *beliefs* of geography student teachers presented in this paper can thus be seen as a first and at the same time necessary step to be able to clarify the complex interplay of aspects of the professional competencies, school framework conditions, and classroom actions of geography teachers regarding experimentation in geography classrooms.

6 References

Aka, E. I. (2016). An Investigation into Prospective Science Teachers' Attitudes towards Laboratory Course and Self-Efficacy Beliefs in Laboratory Use. *International Journal of Environmental and Science Education*, *11*(10), 3319–3331.

Baumert, J., & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift Für Erziehungswissenschaft*, *9*(4), 469–520. https://doi.org/10.1007/978-3-658-00908-3_13

Bevins, S., Price, G., & Booth, J. (2019). The I files, the truth is out there: science teachers' constructs of inquiry. *International Journal of Science Education*, *41*(4), 533–545. https://doi.org/10.1080/09500693.2019.1568605

Blömeke, S. (2003). *Lehrerausbildung - Lehrerhandeln - Schülerleistungen: Perspektiven nationaler und internationaler empirischer Bildungsforschung.* Antrittsvorlesung. Berlin: Humboldt Universität.

Blömeke, S., Müller, C., Felbrich, A., & Kaiser, G. (2008). Epistemologische Überzeugungen zur Mathematik. In S. Blömeke, G. Kaiser, & R. Lehmann (Eds.), *Professionelle Kompetenz angehender Lehrerinnen und Lehrer: Wissen, Überzeugungen und Lerngelegenheiten deutscher Mathematikstudierender und -referendare. Erste Ergebnisse zur Wirksamkeit der Lehrerausbildung* (pp. 219–246). Waxmann.

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2. edition). *Methodology in the social sciences*. The Guilford Press.

Buehl, M. M., & Beck, J. S. (2015). The relationship between teachers' beliefs and teachers' practices. In H. Fives & M. G. Gill (Eds.), *Educational psychology handbook series. International handbook of research on teachers' beliefs* (pp. 66–84). Routledge.

Bussfeld, J. (2013). *Chancen und Barrieren des Einsatzes von experimentellen Arbeitsweisen im Geographieunterricht: Eine explorative Interviewstudie mit Geographielehrerinnen und -lehrern der Realschule* [unpublished master thesis]. Westfälische Wilhelms-Universität Münster, Münster.

Chen, F. F. (2007). Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(3), 464–504. https://doi.org/10.1080/10705510701301834

Cheung, D. (2011). Teacher Beliefs about Implementing Guided-Inquiry Laboratory Experiments for Secondary School Chemistry. *Journal of Chemical Education*, *88*, 1462–1468. https://doi.org/10.1021/ed1008409

Crawford, B. A. (2007). Learning to teach science as inquiry in the rough and tumble of practice. *Journal of Research in Science Teaching*, *44*(4), 613–642. https://doi.org/10.1002/tea.20157

DiBiase, W., & McDonald, J. R. (2015). Science Teacher Attitudes Toward Inquiry-Based Teaching and Learning. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, *88*(2), 29–38. https://doi.org/10.1080/00098655.2014.987717

Fang, S.-C. (2021). Towards Scientific Inquiry in Secondary Earth Science Classrooms: Opportunities and Realities. *International Journal of Science and Mathematics Education*, *19*(4), 771–792. https://doi.org/10.1007/s10763-020-10086-6

Fives, H., & Buehl, M. M. (2012). Spring cleaning for the "messy" construct of teachers' beliefs: What are they? Which have been examined? What can they tell us? In K. R. Harris (Ed.), *APA handbooks in psychology. APA educational psychology handbook, Volume 2: Individual differences and cultural and contextual factors* (1st ed., pp. 471–499). American Psychological Association. https://doi.org/10.1037/13274-019

Fögele, J., Luber, L., & Mehren, R. (2020). Types of student teachers between theory and practice in the subject of geography: reconstructing orientations using the example of a seminar on experiments. *Journal of Geography in Higher Education*, *44*(2), 310–333. https://doi.org/10.1080/03098265.2019.1698524

Gavora, P., & Wiegerová, A. (2019). Development of an instrument to measure preschool teachers´ beliefs in inquiry-based activities. *Journal of Baltic Science Education*, *18*(4), 558–568. https://doi.org/10.33225/jbse/19.18.558

Gimbel, K., Ziepprecht, K., & Mayer, J. (2018). Überzeugungen angehender Lehrkräfte fachspezifisch und inhaltsspezifisch operationalisieren und erfassen. In I. Glowinski, A. Borowski, J. Gillen, S. Schanze, & J. von Meien (Eds.), *Kohärenz in der universitären Lehrerbildung: Vernetzung von Fachwissenschaft, Fachdidaktik und Bildungswissenschaften* (pp. 179–198). Universitätsverlag Potsdam. https://publishup.uni-potsdam.de/files/41861/conference_kohaerenz_179-198.pdf

Graham, J. W. (2009). Missing data analysis: making it work in the real world. *Annual Review of Psychology*, *60*, 549–576. https://doi.org/10.1146/annurev.psych.58.110405.085530

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2014). *Multivariate data analysis* (7. Auflage, Pearson new internat. ed.). *Pearson custom library*. Pearson.

Hemmer, I., & Hemmer, M. (2002). Mit Interesse lernen: Schülerinteresse und Geographieunterricht. *Geographie Heute, 202*, 2–7.

Hemmer, I., & Hemmer, M. (Eds.). (2010). *Schülerinteresse an Themen, Regionen und Arbeitsweisen des Geographie-unterrichts: Ergebnisse der empirischen Forschung und deren Konsequenzen für die Unterrichtspraxis*. Selbstverlag des Hochschulverbandes für Geographie und ihre Didaktik e.V.

Hemmer, I., & Hemmer, M. (2021). Das Interesse von Schülerinnen und Schülern an geographischen Themen, Regionen und Arbeitsweisen – ein Bundeslandvergleich zwischen Bayern und Nordrhein-Westfalen. *Zeitschrift Für Geographiedidaktik*, *49*(3), 68–89. https://doi.org/10.18452/10.18452/22789

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., & Rosseel, Y. (2018). *semTools* (Version 0.5-1) [Computer software].

Kremer, K., Möller, A., Arnold, J., & Mayer, J. (2019). Kompetenzförderung beim Experimentieren. In J. Groß, M. Hammann, P. Schmiemann, & J. Zabel (Eds.), *Biologiedidaktische Forschung: Erträge für die Praxis* (pp. 113–128). Springer Spektrum. https://doi.org/10.1007/978-3-662-58443-9_7

Kuhl, J., Moser, V., Schäfer, L., & Redlich, H. (2013). Zur empirischen Erfassung von Beliefs von Förderschullehrerinnen und -lehrern. *Empirische Sonderpädagogik,* (1), 3–24. https://doi.org/10.25656/01:8907

Kunter, M., Klusmann, U., Baumert, J., Richter, D., Voss, T., & Hachfeld, A. (2013). Professional competence of teachers: Effects on instructional quality and student development. *Journal of Educational Psychology*, *105*(3), 805–820. https://doi.org/10.1037/a0032583

Labudde, P., & Möller, K. (2012). Stichwort: Naturwissenschaftlicher Unterricht. *Zeitschrift für Erziehungswissenschaft*, *15*(1), 11–36. https://doi.org/10.1007/s11618-012-0257-0

Lethmate, J. (2006). Experimentelle Lehrformen und Scientific Literacy. *Praxis Geographie*, *11*, 4–11.

Leuchter, M., Pauli, C., Reusser, K., & Lipowsky, F. (2006). Unterrichtsbezogene Überzeugungen und handlungsleitende Kognitionen von Lehrpersonen. *Zeitschrift für Erziehungswissenschaft*, *9*(4), 562–579. https://doi.org/10.1007/s11618-006-0168-z

Lucero, M., Valcke, M., & Schellens, T. (2013). Teachers' Beliefs and Self-Reported Use of Inquiry in Science Education in Public Primary Schools. *International Journal of Science Education*, *35*(8), 1407–1423. https://doi.org/10.1080/09500693.2012.704430

Lüdtke, O., Robitzsch, A., Trautwein, U., & Köller, O. (2007). Umgang mit fehlenden Werten in der psychologischen Forschung. Probleme und Lösungen. *Psychologische Rundschau*, *58*, 103–117. https://doi.org/10.1026/0033-3042.58.2.103

Maier, M. F., Greenfield, D. B., & Bulotsky-Shearer, R. J. (2013). Development and validation of a preschool teachers attitudes and beliefs toward science teaching questionnaire. *Early Childhood Research Quarterly*, *28*(2), 366–378. https://doi.org/10.1016/j.ecresq.2012.09.003

Mansour, N. (2009). Science Teachers' Beliefs and Practices: Issues, Implications and Research Agenda. *International Journal of Environmental and Science Education*, *4*(1), 25–48.

Marshall, J. C., Horton, R., Igo, B. L., & Switzer, D. M. (2009). K-12 Science and Mathematics Teachers' Beliefs About and Use of Inquiry in the Classroom. *International Journal of Science and Mathematics Education*, *7*(3), 575–596. https://doi.org/10.1007/s10763-007-9122-7

Mellado, V., Bermejo, M. L., Blanco, L. J., & Ruiz, C. (2007). The Classroom Practice of a Prospective Secondary Biology Teacher and His Conceptions of the Nature of Science and of Teaching and Learning Science. *International Journal of Science and Mathematics Education*, *6*(1), 37–62. https://doi.org/10.1007/s10763-007-9081-z

Miener, J. P., & Köhler, K. A. (2013). *Experimentelle Arbeitsweisen im Geographieunterricht. Vorstellungen von Geographielehrern zu Chancen und Barrieren*. AV Akademiker Verlag.

Nespor, J. (1987). The role of beliefs in the practice of teaching. *Journal of Curriculum Studies*, *19*(4), 317–328. https://doi.org/10.1080/0022027870190403

Oser, F., & Blömeke, S. (2012). Überzeugungen von Lehrpersonen. Einführung in den Thementeil. *Zeitschrift für Pädagogik*, *58*(4), 415–421. https://doi.org/10.25656/01:10405

Otto, K.-H. (2009). Experimentieren als Arbeitsweise im Geographieunterricht. *Geographie Und Schule*(180), 4–15.

Otto, K.-H., & Mönter, L. O. (2015). Scientific Literacy im Geographieunterricht fördern: Experimentelle Lehr-/Lernformen und Modellexperimente. *Geographie Heute*, *322*, 2–7.

Pajares, M. F. (1992). Teachers' Beliefs and Educational Research: Cleaning Up a Messy Construct. *Review of Educational Research*, *62*(3), 307–332. https://doi.org/10.3102%2F00346543062003307

Petermann, V., & Vorholzer, A. (2022). Relationship between Beliefs of Teachers about and Their Use of Explicit Instruction When Fostering Students' Scientific Inquiry Competencies. *Education Sciences*, *12*(9), 593. https://doi.org/10.3390/educsci12090593

Ramnarain, U. (2014). Teachers' perceptions of inquiry-based learning in urban, suburban, township and rural high schools: The context-specificity of science curriculum implementation in South Africa. *Teaching and Teacher Education*, *38*, 65–75. https://doi.org/10.1016/j.tate.2013.11.003

Ramnarain, U. (2016). Understanding the influence of intrinsic and extrinsic factors on inquiry-based science education at township schools in South Africa. *Journal of Research in Science Teaching*, *53*(4), 598–619. https://doi.org/10.1002/tea.21315

Ramnarain, U., & Hlatswayo, M. (2018). Teacher beliefs and attitudes about inquiry-based learning in a rural school district in South Africa. *South African Journal of Education*, *38*(1), 1–10. https://doi.org/10.15700/saje.v38n1a1431

Reusser, K., & Pauli, C. (2014). Berufsbezogene Überzeugungen von Lehrerinnen und Lehrern. In E. Terhart, H. Bennewitz, & M. Rothland (Eds.), *Handbuch der Forschung zum Lehrerberuf* (2nd ed., S. 642–661). Waxmann.

Richardson, V. (1996). The role of attitudes and beliefs in learning to teach. In J. Sikula (Ed.), *Handbook of Research on Teacher Education* (2nd ed., pp. 102–119). Macmillan.

Rokeach, M. (1969). *Beliefs, attitudes, and values: a theory of organization and change*. Jossey-Bass.

Rosendahl, N., Hemmer, M., & Schrüfer, G. (2020). Mit Vielfalt experimentieren: Professionalisierung angehender Lehrkräfte im GEO-Lehr-Lern-Labor. In R. Kürten, G. Greefrath, & M. Hammann (Eds.), *Begabungsförderung: Individuelle Förderung und Inklusive Bildung: Vol. 8. Komplexitätsreduktion in Lehr-Lern-Laboren: Innovative Lehrformate in der Lehrerbildung zum Umgang mit Heterogenität und Inklusion* (pp. 137–161). Waxmann.

Rosseel, Y. (2012). Lavaan : An R Package for Structural Equation Modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Saad, R., & BouJaoude, S. (2012). The Relationship between Teachers' Knowledge and Beliefs about Science and Inquiry and Their Classroom Practices. *Eurasia Journal of Mathematics, Science and Technology Education*, *8*(2), 113–128. https://doi.org/10.12973/eurasia.2012.825a

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the Fit of Structural Equation Models: Tests of Significance and Descriptive Goodness-of-Fit Measures. *Methods of Psychological Research Online*, *8*(2), 23–74.

Schubert, J. C. (2016). Kognitiv aktivierend und eigenständig experimentieren: Schülerinnen und Schüler erforschen das Wasserhaltevermögen von Böden. *Geographie Aktuell Und Schule*, *38*(219), 24–34.

Schubert, J. C., & Höhnle, S. (2016). Hindernisse für den Einsatz naturwissenschaftlicher Arbeitsweisen im Geographieunterricht aus Studierendenperspektive - Ausgewählte Ergebnisse einer empirischen Studie mit Lehramtsstudierenden. *GW-Unterricht*, *2-3*(142/143), 153–161.

Siorenta, A., & Jimoyiannis, A. (2008). Physics instruction in secondary schools: An investigation of teachers' beliefs towards physics laboratory and ICT. *Research in Science & Technological Education*, *26*(2), 185–202. https://doi.org/10.1080/02635140802037328

Steinmann, S., & Oser, F. (2012). Prägen Lehrerausbildende die Beliefs der angehenden Primarlehrpersonen? Shared Beliefs als Wirkungsgröße in der Lehrerausbildung. *Zeitschrift für Pädagogik*, *58*(4), 441–459. https://doi.org/10.25656/01:10388

Tabachnick, B. G., & Fidell, L. S. (2007). *Using Multivariate Statistics* (6. Auflage). Pearson.

Velling, H., Schubert, J. C., & Naujoks-Schober, N. (2022). Beliefs angehender Lehrpersonen zum Experimentieren im Geographieunterricht: Entwicklung und Pilotierung eines Fragebogens. *Zeitschrift für Didaktik der Naturwissenschaften*, *28*(1), 1–16. https://doi.org/10.1007/s40573-022-00148-3

Wallace, C. S., & Kang, N.-H. (2004). An investigation of experienced secondary science teachers' beliefs about inquiry: An examination of competing belief sets. *Journal of Research in Science Teaching*, *41*(9), 936–960. https://doi.org/10.1002/tea.20032

Wilhelmi, V. (2012). Die Experimentelle Lernform: Herausforderung des kompetenzorientierten Geographieunterrichts. *Praxis Geographie*, *7*(8), 4–8.

Wilkins, J. L. M. (2008). The relationship among elementary teachers' content knowledge, attitudes, beliefs, and practices. *Journal of Mathematics Teacher Education*, *11*(2), 139–164. https://doi.org/10.1007/s10857-007-9068-2

Worthington, R. L., & Whittaker, T. A. (2006). Scale Development Research. *The Counseling Psychologist*, *34*(6), 806–838. https://doi.org/10.1177/0011000006288127