



# European Journal of Educational Research

Volume 9, Issue 4, 1435 - 1453.

ISSN: 2165-8714

<http://www.eu-jer.com/>

## Arguments for Construct Validity of the Self-Efficacy Beliefs of Interdisciplinary Science Teaching (Self-ST) Instrument

Kevin Handtke\*

University of Goettingen, GERMANY

Susanne Bögeholz

University of Goettingen, GERMANY

Received: May 15, 2020 • Revised: July 20, 2020 • Accepted: August 26, 2020

**Abstract:** Current research on self-efficacy beliefs of interdisciplinary science teaching indicates shortcomings in facing recent teaching challenges in secondary education and corresponding valid instruments. Thus, we designed the Self-Efficacy Beliefs of Interdisciplinary Science Teaching (Self-ST) instrument based on a pedagogical content knowledge model for science teaching. We ensured the factorial validity of ten factors. To bring construct validity down to a round figure, we examined convergent and divergent validity in this paper. For answering the overall research question regarding arguments for the convergent and divergent validity of the interpretation of the Self-ST instrument's test values (and related hypotheses), we analyzed data of pre-service, trainee, and in-service biology, chemistry, and physics teachers ( $n = 590$ ) in a cross-sectional study. While the strong latent correlations of the ten Self-ST factors with self-efficacy beliefs of interdisciplinary science teaching in primary education ( $r = 0.40 - 0.63$ ,  $p < 0.01$ ) indicate convergent validity, the rather weak correlations with self-efficacy beliefs of general teaching ( $r = 0.17 - 0.54$ ,  $p < 0.01$ ), self-rated content knowledge in science ( $r = 0.13 - 0.40$ ,  $p < 0.01$ ), and perceived stress ( $r = -0.13 - -0.19$ ,  $p < 0.01$ ) support different divergent validity intensities. Thus, assumed relations within the nomological net surrounding the self-efficacy beliefs of interdisciplinary science teaching construct were confirmed for secondary education. In sum, we shed light on a rarely explored aspect of construct validity in science education research regarding self-efficacy beliefs. Doing so, we gained strong arguments that the Self-ST instrument's test values can serve as indicators of self-efficacy beliefs of interdisciplinary science teaching in secondary education.

**Keywords:** *Self-efficacy belief, science education, teacher education, construct validity, self-rated content knowledge.*

**To cite this article:** Handtke, K., & Bögeholz, S. (2020). Arguments for construct validity of the self-efficacy beliefs of interdisciplinary science teaching (Self-ST) instrument. *European Journal of Educational Research*, 9(4), 1435-1453. <https://doi.org/10.12973/eu-jer.9.4.1435>

### Introduction

Nowadays, interdisciplinary science teaching is a serious challenge because of possible shortcomings in formal education for prospective teachers (Carlson & Daehler, 2019). For example, German science teacher education programs for secondary education are mostly disciplinary, i.e., studies focus on biology, chemistry, and physics (Neumann et al., 2017). In contrast, lower secondary science education is interdisciplinary at comprehensive schools, e.g., in Lower Saxony (Niedersächsisches Kultusministerium, 2020). Thus, the discipline-specific focus of teacher education has led to criticisms that teacher education for interdisciplinary science is insufficient (Bröll & Friedrich, 2012) and needs to be improved (Dörge, 2001).

Shulman (2015, p. 9) emphasized that affective teaching components are highly important as they are connected with "what teachers 'know and do'". Baumert and Kunter (2013) also underpinned the importance of affective components, e.g., by integrating Motivational Orientations as one of four aspects of their model of teachers' professional competence. One affective construct within this Motivational Orientations is self-efficacy beliefs (Baumert & Kunter, 2013). This construct includes beliefs in one's skills to perform specific actions despite particular difficulties and obstacles (Bandura, 1997). Since self-efficacy beliefs serve to evaluate the success of teacher education (Forsthuber et al., 2011), measures of self-efficacy beliefs are informative for improving science teacher education.

Filling in a research gap, we developed a theory-based, literature-based, and curricular valid instrument for interdisciplinary science teaching in secondary education (grammar and comprehensive schools) that systematically

\* Corresponding author:

Kevin Handtke, Waldweg 26, 37073 Goettingen, Germany. ✉ [khandtk@gwdg.de](mailto:khandtk@gwdg.de)



considers the obstacle interdisciplinary science teaching (Handtke & Bögeholz, 2019a). The Self-Efficacy Beliefs of Interdisciplinary Science Teaching (SELF-ST) instrument operationalizes Park and Chen's (2012) pedagogical content knowledge (PCK) model for science teaching. An exploratory factor analysis ( $n = 114$ ) of a preliminary study revealed ten reliable factors that are important for interdisciplinary science teaching (Handtke & Bögeholz, 2019a). Subsequently, the factorial validity was confirmed using a confirmatory factor analysis investigating a new independent sample (Handtke & Bögeholz, 2019b).

The aim of this paper was to identify additional arguments supporting construct validity of the interpretation of the SELF-ST instrument's test values to gain further knowledge on and inform future users about the adequateness of the measure.

## Literature Review

### *Self-Efficacy Beliefs: Characteristics and Sources*

Self-efficacy beliefs are part of teachers' professional competence (Baumert & Kunter, 2013). They were defined by Bandura (1997) as beliefs in one's own ability to perform actions necessary for a demanded performance. Self-efficacy beliefs vary regarding the generality of the beliefs (e.g., overall or specific context), the level of the given task (e.g., easy or difficult), and the strength of the beliefs against problems (e.g., easy to impair or not) (Bandura, 1997). Due to these characteristics, self-efficacy beliefs differ for every task, scope, and difficulty and are context-specific (Park & Oliver, 2008; Tschannen-Moran et al., 1998). They benefit from being examined in terms of specific contexts instead of being examined only as a global construct (Bandura, 1997). Integrating demanding situations in the form of obstacles into the operationalization of self-efficacy beliefs is essential to contribute more variance in response behavior and overcome potential ceiling effects (Bandura, 1997, 2006). Also, self-efficacy beliefs refer to present skills, not future ones (Bandura, 1997).

According to Bandura (1997), mastery (direct) experience, vicarious (indirect) experience, verbal persuasion, and physiological and affective states are sources of information to derive self-efficacy beliefs. Bandura (1995) states that direct experience is the most powerful source, while humans "rely partly on their physiological and emotional states" (Bandura, 1995, p. 4). According to Schwarzer and Jerusalem (2002), these physiological and emotional states are the weakest sources of self-efficacy beliefs.

### *Importance of Self-Efficacy Beliefs*

Self-efficacy beliefs are key enabling factors for teachers, teaching, and students. For example, teachers' higher self-efficacy beliefs of science teaching resulted in ...

... teachers

- experiencing science as fun and interesting (de Laat & Watters, 1995),
- willing to better science teaching for (prospective) teachers and students (Ramey-Gassert et al., 1996),
- using more hands-on activities (Appleton & Kindt, 2002), and

... better student performance in school (Lumpe et al., 2012).

Also, teacher education courses had a positive impact on pre-service teachers' self-efficacy beliefs of science teaching (e.g., in primary education: Palmer et al., 2015), indicating their importance for teacher education.

### *Measures of Self-Efficacy Beliefs of Science Teaching*

#### *Previous Measures*

Handtke and Bögeholz (2019a) shed light on the need for a new measure of self-efficacy beliefs of interdisciplinary science teaching. To fill this gap, we operationalized self-efficacy beliefs of interdisciplinary science teaching considering three requirements: we responded to the request for a multidimensional approach (Bandura, 1997) by using a science teaching model as the theoretical base (Handtke & Bögeholz, 2019a). We addressed (evidence-based) best practice by designing a literature-based measure, and considered teacher and learner requirements by integrating standards and curricular-valid contents and competencies into the instrument (Handtke & Bögeholz, 2019a).

To identify the need for a new measure, we analyzed published measures referencing self-efficacy beliefs in the science context (biology, chemistry, physics, or science) (Handtke & Bögeholz, 2019a). It turned out that more measures exist for primary science education than for secondary science education (Handtke & Bögeholz, 2019a). Several instruments were built on the *Science Teaching Efficacy Belief Instrument* (STEBI; Enochs & Riggs, 1990; Riggs & Enochs, 1990) for primary education (Handtke & Bögeholz, 2019a), an influential and convenient measure in those days (for 25 years of history of the STEBI, see Deehan, 2017).

In Handtke and Bögeholz (2019a), we revealed that subsequently, several literature-based measures were developed (Mavrikaki & Athanasiou, 2011; Pruski et al., 2013; Rabe et al., 2012; Ritter, 1999; Roberts & Henson, 2000; Vidwans, 2016; Walan & Chang Rundgren, 2014); among them, only a few considered curricula requirements additionally (Mavrikaki & Athanasiou, 2011; Walan & Chang Rundgren, 2014). Only Vidwans' (2016) measure was driven by an existing model (theory-based), but it is rather generic and needs further factor analytic testing (Handtke & Bögeholz, 2019a). In addition, according to Handtke and Bögeholz (2019a), the vast majority of self-efficacy belief instruments in the context of teaching science (subjects) measured rather globally (Barros et al., 2010; Enochs & Riggs, 1990; Mavrikaki & Athanasiou, 2011; Pruski et al., 2013; Riese, 2009; Ritter, 1999; Roberts & Henson, 2000; Savran & Cakiroglu, 2001; Vidwans, 2016; Walan & Chang Rundgren, 2014); only one measure integrated obstacles into all items (Meinhardt et al., 2018).

### *Validation Types and Purposes*

When developing and evaluating a new instrument, validity is the most important quality criterion (Hartig et al., 2012). Different types of validity exist, such as content, criterion, and construct validity (DeVellis, 1991; Hartig et al., 2012; Moosbrugger & Kelava, 2012). While content validity focuses on the appropriate representation of the construct in the measure, as judged by experts (DeVellis, 1991; Hartig et al., 2012; Moosbrugger & Kelava, 2012), criterion validity tests the relation of a measure to external criteria that are relevant for diagnosis and practical issues (DeVellis, 1991; Hartig et al., 2012; Moosbrugger & Kelava, 2012). Criterion validity is important for diagnostic decisions or statements based on the test values of a measure (Hartig et al., 2012). Criterion validity also focuses on whether a test can be used to predict a specific behavior (Moosbrugger & Kelava, 2012).

Construct validity is the central validity aspect (Hartig et al., 2012), especially for new measures. Construct validity ensures that the interpretation of the test values of a measure is consistent with the intended construct (Hartig et al., 2012). A necessary prerequisite of construct validity is testing factorial validity (i.e., the factorial structure of a construct), such as through confirmatory factor analysis (Bühner, 2011; Hartig et al., 2012). However, arguments for factorial validity only are insufficient to assume construct validity because factorial validity makes no statement about the content of the factors (Hartig et al., 2012). According to Cronbach and Meehl (1955, p. 291), "to validate a claim that a test measures a construct, a nomological net surrounding the concept must exist." Thus, theoretical assumptions about the relations between the measured construct and other constructs are needed (Bühner, 2011; DeVellis, 1991; Hartig et al., 2012).

Construct validity is split into convergent and divergent validity (DeVellis, 1991; Hartig et al., 2012). Convergent validity expects a strong correlation with the same or a similar construct that shares decisive features with the investigated construct (Bühner, 2011; DeVellis, 1991; Hartig et al., 2012; Moosbrugger & Kelava, 2012). The aim of divergent validity is to distinguish the measured construct from validation constructs that are supposed to be different (Bühner, 2011; Moosbrugger & Kelava, 2012). Thus, divergent validity expects no correlation or a weak correlation with a different validation construct (Bühner, 2011; DeVellis, 1991; Hartig et al., 2012; Moosbrugger & Kelava, 2012). It is important that the validation constructs do not obviously measure something completely different from the construct under investigation (Bühner, 2011; Moosbrugger & Kelava, 2012). The validation construct should be related to the measured construct in some way (Bühner, 2011; Moosbrugger & Kelava, 2012). Some constructs are more divergent from the self-efficacy beliefs of interdisciplinary science teaching than others. Thus, we differentiate between different divergent validity intensities in this paper for enhanced precision. If convergent and divergent validity assumptions can be confirmed empirically, this supports the idea that the test values represent the intended construct (Hartig et al., 2012).

### *Validation Studies of Previous Measures*

Next, we present which validation types were used in previous measure developments concerning the self-efficacy beliefs of teaching science (subjects). Several studies integrated criterion validity arguments with, e.g., years of teaching experience or subject preference (Enochs & Riggs, 1990; Rabe et al., 2012; Riggs & Enochs, 1990; Ritter, 1999). Regarding construct validity, some studies only investigated arguments for factorial validity (Pruski et al., 2013; Riese, 2009; Roberts & Henson, 2000). Only a few researchers integrated divergent validity arguments with, e.g., self-efficacy beliefs of general teaching or outcome expectancy (Mavrikaki & Athanasiou, 2011; Rabe et al., 2012; Riggs & Enochs, 1990). To date, an examination of convergent validity has apparently not been integrated into any development of a self-efficacy beliefs of science (subjects) teaching measure. Overall, there has been little integration of divergent and (especially) convergent validity into measure development regarding self-efficacy beliefs of science (subjects) teaching to ensure the quality of the measures.

### *The SELF-ST Instrument*

Due to the desideratum, we designed a new science teaching-specific instrument for self-efficacy beliefs based on Park and Chen's (2012) PCK model. Handtke and Bögeholz (2019a) presented more information regarding other and more recent PCK models and regarding the decision for the chosen model to create the SELF-ST instrument. Only the

Orientations to Teaching Science category from the PCK model (Park & Chen, 2012) was not integrated into our measure; it was not considered reasonable for our purposes because as the only one it did not include competencies that could be learnt (for the detailed description of the measure development, see Handtke & Bögeholz, 2019a). We operationalized all other categories of the PCK model (Knowledge of Students' Understanding in Science, Knowledge of Science Curriculum, Knowledge of Assessment of Science Learning, and Knowledge of Instructional Strategies for Teaching Science) with its different subcategories (Handtke & Bögeholz, 2019a; Park & Chen, 2012). These categories cover central aspects of PCK of science teaching in order to diagnose (prospective) teachers' self-efficacy beliefs of interdisciplinary science teaching. We identified nine factors in line with the PCK model (and its subcategories) and one factor, additional to the PCK model, including teaching ethical issues (Handtke & Bögeholz, 2019a).

After an exploratory factor analysis ( $n = 114$ ) (Handtke & Bögeholz, 2019a), we confirmed the factorial validity of these ten factors of the self-efficacy beliefs of interdisciplinary science teaching with a confirmatory factor analysis ( $n = 590$ ) (RMSEA = 0.06, 90% confidence interval: 0.05 – 0.06, CFI = 0.95, TLI = 0.95; Handtke & Bögeholz, 2019b). All factors had satisfying values ( $> 0.7$ ) for Cronbach's Alpha (Handtke & Bögeholz, 2019b).

The ten factors of self-efficacy beliefs of interdisciplinary science teaching (Handtke & Bögeholz, 2019a) can be defined as follows:

- *Surveying Dimensions of Scientific Literacy* (Factor 1): This **science teaching-specific** factor comprises the ability to survey students' different scientific literacy dimensions in science education, such as scientific reasoning or experimenting (competence).
- *Applying Media* (Factor 2): This **generic teaching** factor comprises using and applying different media for teaching, such as interactive whiteboards, (e-)books, and digital media.
- *Teaching Ethically Relevant Issues of Applied Science* (Factor 3): This **science teaching-specific** factor comprises different facets of teaching ethical issues in applied natural sciences, including fostering students' handling of ethical complexity cumulatively and systematically.
- *Differentiated Fostering of Scientific Inquiry and Communication in Science* (Factor 4): This **science teaching-specific** factor comprises the cumulative and systematic fostering of students' competencies related to scientific inquiry and communication in science education and the consideration of different learning requirements when teaching science.
- *Using Subject-specific Materials in Science* (Factor 5): This **science teaching-specific** factor comprises using science-specific materials for teaching, such as curricula, education program materials or textbooks, and teachers' materials presented in journals.
- *Applying Natural Scientific Working Methods* (Factor 6): This **science teaching-specific** factor comprises applying and guiding typical working methods in science teaching, such as experimenting, using models, observing, and comparing.
- *Applying Methods of Evaluation* (Factor 7): This **generic teaching** factor comprises using summative and formative evaluation and choosing the right method by means of balancing advantages and disadvantages.
- *Considering Learning Difficulties and Needs of Students in Science* (Factor 8): This **science teaching-specific** factor comprises the consideration of different learning difficulties and needs of science education students regarding everyday ideas, mistakes when experimenting, difficulties with models, and students' interest and motivation.
- *Including Science-specific and General Instructional Strategies* (Factor 9): This partly **science teaching-specific**, partly **generic teaching** factor comprises applying rather general strategies in science teaching, i.e., the hypothetical-deductive method, activity and problem orientation, and the model of educational reconstruction.
- *Surveying and Fostering Natural Scientific Content Knowledge* (Factor 10): This **science teaching-specific** factor comprises fostering content knowledge in science cumulatively and systematically and surveying such knowledge as a part of scientific literacy in science education.

#### *Nearby Constructs of Self-Efficacy Beliefs of Interdisciplinary Science Teaching*

Besides PCK, one important facet of professional competence is content knowledge (Baumert & Kunter, 2013). Regarding biology, researchers found that content knowledge impacted the PCK of teachers (Käpylä et al., 2009). Schoon and Boone (1998) showed that content knowledge in science of pre-service primary education teachers was positively associated with self-efficacy beliefs of science teaching. The same relation was found for self-rated content knowledge in science ( $r = 0.37, p < 0.01$ ), investigating pre- and in-service primary education teachers (Velthuis et al., 2014; Yangin & Sidekli, 2016).

When identifying nearby self-efficacy beliefs, the context specificity feature (Park & Oliver, 2008; Tschannen-Moran et al., 1998) of self-efficacy beliefs plays a major role. Self-efficacy beliefs focusing on the same subject (e.g., science teaching) are more related to each other than self-efficacy beliefs of teaching different subjects or with different scopes.

One source of self-efficacy beliefs is physiological and affective states (Bandura, 1995). People interpret these states, such as a pounding heart, “as signs of vulnerability to dysfunction” (Bandura, 1997, p. 106). For example, such states occur in stressful situations (Bandura, 1997). Thus, perceived stress seems to have a negative influence on self-efficacy beliefs (Bandura, 1997). Research revealed that the higher the self-efficacy beliefs of general teaching are, the lower are teachers’ burnout and perceived stress (Schmitz, 1999). Self-efficacy beliefs are a protective factor against job stress; they predicted job stress negatively, and job stress predicted burnout positively (Schwarzer & Hallum, 2008). In addition, the correlation between self-efficacy beliefs of general teaching and job stress showed clear negative results but without consistent strength for two teacher subgroups:  $r_{\text{Syrian}} = -0.25$  and  $r_{\text{German}} = -0.52$  (Schwarzer & Hallum, 2008). In sum, research has identified a negative relationship in both directions. On the one hand, stress is a negative source of self-efficacy beliefs (Bandura, 1997). On the other hand, self-efficacy beliefs are a protective factor against stress (Schwarzer & Hallum, 2008).

### *Research Question and Hypotheses*

In their paper, Handtke and Bögeholz (2019a) pointed out the need for a new measure of self-efficacy beliefs of interdisciplinary science teaching. The new measure meets current science education (research) requirements (Handtke & Bögeholz, 2019a). The authors confirmed the factorial validity of the developed measure (Handtke & Bögeholz, 2019b). The ten resulted factors of self-efficacy beliefs of interdisciplinary science teaching (Handtke & Bögeholz, 2019a, 2019b) are mainly based on Park and Chen’s (2012) PCK model.

In addition to the arguments for factorial validity, the interpretation of the factors’ content needs examination (Hartig et al., 2012). Thus, this paper focuses on an in-depth examination of convergent and divergent validity, resulting in our overall research question:

*Research Question: To what extent is it possible to prove the convergent and divergent validity of the interpretation of the self-efficacy beliefs of interdisciplinary science teaching factors’ test values?*

Studies have shown that there is a positive relationship between self-efficacy beliefs of science teaching and self-rated content knowledge in science for primary education (Velthuis et al., 2014; Yangin & Sidekli, 2016). We transferred Yangin and Sidekli’s (2016) result ( $r = 0.37$ ,  $p < 0.01$ ) for primary education to our study. Thus, we assumed a correlation between self-efficacy beliefs of interdisciplinary science teaching and self-rated content knowledge in science for secondary education as well. In doing so, we expected similar correlations up to 0.4 as the different constructs at hand should correlate. However, it must be considered that, in our study, self-rated content knowledge consisted of three different factors: biology, chemistry, and physics (for empirical evidence, see Handtke & Bögeholz, 2020). In contrast, in Yangin and Sidekli’s (2016) study, self-rated content knowledge in science involved one factor. Thus, the correlations should not have exceeded 0.4, and correlations as low as 0.2 seemed possible as well due to the different scopes (science as a subject vs. science as biology, chemistry, and physics). In sum, we expected there to be a correlation between the constructs even though they were different. Thus, hypothesis 1 tested for divergent validity (Moosbrugger & Kelava, 2012).

Since some facets of self-efficacy beliefs of interdisciplinary science teaching and self-rated content knowledge may be more or less divergent than others, we differentiated between different intensities of divergent validity that led to an expected range of correlations:

*Hypothesis 1 (H1): The ten factors of self-efficacy beliefs of interdisciplinary science teaching in secondary education correlate positively with the self-rated content knowledge in biology, chemistry, and physics (between 0.2 – 0.4).*

Considering the context specificity of self-efficacy beliefs (Park & Oliver, 2008; Tschannen-Moran et al., 1998), self-efficacy beliefs of general teaching supposedly differ from self-efficacy beliefs of interdisciplinary science teaching in secondary education. However, they should relate as both constructs measure a type of self-efficacy beliefs. Thus, we expected correlations between 0.2 and 0.4 for the self-efficacy beliefs of general teaching, since this construct includes a greater portion of self-efficacy beliefs of interdisciplinary science teaching than self-rated content knowledge. However, compared to self-rated content knowledge in science subjects (biology, chemistry, and physics), self-efficacy beliefs of general teaching do not contain science content. Instead, they focus on general teaching. Thus, we expected correlations in the same range (from 0.2 to 0.4). Both self-efficacy belief constructs had different foci but should be related. Thus, hypothesis 2 tested for divergent validity (Moosbrugger & Kelava, 2012). In addition, the relation of the ten factors of self-efficacy beliefs of interdisciplinary science teaching to self-efficacy beliefs of general teaching could differ in intensity. As with hypothesis H1, we differentiated the intensities of divergent validity:

*Hypothesis 2 (H2): The ten factors of self-efficacy beliefs of interdisciplinary science teaching correlate positively with the self-efficacy beliefs of general teaching (between 0.2 – 0.4).*

Due to the context specificity (Park & Oliver, 2008; Tschannen-Moran et al., 1998), we also assumed a positive relation between self-efficacy beliefs of interdisciplinary science teaching in secondary education and in primary education. The two constructs were relatively similar. The context-specific primary education self-efficacy beliefs were even more similar to the self-efficacy beliefs in the same context in secondary education than the self-efficacy beliefs of general teaching or self-rated content knowledge in science.

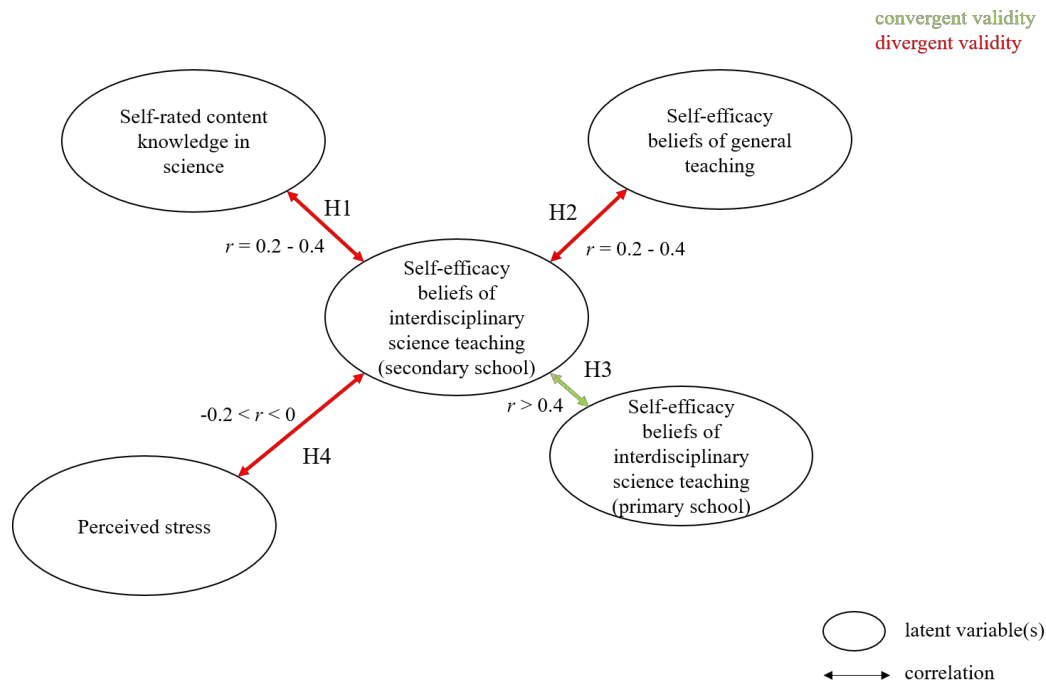
Between the two interdisciplinary science teaching self-efficacy beliefs, the construct self-efficacy beliefs and the teaching domain (science) were the same; only the education level differed. Thus, we expected correlations greater than 0.4. Considering the similarity of the two constructs, our third hypothesis examined convergent validity (Moosbrugger & Kelava, 2012):

**Hypothesis 3 (H3):** *The ten factors of self-efficacy beliefs of interdisciplinary science teaching in secondary education correlate positively with the self-efficacy beliefs of interdisciplinary science teaching in primary education (more than 0.4).*

According to the theory, perceived stress seems to be the weakest and a negative source of self-efficacy beliefs (Bandura, 1995, 1997; Schwarzer & Jerusalem, 2002). In addition, self-efficacy beliefs could be a protective factor against stress (e.g., Schmitz, 1999; Schwarzer & Hallum, 2008). As we measured general perceived stress (neither in terms of teaching nor science) and as research reported negative correlations but no clear result regarding the strength, we expected perceived stress to be a weak source of self-efficacy beliefs in line with the theory (Bandura, 1995, 1997; Schwarzer & Jerusalem, 2002). Thus, we assumed the correlations of self-efficacy beliefs of interdisciplinary science teaching with perceived stress would be less than 0 but greater than -0.2. As perceived stress and self-efficacy beliefs of interdisciplinary science teaching were different but (negatively) linked affective constructs (Bandura, 1997; Schmitz, 1999; Schwarzer & Hallum, 2008), our fourth hypothesis tested for divergent validity (Moosbrugger & Kelava, 2012):

**Hypothesis 4 (H4):** *The ten factors of self-efficacy beliefs of interdisciplinary science teaching in secondary education correlate negatively with perceived stress (greater than -0.2).*

To meet the requirement of construct validity for the interpretation of our measure's test values (Hartig et al., 2012), we tested the relations of four nearby constructs with the self-efficacy beliefs of interdisciplinary science teaching in secondary education (H1-4). Figure 1 displays the analyses conducted to test the relations of self-efficacy beliefs of interdisciplinary science teaching in secondary education to proximal and distal constructs within the nomological net (for general information on the nomological net, see Cronbach & Meehl, 1955). The figure highlights the specific validation purposes of the analyses and provides an overview of the expected correlations for each hypothesis.



*Figure 1: Overview of the analyses for testing the hypotheses – construct validation of the self-efficacy beliefs of interdisciplinary science teaching with nearby constructs. The lengths of the double arrows represent the distances.  $r$  = expected latent correlation(s).*

## Methodology

### Sample and Data Collection

In the present study,  $n = 590$  pre-service, trainee, and in-service teachers from five federal states in Germany took part (Table 1). We conducted a cross-sectional study design from December 2017 to December 2018 using a paper-pencil questionnaire. The study is part of a nationwide program to further develop German teacher education quality ("Qualitaetsoffensive Lehrerbildung"). Within the program, several studies took place. Thus, we did not establish the sample composition at random or systematically but had to select test persons also based on availability. However, we integrated test persons from different federal states, with different science education subjects studied, and from different phases of teacher education, ensuring the explanatory power of the sample.

Table 1: Sample composition ( $n = 590$ ). If all forms of a variable in total include more/less than 100% (e.g., 99.99% or 100.1%), the difference is due to missing values, "no indication", or rounding.

Variable	Percentage
Gender	
<i>Female</i>	57.63%
<i>Male</i>	41.86%
Federal state	
<i>Lower Saxony</i>	73.06%
<i>Others</i>	20.68%
Subject(s)	
<i>Biology</i>	28.81%
<i>Chemistry</i>	23.56%
<i>Physics</i>	16.78%
<i>Biology and chemistry</i>	24.24%
<i>Biology and physics</i>	1.19%
<i>Chemistry and physics</i>	5.08%
<i>Biology, chemistry, and physics</i>	0.17%
<i>Science</i>	0.17%
Phase of teacher education	
<i>Undergraduate</i>	48.64%
<i>Master of education</i>	23.39%
<i>State examination</i>	7.29%
<i>Trainee teacher</i>	11.53%
<i>In-service teacher</i>	6.44%
Type of education	
<i>Secondary education</i>	88.47%
<i>Another type of education</i>	7.29%

Almost three-fifths of the test persons were female, and three quarters studied or taught in Lower Saxony. The test persons studied at least one of the three science education subjects (Table 1). Nearly a quarter each studied only biology, only chemistry, or biology together with chemistry. One-sixth of the test persons studied only physics. Nearly half of the test persons were undergraduate students, a quarter was in a master's degree program for teacher education (two test persons had finished), and one-sixth were trainee or in-service teachers. The vast majority of the test persons studied to teach in secondary education ( $n = 522$ ).

### Measurement Instruments

In this section, we present the constructs in the nomological net (Figure 1) and their operationalization. At the center of interest were the self-efficacy beliefs of interdisciplinary science teaching in secondary education.

#### Self-Efficacy Beliefs of Interdisciplinary Science Teaching (SElf-ST) Instrument

We operationalized the construct self-efficacy beliefs of interdisciplinary science teaching in secondary education with the subscales of the SElf-ST instrument, first presented in Handtke and Bögeholz (2019a). The established four-point response scale (e.g., Schulte et al., 2011) ranged from "Is not right" (1) to "Is a little right" (2) to "Is rather right" (3) to "Is exactly right" (4) (Handtke & Bögeholz, 2019a, p. 8). The only change made after the Handtke and Bögeholz (2019a) study was that we integrated an initial situation description instead of the perennial footnote about the meaning of interdisciplinary science teaching (Handtke & Bögeholz, 2019a). The newly designed fictive situation asked the test persons: "Imagine that you currently teach at least one subject out of biology, chemistry, and physics. As of now, you are employed to teach interdisciplinary science at a comprehensive school. Please rate your skills for interdisciplinary science teaching below." The following is an example of an item of factor 3:



**“Even in natural scientific teaching [= obstacle], I can ...**

... consider students’ difficulties with ethically complex questions (for example, regarding the topics animal testing, climate change, atomic energy).” (Handtke & Bögeholz, 2019a, p. 8)

*Self-Rated Content Knowledge in Biology, Chemistry, and Physics*

For self-rated content knowledge in biology, chemistry, and physics, we used a three factorial instrument that considered Lower Saxony’s (Germany) curricula of the three science education subjects (Niedersächsisches Kultusministerium, 2015). The mentioned federal state curricula are a refined version of the National Educational Standards for students (e.g., for biology, see Kultusministerkonferenz, 2004). The development of the measure and its items are described in Handtke and Bögeholz (2020). One example, translated into English, is as follows: “*I know very much about the core idea of control and regulation (e.g., physiological regulation, such as regarding body temperature or hormones, and ecological interactions such as predator-prey relationship).*” (Handtke & Bögeholz, 2020, p. 66). We used a four-point response scale ranging from “Do not agree at all” (1) to “Do rather not agree” (2), to “Do rather agree” (3), to “Fully agree” (4) (Handtke & Bögeholz, 2020, p. 55).

*Multidimensional Scale of Self-Efficacy Beliefs of General Teaching*

The second validation construct comprised self-efficacy beliefs of general teaching. For operationalization, we used the 16-item short version of the Multidimensional Scale of Self-Efficacy Beliefs of General Teaching (response scale: same as described for self-efficacy beliefs of interdisciplinary science teaching in secondary education) (Schulte et al., 2011). One example, translated into English, is as follows: “*I know even for most different situations, how to apply various media appropriately for the situation*” (Schulte et al., 2011, p. 245; translated by the authors). The instrument contained five subscales based on standards for teacher education in the educational sciences: i) *Coping*, ii) *Communication and Conflict Resolution*, iii) *Diagnosis of Learning Conditions*, iv) *Performance Assessment*, and v) *Teaching* (Schulte et al., 2011; translated by the authors).

*Science Teaching Efficacy Belief Instrument-B*

The third validation construct was the self-efficacy beliefs of interdisciplinary science teaching in primary education. For operationalization, we used the *Personal Science Teaching Efficacy Belief Scale* of the STEBI-B for pre-service primary education teachers (Enochs & Riggs, 1990). It contained 13 items on a five-point response scale: “Strongly disagree” (1), “Disagree” (2), “Uncertain” (3), “Agree” (4), and “Strongly agree” (5) (Enochs & Riggs, 1990, p. 28). From the two scales of the STEBI-B (Enochs & Riggs, 1990), we applied the *Personal Science Teaching Efficacy Belief Scale* (e.g., “*I will continually find better ways to teach science.*”, Enoch & Riggs, 1990, p. 28). We chose this scale and not the *Science Teaching Outcome Expectancy Scale* (Enochs & Riggs, 1990) because we wanted to examine convergent validity. We translated the instrument into German and then had it independently back translated by an additional professional. We discussed discrepancies until a consensus was reached or until one translator could convince the other. Before analysis, we reversed the polarity of the eight negatively constructed items (Enochs & Riggs, 1990).

*Perceived Stress Scale*

To measure the general perceived stress as the fourth validation construct, we used Büssing’s (2011) German version of the 10-item version of the Perceived Stress Scale created by Cohen and Williamson (1988). The items asked for the perceived stress of the test persons within the last month, e.g., “*In the last month, how often have you been upset because of something that happened unexpectedly?*” (Cohen & Williamson, 1988, p. 65). The five-point response scale included “Never” (1), “Almost never” (2), “Sometimes” (3), “Fairly often” (4), and “Very often” (5) (Cohen & Williamson, 1988, pp. 64–65). We reversed the four negatively formulated items before data analysis (Cohen & Williamson, 1988). As recommended by Cohen and Williamson (1988), we considered the ten items as one factor.

*Procedure*

Regarding the administration of the questionnaire instruments, we first applied the measure of self-rated content knowledge (Handtke & Bögeholz, 2020), followed by the SELF-ST instrument (Handtke & Bögeholz, 2019a), the STEBI (Enochs & Riggs, 1990), the Multidimensional Scale of Self-Efficacy Beliefs of General Teaching (Schulte et al., 2011), and the Perceived Stress Scale (Büssing, 2011). We paid 135 pre-service teachers who participated in the study during free time within courses (e.g., of a practical laboratory course) or outside of courses. All other pre-service teachers participated during course time without payment. All trainee teachers received participation fees, whereas in-service teachers were not paid.

*Data Analysis*

We analyzed the data using *R* with *RStudio* (version 1.1463) and *lavaan* (version 0.6-3) and *psych* (version 1.8.12) packages. To investigate the hypotheses, the measurement models were specified (confirmatory factor analysis). As our



sample was much larger than the one used in Handtke and Bögeholz (2019a), we were able to complete the data computation in ordinal values.

Regarding the confirmatory factor analysis assumptions, we used the WLSMV estimator to handle the ordinal and non-normal data (Brown, 2006). The sample size was sufficient (Little, 2013), and we used at least three items per factor (Kline, 2011). Investigating possible outliers with boxplots, all answers given were possible and allowed. There was no hint on mistyped values. All the test persons were part of the target population because all studied at least one subject out of biology, chemistry, or physics to teach at a school. Thus, we did not delete test persons' answers that could have been possible outliers, as indicated by the boxplots (Flora et al., 2012). Removing these values would have resulted in no answers in the respective (mostly the lowest) response category. This would have skewed the results because some items that contained more than 30 permitted answers in this category would have been deleted. Also, specifying the measurement model without an answer on one response category (e.g., the lowest) would have been problematic; it would have caused additional, possibly problematic zero cells for the correlations (Jöreskog, 2004). In addition, large samples like our sample are less prone to the impact of potential outliers (Brown, 2006). Because of only about 1% missing values, we used pairwise deletion (Rosseel et al., 2020). We used the following guidelines from Little (2013) and Wheaton et al. (1977) as the minimum demand to rate our models' fit:  $\chi^2/df \leq 5$ ,  $CFI > 0.90$ ,  $TLI > 0.90$ ,  $RMSEA < 0.10$ .

We specified the measurement models of our SELF-ST instrument (ten factors), our measure of self-rated content knowledge (three factors), the Multidimensional Scale of Self-Efficacy Beliefs of General Teaching (five factors), the STEBI (one factor), and the Perceived Stress Scale (one factor). By specifying the measurement models of two constructs for each hypothesis, the correlations between the factors were calculated automatically by *lavaan*. Each of the hypotheses H1-4 was tested individually, and the *p*-values of the correlations of the directional hypotheses H1-4 were halved.

Besides the correlations of the single factors, we calculated means of the correlations (in the following: averaged correlations) of each validation construct factor with all self-efficacy beliefs of interdisciplinary science teaching factors. For example, we calculated the averaged correlation of *Coping* (as one factor of self-efficacy beliefs of general teaching) with all ten factors of self-efficacy beliefs of interdisciplinary science teaching, resulting in one averaged correlation instead of ten single correlations. In the end, each validation construct factor had one averaged correlation with self-efficacy beliefs of interdisciplinary science teaching based on the ten single correlations. This approach served as an approximation to provide better overviews. Using *psych*, we applied the Fisher *z*-transformation to transform all correlations of a validation construct factor with the ten factors of self-efficacy beliefs of interdisciplinary science teaching into *z*-values (Leonhart, 2013). After that, we computed the mean of these *z*-values and then retransformed the calculated mean of the *z*-values to a correlation (Leonhart, 2013). This correlation was the mean of the considered correlations (i.e., the averaged correlation), regardless of significance. Significance can only be indicated for the single correlations.

## Results

For hypotheses H1-4, we looked at the overall correlations of the self-efficacy beliefs of interdisciplinary science teaching in secondary education with the four nearby constructs. Table 2 shows that the majority of the correlations corresponded to our expectations without substantial deviations. We present these different correlations in detail below.

Table 2: Overview of correlations of self-efficacy beliefs of interdisciplinary science teaching in secondary education with nearby constructs.

Nearby constructs (number of factors)	Expected correlations	Empirical correlations of all factors (number of correlations)	Empirical averaged correlations (number of correlations)
H1: Self-rated content knowledge in science (3)	0.2-0.4	<u>0.13**</u> - 0.40** (30)	0.22 - 0.33 (3)
H2: Self-efficacy general teaching (5)	0.2-0.4	<u>0.17**</u> - <u>0.54**</u> (50)	0.23 - <u>0.45</u> (5)
H3: Self-efficacy science teaching primary (1)	> 0.4	<u>0.40**</u> - 0.63** (10)	0.54 (1)
H4: Perceived stress (1)	> -0.2, < 0	-0.13** - -0.19** (10)	-0.17 (1)

\*\* =  $p < 0.01$

Notes: The underlined values are outside the range of expected correlations. The grey column shows the averaged correlations.

### Hypothesis H1

The correlations of the ten factors of the self-efficacy beliefs of interdisciplinary science teaching in secondary education and the three factors of self-rated content knowledge in science ( $\chi^2(1691) = 3599.73$ ,  $p < 0.001$ , Ratio = 2.13, CFI = 0.96, TLI = 0.96, RMSEA = 0.04, 90% confidence interval = 0.04 - 0.05) were mostly in line with our expectations

(0.2 – 0.4) and ranged from 0.13 to 0.40 ( $p < 0.01$ ). The averaged correlation of each factor of self-rated content knowledge with the ten factors of self-efficacy beliefs of interdisciplinary science teaching was within the range of 0.2 to 0.4 (Table 3, grey column).

Six out of the 30 correlations were below the expected range (see underlined values in Table 3). The generic teaching factors, *Applying Media* (F2) and *Applying Methods of Evaluation* (F7), had rather low correlations with all factors of self-rated content knowledge in science ( $r = 0.15 - 0.25$ ,  $p < 0.01$ ), including three correlations below 0.2. *Teaching Ethically Relevant Issues of Applied Science* (F3) correlated strongest with the self-rated content knowledge in biology ( $r = 0.40$ ,  $p < 0.01$ ), also in the expected boundaries with self-rated content knowledge in chemistry ( $r = 0.24$ ,  $p < 0.01$ ), but below 0.2 with the self-rated content knowledge in physics ( $r = 0.13$ ,  $p < 0.01$ ; fourth deviation). The correlations of *Differentiated Fostering of Scientific Inquiry and Communication in Science* (F4) with the self-rated content knowledge in physics ( $r = 0.15$ ,  $p < 0.01$ ) and *Using Subject-specific Materials in Science* (F5) with the self-rated content knowledge in biology ( $r = 0.19$ ,  $p < 0.01$ ) were the remaining two correlations below 0.2 (fifth and sixth deviation).

Table 3: Correlation matrix of the ten factors of the self-efficacy beliefs of interdisciplinary science teaching in secondary education and the three factors of self-rated content knowledge in biology, chemistry, and physics ( $n = 590$ ).

	Ø	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10
srCK <sub>biology</sub>	<b>0.28</b>	0.27**	<u>0.15**</u>	0.40**	0.33**	<u>0.19**</u>	0.30**	0.20**	0.31**	0.28**	0.33**
srCK <sub>chemistry</sub>	<b>0.33</b>	0.37**	0.24**	0.24**	0.34**	0.38**	0.40**	<u>0.19**</u>	0.39**	0.31**	0.40**
srCK <sub>physics</sub>	<b>0.22</b>	0.26**	0.25**	<u>0.13**</u>	<u>0.15**</u>	0.31**	0.32**	<u>0.15**</u>	0.22**	0.23**	0.21**

\*\* =  $p < 0.01$

Notes: F1 = *Surveying Dimensions of Scientific Literacy*, F2 = *Applying Media*, F3 = *Teaching Ethically Relevant Issues of Applied Science*, F4 = *Differentiated Fostering of Scientific Inquiry and Communication in Science*, F5 = *Using Subject-specific Materials in Science*, F6 = *Applying Natural Scientific Working Methods*, F7 = *Applying Methods of Evaluation*, F8 = *Considering Learning Difficulties and Needs of Students in Science*, F9 = *Including Science-specific and General Instructional Strategies*, F10 = *Surveying and Fostering Natural Scientific Content Knowledge*. srCK = self-rated content knowledge. The underlined values are outside the range of 0.2 to 0.4. The grey column shows the srCK factors' averaged correlations with the ten factors of self-efficacy beliefs of interdisciplinary science teaching.

### Hypothesis H2

The correlations of the ten factors of the self-efficacy beliefs of interdisciplinary science teaching in secondary education and the five factors of self-efficacy beliefs of general teaching (SEGT) ( $\chi^2(1434) = 2794.47$ ,  $p < 0.001$ , Ratio = 1.95, CFI = 0.97, TLI = 0.96, RMSEA = 0.04, 90% confidence interval = 0.04 – 0.04) were mostly in line with our expectations (0.2 – 0.4) and ranged from 0.17 to 0.54 ( $p < 0.01$ ). Four of the five self-efficacy beliefs of general teaching factors' averaged correlations with the ten factors of self-efficacy beliefs of interdisciplinary science teaching were between 0.2 and 0.4 (Table 4, grey column). Only the averaged correlation of the factor *Teaching* (SEGT5) was greater than 0.4. In more detail, eight correlations of the ten factors of self-efficacy beliefs of interdisciplinary science teaching with *Teaching* (SEGT5) were greater than 0.4. Furthermore, the correlations of *Applying Methods of Evaluation* (F7) and *Including Science-specific and General Instructional Strategies* (F9) with *Performance Assessment* (SEGT4) were above the expected values (F7:  $r = 0.45$ ,  $p < 0.01$ ; F9:  $r = 0.46$ ,  $p < 0.01$ ). In addition to the ten higher correlations ( $> 0.4$ ), two correlations were remarkably low ( $< 0.2$ ) (Table 4, underlined values). The factor *Coping* (SEGT1) of self-efficacy beliefs of general teaching displayed the lowest correlations with the ten factors of the self-efficacy beliefs of interdisciplinary science teaching in secondary education, including the two correlations under 0.2.

Table 4: Correlation matrix of the ten factors of the self-efficacy beliefs of interdisciplinary science teaching in secondary education with the five factors of the self-efficacy beliefs of general teaching (SEGT1 – 5) ( $n = 590$ ).

	Ø	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10
SEGT1	<b>0.23</b>	0.24**	0.22**	0.24**	0.27**	0.21**	<u>0.19**</u>	0.26**	0.22**	0.31**	<u>0.17**</u>
SEGT2	<b>0.34</b>	0.25**	0.33**	0.35**	0.37**	0.32**	0.36**	0.35**	0.37**	0.40**	0.31**
SEGT3	<b>0.29</b>	0.26**	0.24**	0.30**	0.33**	0.22**	0.25**	0.28**	0.29**	0.31**	0.38**
SEGT4	<b>0.38</b>	0.38**	0.37**	0.36**	0.40**	0.34**	0.36**	<u>0.45**</u>	0.30**	<u>0.46**</u>	0.38**
SEGT5	<b>0.45</b>	0.38**	<u>0.53**</u>	<u>0.42**</u>	<u>0.49**</u>	<u>0.43**</u>	<u>0.46**</u>	<u>0.46**</u>	<u>0.41**</u>	<u>0.54**</u>	0.34**

\*\* =  $p < 0.01$

Notes: For the legend of F1-F10, see Table 3. SEGT = Self-efficacy beliefs of general teaching; SEGT1 = *Coping*, SEGT2 = *Communication and Conflict Resolution*, SEGT3 = *Diagnosis of Learning Conditions*, SEGT4 = *Performance Assessment*, SEGT5 = *Teaching*. The underlined values are outside the range of 0.2 to 0.4. The grey column shows the SEGT factors' averaged correlations with the ten factors of self-efficacy beliefs of interdisciplinary science teaching.

### Hypothesis H3

Third, we examined the correlations of the self-efficacy beliefs of interdisciplinary science teaching in secondary education with those in primary education ( $\chi^2(1322) = 2853.22, p < 0.001$ , Ratio = 2.16, CFI = 0.95, TLI = 0.95, RMSEA = 0.04, 90% confidence interval = 0.04 – 0.05). All correlations were at least 0.40 (Table 5) and ranged from 0.40 to 0.63 ( $p < 0.01$ ) with only one correlation just reaching 0.40. The averaged correlation (Table 5, grey column) supported our expectations. Four factors had correlations that were less than 0.5: *Surveying Dimensions of Scientific Literacy* (F1), *Applying Media* (F2), *Teaching Ethically Relevant Issues of Applied Science* (F3), and *Applying Methods of Evaluation* (F7) ( $r = 0.40 - 0.48, p < 0.01$ ). The other factors displayed correlations that were greater than 0.5 (F4, F5, F6, F8, F9, F10:  $r = 0.55 - 0.63, p < 0.01$ ).

Table 5: Correlation matrix of the ten factors of the self-efficacy beliefs of interdisciplinary science teaching in secondary education with the self-efficacy beliefs of interdisciplinary science teaching in primary education ( $n = 590$ ).

	$\emptyset$	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10
SEST	<b>0.54</b>	0.48**	0.43**	0.45**	0.56**	0.61**	0.63**	<u>0.40**</u>	0.63**	0.55**	0.61**

\*\* =  $p < 0.01$

Notes: For the legend of F1-F10, see Table 3. SEST: Self-efficacy beliefs of interdisciplinary science teaching in primary education. The underlined value is not greater than 0.4. The grey column shows the SEST factor's averaged correlation with the ten factors of self-efficacy beliefs of interdisciplinary science teaching.

### Hypothesis H4

Fourth, the correlations between the ten factors of the self-efficacy beliefs of interdisciplinary science teaching in secondary education and the factor of perceived stress ( $\chi^2(1169) = 2507.07, p < 0.001$ , Ratio = 2.14, CFI = 0.96, TLI = 0.96, RMSEA = 0.04, 90% confidence interval = 0.04 – 0.05) displayed low and negative correlations from -0.13 to -0.19 ( $p < 0.05$ ) (Table 6). Thus, the averaged correlation was also negative and greater than -0.2 (Table 6, grey column).

Table 6: Correlation matrix of the ten factors of the self-efficacy beliefs of interdisciplinary science teaching in secondary education with perceived stress ( $n = 590$ ).

	$\emptyset$	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10
PS	<b>-0.17</b>	-0.15**	-0.18**	-0.19**	-0.17**	-0.17**	-0.18**	-0.13**	-0.18**	-0.14**	-0.19**

\*\* =  $p < 0.01$

Notes: For the legend of F1-F10, see Table 3. PS = Perceived stress. The grey column shows the PS factor's averaged correlation with the ten factors of self-efficacy beliefs of interdisciplinary science teaching.

## Discussion

In this paper, we focus on the validation of the interpretation of the test values of the new theory-based, literature-based, and curricular valid Self-ST instrument. After revealing factorial validity (Handtke & Bögeholz, 2019a, 2019b) – a necessary prerequisite of construct validity (Hartig et al., 2012) – we added an important further step regarding the interpretation of the measure's test values. We examined the convergent and divergent validity of the interpretation of the ten factors test values' of our construct self-efficacy beliefs of interdisciplinary science teaching in-depth (in accordance with our research question). Thereby, we tested whether the test values can be considered to measure the intended construct (Hartig et al., 2012).

As shown in Table 2, most of the (averaged) correlations matched or only slightly deviated from our expectations. The results reflect the different similarities between the nearby constructs and the self-efficacy beliefs of interdisciplinary science teaching. Overall, these results provide strong arguments for the convergent and divergent validity of the interpretation of the Self-ST instrument's test values and the structure of the nomological net. We explain noteworthy or deviating correlations compared to our hypotheses below.

### Ad Hypothesis H1: Relation between Self-Efficacy Beliefs of Interdisciplinary Science Teaching and Self-Rated Content Knowledge

Regarding the ten factors of the self-efficacy beliefs of interdisciplinary science teaching and the three factors of self-rated content knowledge, according to our expectations, 80% of the correlations ranged from 0.20 to 0.40 ( $r = 0.13 - 0.40, p < 0.01$ ). Due to context specificity (Park & Oliver, 2008; Tschannen-Moran et al., 1998), the generic teaching factors, *Applying Media* (F2) and *Applying Methods of Evaluation* (F7), had rather low correlations with all factors of self-rated content knowledge, including three correlations that were less than 0.2 ( $r = 0.15 - 0.25, p < 0.01$ ).

The relatively strong correlation of self-rated content knowledge in biology with *Teaching Ethically Relevant Issues of Applied Science* (F3) ( $r = 0.40, p < 0.01$ ) compared to the correlation of physics with F3 is comprehensible. The biology curriculum offers various topics for socioscientific reasoning and decision making in school (e.g., risks of smoking,

sexuality, Sustainable Development) even in lower secondary education (Niedersächsisches Kultusministerium, 2015). Meanwhile, the physics curriculum states that opportunities to develop competencies regarding socioscientific reasoning and decision making are limited and complex (Niedersächsisches Kultusministerium, 2015).

The relatively low correlation of *Differentiated Fostering of Scientific Inquiry and Communication in Science* (F4) with the self-rated content knowledge in physics ( $r = 0.15, p < 0.01$ ) could also be explained by perceived limited opportunities to foster these competencies in physics teaching. Like socioscientific reasoning and decision making, scientific inquiry and communication are process-related competencies (Niedersächsisches Kultusministerium, 2015). Thus, they are comparable and could suffer from similar limitations. Perhaps the fostering of scientific inquiry and communication is somewhat less affected by self-rated content knowledge in physics, comparable to the aforementioned low correlation with teaching ethically relevant issues due to limited opportunities (Niedersächsisches Kultusministerium, 2015).

The other relatively low correlation of *Using Subject-specific Materials in Science* (F5) (e.g., education program materials, textbooks, or material of teacher education journals) with the self-rated content knowledge in biology ( $r = 0.19, p < 0.01$ ) could be explained by a perceived less difficulty of (self-rated) content knowledge in biology. The usage of science materials could require more (self-rated) content knowledge in chemistry and physics. Also, test persons' self-rated content knowledge in biology could be stronger than in physics or chemistry. Thus, understanding chemistry and physics would be judged more important to use science materials, resulting in higher correlations. Yilmaz-Tuzun (2008) supports this assumption by showing that pre-service elementary science teachers would rather teach biology (and earth science) concepts than physics and chemistry concepts.

Overall, the correlations of the self-efficacy beliefs of interdisciplinary science teaching with self-rated content knowledge in science are according to expectations (Velthuis et al., 2014; Yangin & Sidekli, 2016) and provide arguments for divergent validity. The results reveal different intensities of diverging factors. The reasons are the context specificity (Park & Oliver, 2008; Tschannen-Moran et al., 1998), normative guidelines (e.g., Niedersächsisches Kultusministerium, 2015), or the specific content of a factor. For example, while *Teaching Ethically Relevant Issues of Applied Science* (F3) correlated rather strongly with self-rated content knowledge in biology, the factor correlated clearly lower with the self-rated content knowledge in physics. Both correlations were plausible and, thus, reveal our in-depth examination of divergent validity. The former correlation (F3 with self-rated content knowledge in biology) supports a rather less intensively diverging factor, while the latter correlation (F3 with self-rated content knowledge in physics) reveals a more intensively diverging factor. This plausible difference in intensity of divergent validity could be revealed for correlations of different factors above, i.e., *Applying Media* (F2), *Applying Methods of Evaluation* (F7), *Differentiated Fostering of Scientific Inquiry and Communication in Science* (F4), and *Using Subject-specific Materials in Science* (F5). The findings provided arguments for the overall conclusion of hypothesis H1 of divergent validity with differing intensities of diverging factors.

#### *Ad Hypothesis H2: Relation between Self-Efficacy Beliefs of Interdisciplinary Science Teaching and Self-Efficacy Beliefs of General Teaching*

Regarding the ten factors of the self-efficacy beliefs of interdisciplinary science teaching and the five factors of the self-efficacy beliefs of general teaching, 76% of the single correlations ranged from 0.2 to 0.4, according to our expectations. The correlations of the general teaching factors with the self-efficacy beliefs of interdisciplinary science teaching factors varied in strength. Figure 2 shows a nomological net for hypothesis H2. The figure displays the distance between the five factors of self-efficacy beliefs of general teaching and those of interdisciplinary science teaching. The distances reflect the strength of the averaged correlations. We explain the differences and deviations and why the results indicate arguments for different intensities of divergent validity below.

The ten factors displayed relatively low correlations with *Coping* (SEGT1; mean = 0.23, Figure 2), which contains items regarding knowing and applying research about stress (Schulte et al., 2011). This content is not integrated into the self-efficacy beliefs of interdisciplinary science teaching in any form.

The correlations of the ten factors with *Communication and Conflict Resolution* (SEGT2) were also in the expected range (mean = 0.34, Figure 2). Knowledge about communication in situations with parents (Schulte et al., 2011) is not integrated into our measure. Communication regarding school and teaching situations (Schulte et al., 2011) is indirectly integrated into our measure. For example, communication is important for applying instructional strategies in teaching (Factor 9). Thus, there are partial overlaps.

*Diagnosis of Learning Conditions* (SEGT3; mean = 0.29, Figure 2) focuses on the intellectual giftedness or learning disorders of students (Schulte et al., 2011). Our measure of the self-efficacy beliefs of interdisciplinary science teaching mainly focuses on subject-specific challenges, such as science teaching-specific learning difficulties or the needs of students (Factor 8) instead of more general teaching challenges, as mentioned in Schulte et al. (2011). Thus, the correlations were between 0.2 and 0.4 and not even higher.

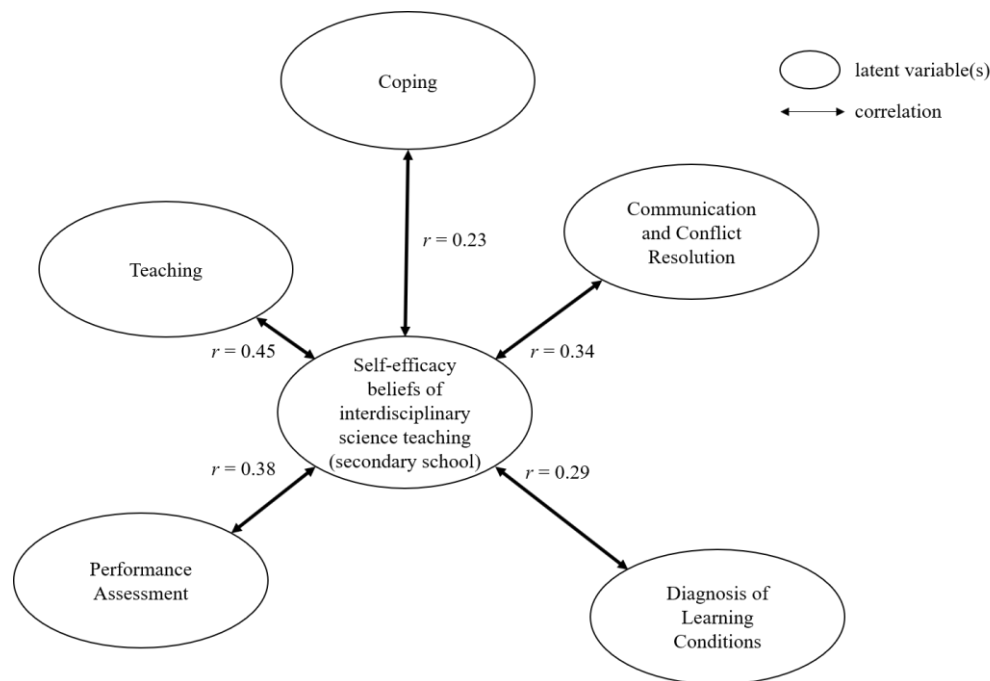


Figure 2: Overview of the averaged correlations of the ten factors of self-efficacy beliefs of interdisciplinary science teaching and the five factors of general teaching self-efficacy beliefs. The length of the double arrows represents the distance of the constructs.  $r$  = averaged correlation.

On average, the self-efficacy beliefs of interdisciplinary science teaching had the strongest correlations with *Performance Assessment* (SEGT4;  $r = 0.38$ ) and *Teaching* (SEGT5;  $r = 0.45$ ) due to several correlations above 0.4. The correlation of *Applying Methods of Evaluation* (F7) with *Performance Assessment* (SEGT4;  $r = 0.45$ ,  $p < 0.01$ ) is plausible. *Applying Methods of Evaluation* (F7) focuses on the usage of different evaluation methods and *Performance Assessment* (SEGT4) focuses on evaluation in general (e.g., performance assessment functions; Schulte et al., 2011). Thus, they take into account the same topic but have different foci and scopes. Therefore, they correlated relatively strongly but did not correlate even stronger. The correlation of *Performance Assessment* (SEGT4) with *Including Science-specific and General Instructional Strategies* (F9) ( $r = 0.46$ ,  $p < 0.01$ ) can be explained because a generic teaching factor correlated with a partly generic teaching factor, resulting in more related contexts (Park & Oliver, 2008; Tschannen-Moran et al., 1998). In addition, both have a more general scope on teaching, resulting in a somewhat higher correlation. For example, *Performance Assessment* (SEGT4) contains a general item about the functions of performance assessment when teaching (Schulte et al., 2011), and *Including Science-specific and General Instructional Strategies* (F9) contains a general item about activity and problem orientation. In sum, the correlations of the self-efficacy beliefs of interdisciplinary science teaching with *Performance Assessment* (SEGT4) can be explained.

Eight correlations of self-efficacy beliefs of interdisciplinary science teaching with *Teaching* (SEGT5) were above 0.4 ( $r = 0.41 - 0.54$ ,  $p < 0.01$ ). This result can be explained by the fact that *Teaching* (SEGT5) of the self-efficacy beliefs of general teaching includes items about, e.g., applying media and methods of cooperative or self-determined learning when teaching (Schulte et al., 2011). Comparable to that, the eight factors of the self-efficacy beliefs of interdisciplinary science teaching include items regarding teaching and its methods and media (e.g., applying media, methods, and instructional strategies and fostering competence). Due to the context specificity of self-efficacy beliefs (Park & Oliver, 2008; Tschannen-Moran et al., 1998) and because the context of the factors (general vs. science) is not always the same, and some factors are more similar than others, the correlation was only slightly above 0.4.

Overall, the correlations of self-efficacy beliefs of interdisciplinary science teaching with those of general teaching provide arguments for divergent validity. The explanations above reveal different intensities of diverging factors, as evident in the correlations (see Figure 2). The differences can be explained reasonably by the item content (Schulte et al., 2011), context specificity (Park & Oliver, 2008; Tschannen-Moran et al., 1998), or the scope of the items/factors. On the one hand, *Coping* (SEGT1) reflects content that is not integrated into our measure. That fact plausibly explains the low correlations. Based on this, we can argue for divergent validity with a more intensively diverging factor. On the other hand, *Teaching* (SEGT5) reflects content that is rather strongly integrated into our measure, primarily differing in terms of the science teaching context. That fact, in turn, explains the rather strong correlations. In light of this relation, we can argue for divergent validity with a less intensively diverging factor. As the other factors of self-efficacy beliefs of general teaching had plausible correlations with those of interdisciplinary science teaching as well, they also indicate different intensities of divergent validity. Thus, the findings support hypothesis H2 by providing arguments for divergent validity with differing intensities of diverging factors.

*Ad Hypothesis H3: Relation between Self-Efficacy Beliefs of Interdisciplinary Science Teaching in Secondary Education and Those in Primary Education*

Regarding the self-efficacy beliefs of interdisciplinary science teaching in secondary education (F1-10) and those in primary education (SEST), nearly all correlations were according to expectations. *Applying Media* (F2) and *Applying Methods of Evaluation* (F7) are generic teaching factors. Thus, they correlated lower ( $r = 0.40 - 0.43, p < 0.01$ ) than the (rather) science teaching-specific factors ( $r = 0.55 - 0.63, p < 0.01$ ). *Surveying Dimensions of Scientific Literacy* (F1) and *Teaching Ethically Relevant Issues of Applied Science* (F3) correlated only slightly higher ( $r = 0.45 - 0.48, p < 0.01$ ). With surveying scientific literacy and teaching ethics, they include topics that are not addressed in the STEBI (Enochs & Riggs, 1990).

In sum, the correlations of hypothesis H3 were as expected or explainable lower and plausible due to the construct measured in combination with its context specificity (Park & Oliver, 2008; Tschannen-Moran et al., 1998). The results indicate arguments for the convergent validity of the interpretation of our measured construct's test values because relatively high correlations confirmed the similarity of the constructs that only differed by the type of education.

*Ad Hypothesis H4: Relation between Self-Efficacy Beliefs of Interdisciplinary Science Teaching and Perceived Stress*

The low and negative correlations of the self-efficacy beliefs of interdisciplinary science teaching with perceived stress were according to our expectations (Bandura, 1995; Schwarzer & Jerusalem, 2002). Perceived stress was not measured in science- or teaching-specific ways in this study (Büssing, 2011), which makes the low correlations even more comprehensible. In sum, the results of hypothesis H4 display arguments for the divergent validity of the interpretation of the measure's test values because, even though the correlations were low, the results indicate that the two different affective constructs are related.

### Conclusion

The findings of our validation study are plausible. The observed deviations from our expectations are explainable by the literature and by concepts of operationalization of the constructs in the instruments applied in the validation study. Thus, the correlations with self-rated content knowledge, self-efficacy beliefs of general teaching, and perceived stress provide arguments in support of the different intensities of divergent validity (H1, H2, H4), and the correlations with the self-efficacy beliefs of interdisciplinary science teaching in primary education provide arguments in support of the convergent validity (H3) of the interpretation of the SELF-ST instrument's test values. Thus, the results indicate that the SELF-ST instrument measures the intended construct of self-efficacy beliefs of interdisciplinary science teaching in secondary education.

These results remarkably complement the measure development of Handtke and Bögeholz (2019a, 2019b). The paper at hand especially focused on the content of the measure and on ensuring that it measures the intended construct. We have provided new comprehensive arguments for construct validity regarding the location of the ten factors of self-efficacy beliefs of interdisciplinary science teaching in a structurally differentiated nomological net (overall: Figure 1, Table 2; hypothesis H2 in-depth: Figure 2).

Unlike previous measure developments, we examined convergent validity. We also checked the rarely investigated divergent validity (e.g., Rabe et al., 2012) regarding self-rated content knowledge in biology, chemistry, and physics, self-efficacy beliefs of general teaching and perceived stress in our validation study. Thus, we brought elaborated measure development down to a round figure unlike previous measure developments in this field. Besides the SELF-ST instrument's theory-based, literature-based, and curricular valid approach (Handtke & Bögeholz, 2019a), these additional results, complementing construct validity, are a further strong argument in support of the new measure of self-efficacy beliefs of interdisciplinary science teaching in secondary education at hand.

### Limitations

For some factors of the validation constructs (e.g., *Teaching* of the self-efficacy beliefs of general teaching), achieving a clear classification regarding divergent/convergent validity was complicated. The overall construct of self-efficacy beliefs of general teaching is divergent to the self-efficacy beliefs of interdisciplinary science teaching. However, this special factor *Teaching* had comprehensible overlaps with the self-efficacy beliefs of interdisciplinary science teaching. Nevertheless, convergent would not be the right classification, as convergent validity refers to a (very) similar or the identical construct (Moosbrugger & Kelava, 2012) with a high correlation (Hartig et al., 2012). Thus, we argued for different intensively diverging factors within a range. This approach facilitated a differentiated communication of our summarized results, which was close to the detailed findings.

As we applied multiple measures together at the same time and with the same test persons, common method bias could be an issue (e.g., Podsakoff et al., 2003). It was not possible to gain information from other sources than the test persons (Podsakoff et al., 2003) because all of the measures were subjective constructs about the test persons themselves.

We tried an *ex ante* approach to prevent a common method bias by other means. We never applied measures with the same response scale consecutively, resulting in varying scales; we ensured anonymity and that there would be no disadvantages for the test persons (to counter possible evaluation apprehension); and we asked for their answers for the purpose of enhancing teacher education and test persons' own training in the future (e.g., to reduce social desirability, ) (Podsakoff et al., 2003). Prior to three instruments, we integrated short descriptions to separate the instruments from each other for the test persons (Podsakoff et al., 2003). Moreover, our resulting correlations do not indicate a strong common method bias because correlations of different expected strengths exist, including low correlations. A powerful common method bias would have rather caused similar correlations between all measures (e.g., by choosing the same answers out of habit throughout the questionnaire).

Furthermore, the one-factorial model of perceived stress, which was recommended by Cohen and Williamson (1988), did not fit the data well in our study; only the CFI and TLI were acceptable ( $> 0.90$ ). We nevertheless applied it since it has been a frequently and for a long time used measure of perceived stress, but the results must be interpreted carefully.

A general effect of the participation fee can be denied. Nine of ten factors were not affected in a structural equation model with participation fee as a dummy-coded variable (0 = not paid, 1 = paid). Only the factor *Applying Media* (F2) seemed to be positively influenced by the participation fee to some degree ( $\beta = 0.15$ ,  $SE = 0.04$ ,  $p < 0.01$ ).

The sample mainly included pre-service teachers for practical reasons, such as that pre-service teachers were better accessible and that the most significant impact on improving teacher education is assumed to be possible in higher education at universities. Also, the most important aim of the "Qualitaetsoffensive Lehrerbildung" is to advance German teacher education at the university level.

### Suggestions

After in-depth validation, we confirmed that the SELF-ST instrument can be used for various purposes in the future. With its multiple dimensions, based on Park and Chen's (2012) PCK model, the instrument can be applied, e.g., for self-reflection among pre-service, trainee, and in-service teachers (Handtke & Bögeholz, 2019a). It gives them the opportunity to reflect on their actual abilities and allows them to gain insights into the requirements of interdisciplinary science teaching. For teacher education, the SELF-ST instrument can be applied to evaluate teacher education in general (Forsthuber et al., 2011; Handtke & Bögeholz, 2019a) or special trainings designed to prepare (prospective) teachers for interdisciplinary science teaching (e.g., Palmer et al., 2015).

In this paper, we presented an in-depth nomological net for (construct) validation of the SELF-ST instrument. Based on our results, it seems reasonable for future researchers to focus more on content-related questions. For example, Ngui and Lay (2020) examined the effect of student teachers' self-efficacy beliefs and other constructs on resilience and practicum stress, while Saputro et al. (2020) investigated the effect of problem-based learning on the self-efficacy beliefs of pre-service teachers. Comparable to these studies, it seems reasonable to examine influential factors on and the effects of self-efficacy beliefs of interdisciplinary science teaching in secondary education in the future. However, future researchers must keep in mind that the validity of an instrument needs to be repeatedly reviewed because it cannot be definitively proven (Hartig et al., 2012).

In addition, adapting the SELF-ST instrument and its theory-based approach to other countries or interdisciplinary subjects seems to be possible as well (Handtke & Bögeholz, 2019a). To use the SELF-ST instrument in other countries, adaptations regarding the country's particular curricula (i.e., especially F3, F4, and F10) or the obstacle could be necessary depending on the country's specific science teacher education approach (Handtke & Bögeholz, 2019a). For countries with discipline-specific science teacher education, such as Germany (Neumann et al., 2017), only an adaptation regarding the curricula could be sufficient (Handtke & Bögeholz, 2019a). For interdisciplinary science teacher education, an additional adaptation regarding the obstacle could be necessary because interdisciplinary science teaching could be an insufficiently challenging obstacle for these (prospective) teachers to overcome (Handtke & Bögeholz, 2019a). Adapting the measure to different interdisciplinary subjects and teaching tasks, such as Education for Sustainable Development, also seems possible (Handtke & Bögeholz, 2019a). For this teaching task, Park and Chen's (2012) PCK model could be transferred, and the obstacle could be adapted to teaching Education for Sustainable Development instead of interdisciplinary science.



### Acknowledgments

This project is part of the “Qualitaetsoffensive Lehrerbildung”, a joint initiative of the Federal Government and the *Laender*, which aims to improve the quality of teacher training. The programme is funded by the Federal Ministry of Education and Research (reference number 01JA1617). The authors are responsible for the content of this publication.



We acknowledge support by the Open Access Publication Funds of the University of Goettingen.

### References

- Appleton, K., & Kindt, I. (2002). Beginning elementary teachers' development as teachers of science. *Journal of Science Teacher Education*, 13(1), 43–61. <https://doi.org/10.1023/A:1015181809961>
- Bandura, A. (1995). Exercise of personal and collective efficacy in changing societies. In A. Bandura (Ed.), *Self-Efficacy in Changing Societies* (pp. 1–45). Cambridge University Press.
- Bandura, A. (1997). *Self-Efficacy: The exercise of control*. W. H. Freeman and Company.
- Bandura, A. (2006). Guide for constructing self-efficacy scales. In F. Pajares, & T. Urdan (Eds.), *Self-Efficacy Beliefs of Adolescents* (pp. 307–337). Information Age Publishing.
- Barros, M. A., Laburú, C. E., & da Silva, F. R. (2010). An instrument for measuring self-efficacy beliefs of secondary school physics teachers. *Procedia Social and Behavioral Sciences*, 2(2), 3129–3133. <https://doi.org/10.1016/j.sbspro.2010.03.476>
- 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: Results from the COACTIV Project* (pp. 25–48). Springer.
- Bröll, L., & Friedrich, J. (2012). Zur Qualifikation der Lehrkräfte für den NWA-Unterricht – eine Bestandsaufnahme in Baden-Wuerttemberg [About teachers' qualification to teach interdisciplinary science – An inventory in Baden-Wuerttemberg]. *STEM Journal/ MNU Journal*, 65(3), 180–186.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. The Guilford Press.
- Bühner, M. (2011). *Einführung in die Test- und Fragebogenkonstruktion* [Introduction into designing tests and questionnaires]. Pearson Studium.
- Büssing, A. (2011). *German perceived stress scale - Translation*. [https://www.cmu.edu/dietrich/psychology/stress-immunity-disease-lab/scales/.doc/german\\_pss\\_10.doc](https://www.cmu.edu/dietrich/psychology/stress-immunity-disease-lab/scales/.doc/german_pss_10.doc)
- Carlson, J., & Daehler, K. R. (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 Singapore.
- Cohen, S., & Williamson, G. M. (1988). Perceived stress in a probability sample of the United States. In S. Spacapan, & S. Oskamp (Eds.), *The Social Psychology of Health: The Claremont Symposium on Applied Social Psychology* (pp. 31–67). Sage Publications.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4), 281–302.
- de Laat, J., & Watters, J. J. (1995). Science teaching self-efficacy in a primary school: A case study. *Research in Science Education*, 25(4), 453–464. <https://doi.org/10.1007/BF02357387>
- Deehan, J. (2017). *The science teaching efficacy belief instruments (STEBI A and B): A comprehensive review of methods and findings from 25 years of science education research*. Springer International Publishing.
- DeVellis, R. F. (1991). *Scale development: Theory and applications*. Sage Publications.

- Dörge, A. (2001). Erfahrungen mit dem integrierten naturwissenschaftlichen Unterricht [Experiences with interdisciplinary science teaching]. *STEM Journal/ MNU Journal*, 54(4), 230–232.
- Enochs, L. G., & Riggs, I. M. (1990, April 8–11). *Further development of an elementary science teaching efficacy belief instrument: A preservice elementary scale* [Paper presentation]. 63rd meeting of the National Association of Research in Science Teaching, Atlanta, GA, United States.
- Flora, D. B., LaBrish, C., & Chalmers, R. P. (2012). Old and new ideas for data screening and assumption testing for exploratory and confirmatory factor analysis. *Frontiers in Psychology*, 3, 55. <https://doi.org/10.3389/fpsyg.2012.00055>
- Forsthuber, B., Horvath, A., de Almeida Coutinho, A. S., Motiejūnaitė, A., & Baidak, N. (2011). *Science education in Europe: National policies, practices and research*. Education, Audiovisual and Culture Executive Agency. <https://doi.org/10.2797/7170>
- Handtke, K., & Bögeholz, S. (2019a). Self-efficacy beliefs of interdisciplinary science teaching (SELF-ST) instrument: Drafting a theory-based measurement. *Education Sciences*, 9(4), 247. <https://doi.org/10.3390/educsci9040247>
- Handtke, K., & Bögeholz, S. (2019b, September 9–12). *Selbstwirksamkeitserwartungen zum Unterrichten von Naturwissenschaften: Messinstrumententwicklung und Einflussfaktoren* [Self-efficacy beliefs of interdisciplinary science teaching: Measurement development and influencing factors] [Presentation]. 22nd International Conference of the Association of the Didactics of Chemistry and Physics, and the Professional Division of the Didactics of Biology, Vienna, Austria.
- Handtke, K., & Bögeholz, S. (2020). Self-rated content knowledge of biology, chemistry, and physics – Developing a measure and identifying challenges for interdisciplinary science teaching. *Research in Subject-matter Teaching and Learning*, 3, 46–67. <https://doi.org/10.23770/rt1832>
- Hartig, J., Frey, A., & Jude, N. (2012). Validität [Validity]. In H. Moosbrugger, & A. Kelava (Eds.), *Testtheorie und Fragebogenkonstruktion* [Test theory and the design of questionnaires] (pp. 143–171). Springer.
- Jöreskog, K. G. (2004). *Structural equation modeling with ordinal variables using LISREL*. [https://www.researchgate.net/publication/254206917\\_Structural\\_Equation\\_Modeling\\_with\\_Ordinal\\_Variables\\_using\\_LISREL](https://www.researchgate.net/publication/254206917_Structural_Equation_Modeling_with_Ordinal_Variables_using_LISREL)
- Käpylä, M., Heikkinen, J.-P., & Asunta, T. (2009). Influence of content knowledge on pedagogical content knowledge: The case of teaching photosynthesis and plant growth. *International Journal of Science Education*, 31(10), 1395–1415. <https://doi.org/10.1080/09500690802082168>
- Kline, R. B. (2011). *Principles and practice of structural equation modeling*. The Guilford Press.
- Kultusministerkonferenz (Ed.). (2004). *Bildungsstandards im Fach Biologie fuer den Mittleren Schulabschluss: Beschluss vom 16.12.2004* [National educational standards for biology teaching regarding the intermediate school-leaving qualification: Resolution from 16th December 2004]. [https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen\\_beschluesse/2004/2004\\_12\\_16-Bildungsstandards-Biologie.pdf](https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2004/2004_12_16-Bildungsstandards-Biologie.pdf)
- Leonhart, R. (2013). *Lehrbuch Statistik: Einstieg und Vertiefung* [Textbook statistics: Entry and deepening]. Hans Huber.
- Little, T. D. (2013). *Longitudinal structural equation modeling*. The Guilford Press.
- Lumpe, A., Czerniak, C., Haney, J., & Belyukova, S. (2012). Beliefs about teaching science: The relationship between elementary teachers' participation in professional development and student achievement. *International Journal of Science Education*, 34(2), 153–166. <https://doi.org/10.1080/09500693.2010.551222>
- Mavrikaki, E., & Athanasiou, K. (2011). Development and application of an instrument to measure Greek primary education teachers' biology teaching self-efficacy beliefs. *Eurasia Journal of Mathematics, Science & Technology Education*, 7(3), 203–213. <https://doi.org/10.12973/ejmste/75197>
- Meinhardt, C., Rabe, T., & Krey, O. (2018). Formulierung eines evidenzbasierten Validitätsarguments am Beispiel der Erfassung physikdidaktischer Selbstwirksamkeitserwartungen mit einem neu entwickelten Instrument [Establishing an evidence-based validity argument for assessing self-efficacy beliefs for teaching physics with a newly developed instrument]. *German Journal of Science Education/ Zeitschrift fuer Didaktik der Naturwissenschaften*, 24(1), 131–150. <https://doi.org/10.1007/s40573-018-0079-6>
- Moosbrugger, H., & Kelava, A. (2012). Qualitätsanforderungen an einen psychologischen Test (Testgütekriterien) [Quality requirements on a psychological test (Criteria of test quality)]. In H. Moosbrugger, & A. Kelava (Eds.), *Testtheorie und Fragebogenkonstruktion* [Test theory and the design of questionnaires] (pp. 7–26). Springer.

- Neumann, K., Härtig, H., Harms, U., & Parchmann, I. (2017). Science teacher preparation in Germany. In J. E. Pedersen, T. Isozaki, & T. Hirano (Eds.), *Model Science Teacher Preparation Programs: An International Comparison of What Works* (pp. 29–52). Information Age Publishing Inc.
- Ngui, G. K., & Lay, Y. F. (2020). The effect of emotional intelligence, self-efficacy, subjective well-being and resilience on student teachers' perceived practicum stress: A Malaysian case study. *European Journal of Educational Research*, 9(1), 277–291. <https://doi.org/10.12973/eu-jer.9.1.277>
- Niedersächsisches Kultusministerium (Ed.). (2020). *Kerncurriculum fuer die integrierte Gesamtschule Schuljahrgaenge 5-10: Naturwissenschaften* [Core curriculum of comprehensive school's school years 5-10: Interdisciplinary science]. <https://cuvo.nibis.de/cuvo.php?p=download&upload=234>
- Niedersächsisches Kultusministerium (Ed.). (2015). *Kerncurriculum fuer das Gymnasium Schuljahrgaenge 5-10: Naturwissenschaften* [Core curriculum of grammar school's school years 5-10: Interdisciplinary science]. <https://cuvo.nibis.de/cuvo.php?p=download&upload=18>
- Palmer, D., Dixon, J., & Archer, J. (2015). Changes in science teaching self-efficacy among primary teacher education students. *Australian Journal of Teacher Education*, 40(12), 27–40. <https://doi.org/10.14221/ajte.2015v40n12.3>
- Park, S., & Chen, Y.-C. (2012). Mapping out the integration of the components of pedagogical content knowledge (PCK): Examples from high school biology classrooms. *Journal of Research in Science Teaching*, 49(7), 922–941. <https://doi.org/10.1002/tea.21022>
- Park, S., & Oliver, J. S. (2008). Revisiting the conceptualisation of pedagogical content knowledge (PCK): PCK as a conceptual tool to understand teachers as professionals. *Research in Science Education*, 38(3), 261–284. <https://doi.org/10.1007/s11165-007-9049-6>
- Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *The Journal of Applied Psychology*, 88(5), 879–903. <https://doi.org/10.1037/0021-9010.88.5.879>
- Pruski, L. A., Blanco, S. L., Riggs, R. A., Grimes, K. K., Fordtran, C. W., Barbola, G. M., Cornell, J. E., & Lichtenstein, M. J. (2013). Construct validation of the self-efficacy teaching and knowledge instrument for science teachers-revised (SETAKIST-R): Lessons Learned. *Journal of Science Teacher Education*, 24(7), 1133–1156. <https://doi.org/10.1007/s10972-013-9351-2>
- Rabe, T., Meinhardt, C., & Krey, O. (2012). Entwicklung eines Instruments zur Erhebung von Selbstwirksamkeitserwartungen in physikdidaktischen Handlungsfeldern [Teachers' self efficacy beliefs for teaching physics: Development of a new instrument]. *German Journal of Science Education/ Zeitschrift fuer Didaktik der Naturwissenschaften*, 18, 293–315.
- Ramey-Gassert, L., Shroyer, M. G., & Staver, J. R. (1996). A qualitative study of factors influencing science teaching self-efficacy of elementary level teachers. *Science Education*, 80(3), 283–315. [https://doi.org/10.1002/\(SICI\)1098-237X\(199606\)80:3%3C283::AID-SCE2%3E3.0.CO;2-A](https://doi.org/10.1002/(SICI)1098-237X(199606)80:3%3C283::AID-SCE2%3E3.0.CO;2-A)
- Riese, J. (2009). *Professionelles Wissen und professionelle Handlungskompetenz von (angehenden) Physiklehrkraeften* [(Prospective) physics teachers' professional knowledge and professional competence]. Logos.
- Riggs, I. M., & Enochs, L. G. (1990). Toward the development of an elementary teacher's science teaching efficacy belief instrument. *Science Education*, 74(6), 625–637. <https://doi.org/10.1002/sce.3730740605>
- Ritter, J. M. (1999). *The development and validation of the self-efficacy beliefs about equitable science teaching and learning instrument for prospective elementary teachers* [Doctoral dissertation, Pennsylvania State University]. [https://etda.libraries.psu.edu/files/final\\_submissions/2881](https://etda.libraries.psu.edu/files/final_submissions/2881)
- Roberts, J. K., & Henson, R. K. (2000, November 17–19). *Self-Efficacy Teaching and Knowledge Instrument for Science Teachers (SETAKIST): A Proposal for a New Efficacy Instrument* [Paper presentation]. 28th Meeting of the Mid-South Educational Research Association, Bowling Green, KY, United States.
- Rosseel, Y., Jorgensen, T. D., Rockwood, N., Oberski, D., Byrnes, J., Vanbrabant, L., Savalei, V., Merkle, E., Hallquist, M., Rhemtulla, M., Katsikatsou, M., Barendse, M., & Scharf, F. (2020). *Package 'lavaan'*. <https://cran.r-project.org/web/packages/lavaan/lavaan.pdf>
- Saputro, A. D., Atun, S., Wilujeng, I., Ariyanto, A., & Arifin, S. (2020). Enhancing pre-service elementary teachers' self-efficacy and critical thinking using problem-based learning. *European Journal of Educational Research*, 9(2), 765–773. <https://doi.org/10.12973/eu-jer.9.2.765>
- Savran, A., & Cakiroglu, J. (2001). Preservice biology teachers' perceived efficacy beliefs in teaching biology. *Hacettepe University Journal of Education/ Hacettepe Universitesi Egitim Fakultesi Dergisi*, 21, 105–112.

- Schmitz, G. S. (1999). *Zur Struktur und Dynamik der Selbstwirksamkeitserwartungen von Lehrern: Ein protektiver Faktor gegen Belastung und Burnout?* [Structure and dynamics of teacher self-efficacy: A protective factor against strain and burnout?] [Doctoral dissertation, Free University of Berlin]. [http://www.diss.fu-berlin.de/diss/receive/FUDISS\\_thesis\\_000000000315](http://www.diss.fu-berlin.de/diss/receive/FUDISS_thesis_000000000315)
- Schoon, K. J., & Boone, W. J. (1998). Self-Efficacy and alternative conceptions of science of preservice elementary teachers. *Science Education*, 82(5), 553–568. [https://doi.org/10.1002/\(SICI\)1098-237X\(199809\)82:5<553::AID-SCE2>3.0.CO;2-8](https://doi.org/10.1002/(SICI)1098-237X(199809)82:5<553::AID-SCE2>3.0.CO;2-8)
- Schulte, K., Watermann, R., & Bögeholz, S. (2011). Ueberpruefung der faktoriellen Validitaet einer multidimensionalen Skala der Lehrer-Selbstwirksamkeitserwartung [Examination of the factorial validity of a multidimensional scale of self-efficacy beliefs of general teaching]. *Empirical Pedagogy/ Empirische Paedagogik*, 25(2), 232–256.
- Schwarzer, R., & Hallum, S. (2008). Perceived teacher self-efficacy as a predictor of job stress and burnout: Mediation analyses. *Applied Psychology*, 57, 152–171. <https://doi.org/10.1111/j.1464-0597.2008.00359.x>
- Schwarzer, R., & Jerusalem, M. (2002). Das Konzept der Selbstwirksamkeit [The concept of self-efficacy]. *Journal of Pedagogy/ Zeitschrift fuer Paedagogik*, 44. Supplement/ Beiheft: *Selbstwirksamkeit und Motivationsprozesse in Bildungsinstitutionen* [Self-efficacy and motivation processes in educational institutions], 28–53.
- Shulman, L. S. (2015). PCK: Its genesis and exodus. In A. Berry, P. Friedrichsen, & J. Loughran (Eds.), *Teaching and learning in science series. Re-examining pedagogical content knowledge in science education* (pp. 3–13). Routledge.
- Tschannen-Moran, M., Woolfolk Hoy, A., & Hoy, W. K. (1998). Teacher efficacy: Its meaning and measure. *Review of Educational Research*, 68(2), 202–248. <https://doi.org/10.3102/00346543068002202>
- Velthuis, C., Fisser, P., & Pieters, J. (2014). Teacher training and pre-service primary teachers' self-efficacy for science teaching. *Journal of Science Teacher Education*, 25(4), 445–464. <https://doi.org/10.1007/s10972-013-9363-y>
- Vidwans, M. (2016). *Exploring science teachers' self-efficacy perceptions to teach in Ontario's diverse classrooms: A mixed-methods investigation* [Doctoral dissertation, University of Western Ontario]. <https://ir.lib.uwo.ca/cgi/viewcontent.cgi?article=5460&context=etd>
- Walan, S., & Chang Rundgren, S.-N. (2014). Investigating preschool and primary school teachers' self-efficacy and needs in teaching science: A pilot study. *CEPS Journal*, 4(1), 51–67.
- Wheaton, B., Muthen, B., Alwin, D. F., & Summers, G. F. (1977). Assessing reliability and stability in panel models. *Sociological Methodology*, 8, 84–136. <https://doi.org/10.2307/270754>
- Yangin, S., & Sidekli, S. (2016). Self-efficacy for science teaching scale development: Construct validation with elementary school teachers. *Journal of Education and Training Studies*, 4(10), 54–69. <https://doi.org/10.11114/jets.v4i10.1694>
- Yilmaz-Tuzun, O. (2008). Preservice elementary teachers' beliefs about science teaching. *Journal of Science Teacher Education*, 19(2), 183–204. <https://doi.org/10.1007/s10972-007-9084-1>