Article

Self-Efficacy Beliefs of Interdisciplinary Science Teaching (SElf-ST) Instrument: Drafting a Theory-Based Measurement

Kevin Handtke *[ID] and **Susanne Bögeholz**

Department of Biology Education, University of Goettingen, Waldweg 26, 37073 Goettingen, Germany
* Correspondence: khandtk@gwdg.de

check for updates

Abstract: Interdisciplinary science teaching is an issue in various countries. One example in Europe is Germany, especially regarding comprehensive schools. At the same time, German teacher education is primarily subject-specific. An examination of data on self-efficacy beliefs is helpful for understanding the qualifications of teachers for interdisciplinary science. Previous measurement instruments for teaching biology, chemistry, physics, and science lack a literature-based, theory-based, or curricular-valid measurement or a systematic obstacle to overcome. Thus, to meet these requirements, this research developed a draft for a new instrument to measure self-efficacy beliefs of interdisciplinary science teaching (SElf-ST). As the theoretical base, the instrument operationalizes a model of pedagogical content knowledge for teaching science and adapts it to self-efficacy beliefs. In a cross-sectional study ($N = 114$ pre-service and trainee teachers), a ten-factor-solution for self-efficacy beliefs resulted from an exploratory factor analysis (Kaiser–Meyer–Olkin-criterion = 0.858, α = 0.70–0.86). Nine factors are linked to the theoretical model. An additional tenth factor emerged: Teaching Ethically Relevant Issues. Nine factors show low and medium correlations with teaching experience. Eight factors show at least low correlations with self-rated content knowledge in no less than one of the three subjects. In general, science-specific factors show rather low or medium correlations, and generic factors (e.g., Applying Media, and Applying Methods of Evaluation) show low or no correlations. This result is in accordance with the context specificity of self-efficacy beliefs. These results meet most of the research expectations and provide initial indications of the concurrent, curricular, and divergent validity of the SElf-ST instrument. The paper argues for the development of a new, theory-based instrument to measure self-efficacy beliefs of interdisciplinary science teaching.

Keywords: self-efficacy beliefs; science; interdisciplinary teaching; measurement instrument development; teacher education; pedagogical content knowledge

1. Introduction

"Teachers may be asked to teach a subject they have not formally studied", Carlson and Daehler [1] (p. 91) recently claimed, particularly regarding science. Shulman [2] similarly argues, regarding primary teachers content knowledge, that science teaching is a serious challenge, for which teachers are not prepared. Thus, science teaching and its preparation show to be a serious challenge. Nations that offer disciplinary and interdisciplinary science in secondary education could be particularly affected by this issue. In Europe, for example, Spain, France, and Germany are countries that offer both types of science teaching in lower secondary education [3].

In this research, we investigate the situation in Germany, where interdisciplinary science teaching is a concrete requirement for pre- and in-service teachers not only but particularly at comprehensive schools (e.g., [4]). Since 2007, the number of secondary modern schools from class 5 to 9 ("Hauptschule")

decreased by approximately 50% and from class 5 to 10 ("Realschule") by approximately 30% [5,6]. Simultaneously, the number of comprehensive schools ("Integrierte Gesamtschule") more than tripled (cf. [5,6]).

In particular, comprehensive schools usually offer the interdisciplinary subject science (i.e., biology, chemistry, and physics as one subject) in lower secondary education (e.g., [4]), which follows Labudde's [7] approach at the level of the timetable. *Integrated* is the term for this form of interdisciplinary teaching [7]. Below, we focus on this level and call it *interdisciplinary teaching*. Consequently, the minimum requirement for pre- and in-service teachers is competence in teaching biology, chemistry, and physics. However, to date, (prospective) teachers mostly seem to be only slightly prepared for the challenges of interdisciplinary teaching in Germany.

For a teacher education based upon research evidence and best practices, it is important to know self-efficacy beliefs of interdisciplinary science teaching because they can provide insights into the learning outcomes of teacher education [3]. The most relevant measurement instrument for science teaching is the primary education Science Teaching Efficacy Beliefs Instrument (STEBI) [8]. However, the STEBI includes few findings of the current educational research and practice—core curricula, standards for teaching and teacher education, etc. Overall, primary schools are the focus of research regarding self-efficacy beliefs of science teaching (Section 1.1). In addition, theory-based (i.e., based on an existing model) measurement instruments are lacking (Section 1.1). Thus, a desideratum exists for a theory-based, internationally compatible measurement instrument suitable for the German teacher education. The new measurement instrument should include normative guidelines for teacher education at grammar and comprehensive schools as well as current research findings regarding teacher education. The measurement can be adapted to and used in other countries where science is taught either interdisciplinarily and disciplinarily in lower secondary education (e.g., France, Spain [3]). Moreover, with more extensive adjustments, it could monitor how effective teacher education is in countries with interdisciplinary science teaching only (e.g., Turkey, Italy [3]). The theory-based approach ensures testing for different facets of relevant teaching issues.

1.1. Theory of Self-Efficacy Beliefs, Empirical Findings, and Previous Measurement Instruments

Self-efficacy beliefs are part of Bandura's [9] social cognitive theory, which are the "beliefs in one's capabilities to organize and execute the courses of action required to produce given attainments." [9] (p. 3). Thus, they are of fundamental importance in regard to human action [9]. The level of a task, as well as the generality and strength of a belief, are distinguishers of self-efficacy beliefs [9]. Consequently, they vary depending on the specific situation and must receive context-specific consideration [10]. To ensure the measurement of self-efficacy beliefs is compliant with the theory, the utilization of obstacles is necessary for operationalizing demanding situations [9,11]. Distinct from but related to self-efficacy beliefs are outcome expectations and collective self-efficacy beliefs. Outcome expectations relate to judging the impacts of the attainments produced by one's behavior [9]. Thus, they assume the product of self-efficacy beliefs. Collective self-efficacy beliefs are the shared dynamic beliefs in a group's capability [9]. They are not just the total of everyone's self-efficacy beliefs, but they are "an emergent group-level attribute that is the product of coordinative and interactive dynamics" [9] (p. 7). On the one hand, individual self-efficacy beliefs focus on every (prospective) teachers' different characteristics. On the other hand, self-efficacy beliefs are useful to examine individual (prospective) teachers' beliefs in regard to science teaching within the current state of teacher education.

Self-efficacy beliefs are one aspect of professional competence (Motivational Orientations), such as pedagogical content knowledge (PCK) or content knowledge (CK) (Professional Knowledge) [12]. Attitudes and beliefs are capable of influencing the development of teachers' PCK [1]. Shulman [2] further highlights the importance of non-cognitive constructs, stating that "a lot of what teachers 'know and do' is connected to their own affective and motivation states" [2] (p. 9). Thus, non-cognitive aspects like self-efficacy beliefs should be of interest in teacher education due to their connection to the cognitive aspects (cf. [2]). This statement is supported, as various courses in teacher education affect

self-efficacy beliefs (e.g., science methods courses that address sources of self-efficacy beliefs [13], or science content courses [14]). Therefore, self-efficacy beliefs can be used for evaluation of interventions or coursework addressing CK or PCK as well (cf. [3]).

Various research findings emphasize the value of self-efficacy beliefs of science teaching. Many findings refer to primary school. Due to comprehensibility, we do not distinguish findings for science regarding primary and secondary education in this paper. Teachers with higher self-efficacy beliefs of science teaching are more open to new challenges [15]; their involvement is greater in organizations for teaching science, as well as in the development of curricula and standards for science [15]. Teachers with higher self-efficacy beliefs of science teaching are more confident and view science as interesting and fun [16]. In addition, they seem to apply hands-on-activities more often in the classroom [17] and seem to teach in a student-centered manner more often [16]. Teachers' self-efficacy beliefs of science teaching are a positive predictor for student achievement [18]. These impacts underline the importance of self-efficacy beliefs for teachers. Because self-efficacy beliefs can be influenced by courses in teacher education [13,14], they should receive consideration within teacher education to support professional development. To do so, the imperative to provide starting points for evidence-based teacher education necessitates an instrument to measure actual self-efficacy beliefs.

To expose any gaps, this research utilizes three categories: theory-based, literature-based, and curricular valid. An instrument is theory-based if it is operationalized from an existing model regarding (science) teaching. Bandura [9] recommended a multidimensional measurement of self-efficacy beliefs. Integrating a model for measuring self-efficacy beliefs should prevent global measurements and include various aspects of teaching. A similar theory-based approach can be found in an unpublished measurement instrument for biology teaching ([19], as the items are not published yet, it cannot be further examined). Established models can provide a solid base to operationalize self-efficacy beliefs. The aspects of the model can be further specified. The accordance with (science) teaching models allows for the monitoring of the learning outcomes within teacher education. An instrument is literature-based if its operationalization derives from research findings, which ensures current and best practices. An instrument is curricular valid if it expresses the requirements for teachers in the curriculum and standards for students or the standards for teacher education. This consideration ensures that the measurement aligns with the curriculum and sets forth requirements expected at a school or in teacher education (curricular validity [20]).

Overall, in the context of natural science subjects, more measurement instruments for teaching at the primary level exist than for teaching at the secondary level (Table A1). To examine multidimensionality, the scope of instruments is of particular interest (global or more specific, see [9]). The primary school measurement instrument STEBI [8] is the basis for many measurement instruments (Table A1; further information on the instruments). The STEBI was one of the first instruments to measure self-efficacy beliefs of science teaching. It allows global ratings of self-efficacy beliefs. It "was a useful tool in its time", but the dynamic changes in science education require new instruments [21] (p. 144). A more specific and multidimensional [9] scope could provide more specific insights. Subsequently, some measurement instruments utilized a literature-based conceptualization [22–25], while only a few researchers developed curricular-valid instruments [22,25].

New focuses on diversity [23] and content knowledge [22,24] were established. Based on the STEBI, these instruments [22–24] measure more globally and do not focus on different facets of specific science or biology teaching issues. Two instruments adapted the STEBI in a straightforward way for physics and biology as a brief global measurement [26,27]. The literature-based measurement instrument for teaching biology [22] would necessitate an adaption for science on a more specific level, which would include more facets of teaching. Moreover, as with the biology curriculum, concepts for chemistry and physics would be necessary. The literature-based as well as curricular valid instrument of Walan and Chang Rundgren [25], independent of the STEBI, integrates current research and curricula in Sweden for pre- and primary school in a more specific way. However, the researchers integrated only three of the various dimensions into the pre- and primary school instruments [25], which resulted

in a limited scope. The 3-point Likert-scale employed in the instrument [25] also provides only limited information. Despite constant enhancement (diverse focuses and increasing specificity), theory-based approaches are generally lacking in the development of a measurement instrument. The inclusion of obstacles during item construction (systematically) should also receive consideration.

Moreover, both deficits usually occur within research into self-efficacy beliefs for teaching biology, chemistry, physics, and science in secondary education (Table A2; further information on the instruments). Except for the rather theory-based approach of Vidwans [28], which needs more research concerning the empirical operationalization and the factor-analytic foundation of the theoretical components, the present measurement instruments do not seem to be theory-based [29–31]. Vidwans' instrument [28] newly focuses on culturally responsive pedagogy in science. However, it does not focus on the range of science-specific teaching issues [28]. The utilization of the STEBI and other instruments provides advancement for measuring self-efficacy beliefs of physics teaching [29]. Like the STEBI, the instrument is on a global level [29]. Pruski et al. [31] provide an improvement of the measurement instrument SETAKIST [24] with ambitious statistical modeling, but it remains a global measurement and suffers from psychometric issues. The most recent and independent measurement is for physics teaching [30]. It is more specific than previous instruments and integrates obstacles, albeit unsystematically [30]. It investigates four dimensions in-depth but could include a broader range of teaching issues [30]. The distinction between planning and performing in this instrument [30] is questionable (cf. [32]), and the competencies derive from different research projects, not based on one established model [33]. Especially, when integrating biology and chemistry, an adaption of the instrument to science would need to include more disciplinary aspects for teaching these subjects, e.g., instructional strategies. None of these measurement instruments for secondary education seem to be both literature-based and curricular valid. Researchers only recently incorporated obstacles into the measurement instrument for physics teaching [30] but relatively unsystematically. For interpreting single items or factors, systematical or standardized obstacles could be better.

Thus, an adaptation of existing instruments would be laborious. Only one instrument could be theory-based [28], while the majority of instruments produce rather global measurements [8,22–29,31], are not specific enough [30] or are for primary education [8,22–27]. Only one instrument [30] includes obstacles but in an alternative way and for one subject. No theory-based instrument with obstacles exists for more than one subject. Consequently, a new theory- and literature-based, as well as a curricular valid, measurement instrument, that systematically considers obstacles is required in the field of self-efficacy beliefs of science teaching at grammar and comprehensive schools. Moreover, for primary and secondary education the empirical factors from previous measurement instruments only range from one [8] to four content-related factors, which are distinguished in regard to planning and performance [30]. This limited scope further complicates the differentiated consideration of learning capabilities and successes of education.

1.2. Model of Pedagogical Content Knowledge for Teaching Science

Defining pedagogical content knowledge (PCK) and assuming its components have a long tradition in research (cf. [34] for a summary). This research began with Shulman [35,36] and continues today with the consensus [37] and refined consensus model [1]. Of the many models, we highlight one recent model that addresses the components of PCK for teaching science in detail, unlike the consensus [37] or refined consensus model [1], which is the pentagon model of PCK for teaching science [38]. The basis for this model derives from Magnusson et al. [39], which considers PCK in the context of science teaching. The hierarchic model [39] contains five categories with different subcategories. The four categories (Knowledge of Students' Understanding of Science, Knowledge of Science Curricula, Knowledge of Assessment of Scientific Literacy and Knowledge of Instructional Strategies) display an interaction with the superordinate category of Orientation to Teaching Science [39]. The model contains knowledge and beliefs for each category [39]. For our purposes, we focused on knowledge, like Park and Chen [38].

Based on empirical results, Park [40], Park and Oliver [41,42], and Park and Chen [38] developed the model from Magnusson et al. [39] into the pentagon model of PCK for teaching science [38]. For this, they qualitatively analyzed, among other things, classroom observations and interviews with teachers (e.g., [38]). The model from Park and Chen [38] includes five equivalent and relational categories with further subcategories:

1. Orientation to Teaching Science includes the beliefs of teachers regarding the goals and purposes of science teaching [41], derived from Grossman [43].
2. Knowledge of Students' Understanding in Science includes the subcategories of knowledge of students' Misconceptions, Learning Difficulties, Motivation, Interest, and Need [41].
3. Knowledge of Science Curriculum, derived from Grossman [43], comprises three subcategories: Curriculum Materials, Horizontal Curriculum, and Vertical Curriculum [41]. The Horizontal Curriculum knowledge includes the goals of the subject for the topics, and the Vertical Curriculum includes the sequence of the goals over the school time [39,43]. Curriculum Materials are curricular-valid materials for teaching specific topics [41,43].
4. Knowledge of Assessment of Science Learning, derived from Tamir [44], includes knowledge divided into two subcategories: Dimensions of Science Learning to Assess and appropriate Methods of Assessing Science Learning [41].
5. Knowledge of Instructional Strategies (and Representations) for Teaching Science contains three subcategories: knowledge of Subject/Science-specific Strategies, (Topic-specific Strategies:) Activities, and Representations [39,41]. Subject/Science-specific Strategies are general instructions of science teaching like conceptual change strategies [41]; Activities are natural scientific methods like experiments and Representations are, for example, models [39].

1.3. Relationship of Experience and Content Knowledge with Self-Efficacy Beliefs

According to Bandura, there are four sources of self-efficacy beliefs—direct (enactive mastery, respectively) experience, vicarious experience, verbal persuasion, and physiological and affective states [9]. Direct and positive practical experience is considered the most important source of self-efficacy beliefs [9]. In addition, practical experience in science teaching is a predictor ($\beta = 0.26$, controlled for gender) for self-efficacy beliefs of science teaching [18]. Considered in terms of groups with diverse levels of science teaching experience, more teaching experience relates to higher self-efficacy beliefs of science teaching (medium and strong effect size) [45]. These findings serve as examples to illustrate this relationship. No correlation with, or effect of, practical experience appeared, for example, with general teaching self-efficacy beliefs [46,47] or self-efficacy beliefs of science teaching (pre-post-design [48]).

Another meaningful part of teachers' professional competence, besides PCK, is content knowledge (CK) [12]. There are suggestions that having more natural scientific CK results in higher self-efficacy beliefs of science teaching [49]. In addition, research shows that self-rated CK (srCK) in natural sciences relates to the self-efficacy beliefs of science teaching [45,50].

1.4. Research Question and Hypotheses

We highlighted a need for a theory-based instrument to measure the self-efficacy beliefs of science teaching that meets the current challenges and the state of research. The divergent validity is determined by the relative weakness of the correlation of the factors, as they should be distinct (cf. [20]). From these findings, we derived the following research question to generate the first hypotheses about the factor structure and to investigate the divergent validity of the measurement instrument:

Which evidence can be identified to prove the subcategories based on the PCK model for self-efficacy beliefs of science teaching empirically?

To provide more evidence, we formulated two hypotheses to investigate possible indicators of the concurrent and divergent validity of the measurement instrument. The theoretical base [9], as well as several positive findings in the specific natural scientific context (e.g., [18,45]) resulted in the assumption

of a positive correlation between direct, practical science teaching experience and self-efficacy beliefs of science teaching. The correlation between the instrument and a criterion (measured at the same time) tests for concurrent validity [20]. The exemplary negative finding in the science context only surveyed advanced pre-service teachers using a pre-post-design [48]. Thus, the relatively high experience could explain the missing correlation. Therefore, the derivation of the following hypothesis regarding concurrent validity, as a type of criterion validity, excludes the latter finding.

Hypothesis 1 (H1). *Self-efficacy beliefs of science teaching and practical teaching experience are positively correlated.*

In research, self-efficacy beliefs of science teaching and srCK in science display a positive correlation [45,50]. Thus, we assumed a positive correlation of self-efficacy beliefs of science teaching with srCK of curricular-valid content of biology, chemistry, and physics teaching in Germany as well. Despite the expected correlation, both constructs are, nevertheless, substantially different. Thus, the correlations examine divergent validity, as the constructs should be correlated, but low, as they are distinct constructs [20]. These findings result in the following hypothesis regarding divergent validity.

Hypothesis 2 (H2). *Self-efficacy beliefs of science teaching and srCK in biology, chemistry, and physics are positively correlated.*

2. Materials and Methods

2.1. Sample

In a cross-sectional design, $N = 114$ participants (pre-service and trainee teachers) completed a survey between July and October 2017 as part of the "Qualitätsoffensive Lehrerbildung". The "Qualitätsoffensive Lehrerbildung" is a joint program between Germany and its federal states to improve teacher education. The "Qualitätsoffensive Lehrerbildung", among other things, seeks to address the issue of interdisciplinary (science) teaching. The list of selected universities resulted from the restrictions of numerous surveys of several universities within the nationwide program "Qualitätsoffensive Lehrerbildung". Restrictions were mandatory to ensure that the pre-service teachers would not receive an overload of testing. It was not possible to survey systematically every type of pre-service teacher at each of the universities. However, the sample is sufficient for the exploratory approach, as all necessary subjects and phases of teacher education at the university are addressed. Assuming a close middle effect (0.3) for the first hypothesis, calculations with G*Power (version 3.1.9.2; Heinrich Heine University Düsseldorf, Düsseldorf, Germany) showed, in advance, that 115 participants were necessary for the mathematically comparable Pearson's correlation. Thus, the present sample was sufficient for detecting substantial effects. A total of 62.3% of the test participants were women, 34.2% were men, and 4 provided no statement. The test participants were from Lower Saxony (44.7%), Rhineland-Palatinate (32.5%), North Rhine-Westphalia (15.8%), and Hesse (7.0%). A total of 36.8% studied biology, 7.0% chemistry, 14.9% physics, 26.3% earth science, 12.3% biology and chemistry, 0.9% chemistry and physics, and 1.8% biology and earth science. Pre-service teachers only studying earth science completed the survey due to their partial natural science studies. A total of 75.4% of the test participants studied to teach at grammar or comprehensive schools, while 23.7% prepared for other types of school contexts. Overall, 42.1% were undergraduate students, and 40.4% were Master of Education students (six test participants completed their degrees shortly before the study). The studies of 7% of the test participants ended with the first state examination, and 5.3% were trainee teachers. Undergraduate students completed 4.41 semesters (SD = 1.21) and taught 1.17 lessons (SD = 1.79) of 45-min in duration in the natural sciences (i.e., biology, chemistry, physics, science). Master of Education students completed 11.37 semesters (SD = 3.04) in total and taught 23.54 45-min lessons (SD = 55.89) in the natural science subjects (see above).

2.2. Measurement Instruments

2.2.1. The Self-Efficacy Beliefs of Interdisciplinary Science Teaching (SElf-ST) Instrument

We aimed to develop a draft for an instrument to measure self-efficacy beliefs regarding the demands of interdisciplinary science teaching for pre- and in-service teachers (according to Labudde [7], the definition of integrated science includes biology, chemistry, and physics as one subject). The construct of interest is self-efficacy beliefs relating to the beliefs to execute actions [9]. We base them on the PCK model, which contains knowledge [38]. Consequently, it was necessary to adapt the PCK to action to measure self-efficacy beliefs. We derived actions for self-efficacy beliefs from the subcategories of PCK [38]. Thus, we chose a PCK model [38] to operationalize the skills of all science subjects needed by pre- and in-service teachers for a theory-based instrument of self-efficacy beliefs. We used only interdisciplinary items to measure the real demand in interdisciplinary teaching. Disciplinary items would be too numerous and would not explicitly measure the actual challenge, i.e., interdisciplinary teaching of all three subjects. Thus, the PCK of all three subjects is integrated simultaneously, and interdisciplinary science teaching is used as the obstacle. Therefore, disciplinary issues are integrated nevertheless. Moreover, disciplinary items would complicate a consistent obstacle.

The researchers did not choose the consensus model [37] as it does not specify the components of PCK in detail [1,34] and focuses more on the impacts of PCK and its context factors (e.g., other parts of professional knowledge) [1]. We aimed to operationalize the facets of PCK for self-efficacy beliefs. Even the refined consensus model [1] does not specify the components of PCK, as did Park and Chen [38]. The refined consensus model is no replacement for previous models [1]. In the light of the refined consensus model, one could name our PCK the collective PCK (cPCK) [1] as it is in line with "what teachers need to know" [51] (p. 122). This cPCK has not been sufficiently defined to date [51]. We use the model from Park and Chen [38] to specify cPCK. The test participants rate their self-efficacy beliefs regarding the cPCK. Thereby, they consider their personal PCK (pPCK). Rating the use of this cPCK is the focus, as self-efficacy beliefs focus on action.

Consequently, we operationalized the in-depth described subcategories from four categories of the (c)PCK model of Park and Chen [38] (Section 1.2) for actions regarding teaching. Thus, the items measure self-efficacy beliefs. Only the subcategories of Misconceptions, Motivation, Interest, and Needs were summarized as the subcategory Needs. We did not operationalize the fifth category, Orientation to Teaching Science. It does not contain a learnable competence, and a challenging situation to overcome is, therefore, not reasonable to formulate. However, both issues are mandatory for the operationalization of self-efficacy beliefs, according to Schwarzer and Jerusalem [52] and Bandura [9].

With the PCK model [38] as the guiding model, current research findings (literature-based), as well as curricula and (teacher) educational standards (regarding biology, chemistry, and physics; curricular valid), were used to specify the adapted subcategories for self-efficacy beliefs. The researchers used the following sources for literature-based and curricular valid item development of the self-efficacy beliefs of interdisciplinary science teaching (SElf-ST) instrument: requirements for teacher education [53]; two books about science education [54,55]; the educational standards for learning outcomes in biology [56], chemistry [57], and physics [58]; the core curriculum for teaching science in grammar school [59], as it distinguishes between biology, chemistry, and physics; as well as an article about curricular materials [60].

The researchers also used established approaches to measure self-efficacy beliefs: (1) the wording "I can … " at the beginning of phrases for rating their skills [9]; (2) the use of present tense instead of future tense to measure present skills, not potential future skills [9]; and (3) the use of an obstacle or demanding situation to overcome, so that the task is not too easy [9,11].

To avoid the problem of unsystematic variation of obstacles, we standardized the obstacle on a content-related level. Interdisciplinary science teaching is the obstacle for all items of the SElf-ST instrument (in three slightly different wordings), which ensures a certain level of difficulty for all items. Based on this standardization, the different tasks of the items have varying levels of difficulty. The content-related standardized obstacle allows interpretations of the items only relating to the

task in the context of interdisciplinary science teaching. Unsystematic varying obstacles additional to interdisciplinary science teaching would further complicate this interpretation. Overall, this standardization is a parsimonious solution for integrating obstacles. To ensure the interpretation of the obstacle, it is always labeled with a footnote about what interdisciplinary science teaching means: "*Imparting biological, chemical, **and** physical as well as natural scientific concepts, and ways of thinking and working.*" To avoid measuring knowledge (of technical terms), we explained the terms in footnotes, when deemed necessary. If reasonable and possible, the researchers provided illustrated items, by means of examples from the three natural scientific subjects:

"**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)." (Table A3, item b_4)

. . . use models as research tools (for example, hypothesizing and hypothesis testing with the atomic model, the model of a nerve cell, the model of a wind tunnel)." (Table A3, item h_7)

Based on the subcategories of the PCK model ([38], Section 1.2), we developed six self-efficacy beliefs items for the subcategory Curriculum Materials, five for the subcategory Learning Difficulties, five for the summarized subcategory Needs, five for the subcategory Vertical, five for the subcategory Horizontal Curriculum, seven for the subcategory Dimensions of Science Learning/Scientific Literacy to Assess, five for the subcategory Methods of Assessing Science Learning, ten for the subcategory Subject/Science-specific Strategies, ten for the subcategory Activities, and five for the subcategory Representations. In sum, the researchers developed 63 items for the first version of the SElf-ST instrument based on and summed under the subcategories of the PCK model [38]. Some items were more science-specific than others. This development results in mostly science-specific items, as well as a few that were more generic, as they apply to teaching all subjects (e.g., methods of evaluation or some instructional strategies).

In accordance with two studies about general teaching self-efficacy [61,62], the researchers applied a four-point response scale, which included "Is not right" (1), "Is a little right" (2), "Is rather right" (3), and "Is exactly right" (4).

Initially, a pilot study of the SElf-ST instrument tested $N = 10$ pre-service teachers to identify potential improvements and to solicit remarks. A total of 80% were women, 80% were Master of Education students, 80% studied biology, and 20% studied biology and chemistry. A total of 62 items showed no problems in understanding. Only one item (Table A3, item f_3) required adjustment based on evidence. Thus, the researchers deemed the items to be comprehensible and appropriate. The resulting second version of the SElf-ST instrument contained 63 items [63].

2.2.2. Validation Measurement Instruments

First, researchers conducted a survey of the number of lessons in biology, chemistry, physics, and science that pre-service teachers taught as a part of practical training in a school. The survey included the number of lessons taught by trainee teachers. The researchers, then, aggregated the disciplinary values for each person.

Second, we applied the factor-analytic tested, curricular-valid measurement instrument of Handtke et al. [64] for srCK in the natural sciences, which has foundations in the curricular requirements for grammar schools [59] and contains 20 items. They load on three factors: Biology (eight items, Cronbach's $\alpha = 0.92$, $\lambda = 0.640$ to 0.881), Chemistry (five items, $\alpha = 0.94$, $\lambda = -0.727$ to -0.902), and Physics (seven items, $\alpha = 0.95$, $\lambda = 0.728$ to 0.903). The four-point response scale includes, "Do not agree at all" (1), "Do rather not agree" (2), "Do rather agree" (3), and "Fully agree" (4). This measurement instrument includes the core ideas of the core curriculum and three general assessments on whether the test participants believe that they have the CK to teach the biological, chemical, and physical parts involved in interdisciplinary science teaching. One example regarding the core ideas is:

"I know very much about the core idea . . .

. . . history and relationship (e.g., groups of vertebrates, pedigree analysis as well as homology, and analogy)".

The questionnaire always began by requesting personal data (e.g., subjects, finished practical training, course of study), followed by the measurement instruments for srCK in science [64] and for self-efficacy beliefs of interdisciplinary science teaching [63] (the second version of the SElf-ST instrument) and the self-efficacy beliefs of teaching education for sustainable development instrument. Some test participants completed a survey about their knowledge concerning biodiversity and climate change following the Self-ST instrument. The test participants completed surveys within courses, as well as beyond courses. They received a financial reward for completing beyond courses. The reward was €15 for completing everything and €10 for those not completing the survey on knowledge concerning biodiversity and climate change. The average time required for completing the questionnaire about srCK in science and the SElf-ST instrument (second version) was approximately 20 min.

2.3. Analysis

The researchers analyzed the data with the statistics program SPSS Statistics 24 (IBM, Armonk, NY, USA). The measurement instrument basis is from a PCK model [38], which has a design to measure for knowledge and not self-efficacy beliefs. The appropriateness of the model for self-efficacy beliefs required testing by an exploratory factor analysis (EFA). In addition, the exploratory approach directly reveals concrete discrepancies between the theoretical and empirical (adapted) model. The PCK model development has a basis in qualitative data [e.g., 38]. A previous version of the PCK model contains self-efficacy beliefs as an independent category [41]. Likely due to insufficient evidence, removal of the category was, again, necessary [42]. In general, the EFA can be appropriate for measurement instrument development despite an existing theory [65]. The strength of the theory is decisive [65], and as argued before, the PCK model has some weaknesses. For our measurement instrument of self-efficacy beliefs, the PCK model needs to be adapted to self-efficacy beliefs first. In sum, as a first draft for a new measurement instrument for self-efficacy beliefs of science teaching, the EFA seems to be appropriate for generating hypotheses about the factor structure regarding self-efficacy beliefs. After adapting the PCK model to self-efficacy beliefs, this draft can be checked with a confirmatory factor analysis in a subsequent study for further examination [66].

According to Bühner [67], the sample size is sufficient for an EFA. After checking the prerequisites (e.g., Kaiser–Meyer–Olkin (KMO) value, Bartlett's Test [68]), the researchers chose the principal factor analysis (PFA) as the extraction method because the PFA aims at determining latent variables [69]. Due to a lack of a normal distribution, the maximum-likelihood method could not be applied [70]. The eigenvalues (Kaiser–Guttman criterion [71,72]) and the scree plot [73] results provided identification of the number of factors, although researchers considered the scree plot as less important due to its subjectivity [67]. The usage of oblique rotation (cf. [66]) direct oblimin allowed for correlations between the factors, as they seem possible from a theoretical point of view (cf. [38]). After checking the parameter, the pattern matrix received consideration by the researchers (pairwise deletion, display of loadings ≥0.3). Besides the loading (≥0.3) [68], the content-related fit of the items to the respective factors also received consideration [67]. With each item removal from the analysis, the researchers reran the EFA with the remaining items [68]. For computing the factor scores, researchers used the weighted mean [74]. As an exploratory approach, we completed the data computation in metric values. Due to the lack of normal distribution and outliers in teaching experience (Section 2.1), Spearman's correlation was applied [68]. Then, the researchers interpreted the correlations according to Cohen [75] ($r \geq 0.1$: small, $r \geq 0.3$: medium, $r \geq 0.5$: large). Due to multiple significance testing, we used the Benjamini–Hochberg method in RStudio (version 1.2.1335) to adjust the p-values [76]. The approach has greater power than the Bonferroni method, for example, [76] and thus is appropriate for an exploratory approach to identify possible relations and reduce the beta error.

3. Results

3.1. The Self-Efficacy Beliefs of Interdisciplinary Science Teaching (SElf-ST) Instrument

The KMO value confirms the adequacy of the sample for the PFA; KMO = 0.858 ("meritorious" according to Kaiser and Rice [77] (p. 112)). Bartlett's test of sphericity is significant, χ^2 (820) = 2496.87, $p < 0.001$. After six reruns due to the removal of items, 41 of the 63 constructed items remained, fitting the demands and being content-related interpretable. The scree plot was ambiguous, as the only clear "elbow" was between the first and second factor. However, with the aim of measuring self-efficacy beliefs as multidimensional, as recommended by Bandura [9], a one-factor solution is not appropriate. Thus, and because of the subjectivity of the scree plot [67], we applied only the eigenvalues. In the final sixth PFA, ten factors had an eigenvalue above Kaiser's criterion of one. After extraction, and through these ten factors, we explained 58.79% of the variance in combination. Lists of the ten factors of the SElf-ST instrument (final version) after rotation, their basic indices, and the accordance with the underlying knowledge subcategories of the PCK model are in Table 1. The pattern matrix, after six reruns of the PFA, and the item loadings (≥ 0.3) are in Table A3.

Table 1. Ten factors of the self-efficacy beliefs of interdisciplinary science teaching (SElf-ST) instrument (final version, sorted by eigenvalues). Pedagogical content knowledge (PCK) model according to Park and Chen [38]. α = Cronbach's alpha, n = number of items, M = observed mean, SD = standard deviation and λ = factor loadings. Accordance of self-efficacy belief factors with the (pedagogical content) knowledge model: ● = coincident ◖= partially coincident. S = science-specific factor, G = generic factor.

Subcategory(-ies) of PCK Model (Knowledge Concerning …)	Factor (Self-Efficacy Beliefs of …)	Accordance: ●/◖	α (n)	M (SD)	λ
Dimensions of Science Learning to Assess	1. Surveying Dimensions of Scientific Literacy (S)	●	0.84 (5)	3.00 (0.62)	0.588 to 0.697
Representations	2. Applying Media (G)	●	0.81 (4)	3.42 (0.59)	0.614 to 0.804
(not considered)	3. Teaching Ethically Relevant Issues of Applied Science (S)	new	0.86 (4)	2.97 (0.67)	0.336 to 0.672
Vertical and Horizontal Curriculum	4. Differentiated Fostering of Scientific Inquiry and Communication in Science (S)	◖ (see factor 10)	0.80 (5)	2.88 (0.51)	0.385 to 0.722
Curriculum Materials	5. Using Subject-specific Materials in Science (S)	●	0.76 (4)	3.02 (0.56)	0.512 to 0.701
Activities	6. Applying Natural Scientific Working Methods (S)	●	0.83 (5)	3.16 (0.59)	0.444 to 0.726
Methods of Assessing Science Learning	7. Applying Methods of Evaluation (G)	●	0.70 (3)	2.93 (0.61)	−0.536 to −0.617
Learning Difficulties and Needs (Misconceptions, Motivation, Interest, Need)	8. Considering Learning Difficulties and Needs of Students in Science (S)	● (summed)	0.79 (5)	3.24 (0.50)	−0.459 to −0.694
Subject/Science-specific Strategies	9. Including Science-specific and General Instructional Strategies (S/G)	●	0.78 (3)	3.04 (0.62)	0.480 to 0.608
Vertical and Horizontal Curriculum	10. Surveying and Fostering Natural Scientific Content Knowledge (S)	◖ (see factor 4)	0.80 (3)	3.00 (0.65)	−0.447 to −0.600

Cronbach's α of the factors are at least acceptable [68] and range from 0.70 to 0.86. The means of the factors of the SElf-ST instrument (final version) range from 2.88 (SD = 0.51) to 3.42 (SD = 0.59), overall, with an observed mean of 3.07 (SD = 0.42). Each subcategory of the four PCK model categories [38] is represented by at least one empirical factor that has three or more items (Table 1). The names of the fitting knowledge subcategories and self-efficacy belief factors differ, as the PCK model contains subcategories of knowledge, and the empirical factors describe action regarding the area of

knowledge. Thus, they can vary in their name but, nevertheless, go together. As previously stated (Section 2.2.1), some factors are (partly) generic due to their meaning for various subjects. Media and evaluation methods, as well as some general instructional strategies, are generic, which leads to two generic (Table 1, factors 2 and 7) and one partly generic (Table 1, factor 9) factor.

One part of the curriculum items (d in Table A3) now loads on the two factors of content- and process-related competencies—content knowledge (Table 1, factor 10; Table A3, items d_4, d_9) and scientific inquiry and communication (Table 1, factor 4; Table A3, items d_1, d_2, d_6, d_7). Both factors are completed with one item each, one concerning the consideration of different learning conditions by differentiation (Table A3, item c_4), while the other concerns the survey of content knowledge (Table A3, item e_1). The remaining curriculum items relate to socio-scientific decision making (Table A3, items d_3, d_8) and load with two further items, which concern consideration of difficulties with ethically complex issues (Table A3, item b_4) and surveying socio-scientific decision making (Table A3, item e_7), on the new factor Teaching Ethically Relevant Issues of Applied Science. The items e_1 and e_7 (Table A3) were the only ones with a substantial second-factor loading (≥0.3). We decided to assign them to their topic-specific factor, as the loadings were similar (Table A3); the first factor was represented well, even without them. From a psychometric view, this ensured that all factors contained at least three items. The items of the separated subcategories Learning Difficulties and Needs of the same category (Section 1.2) now load together on Considering Learning Difficulties and Needs of Students in Science (Tables 1 and A3, factor 8). Another item, regarding the application of models (Table A3, item i_1) of the subcategory Representations, now loads on the factor Applying Natural Scientific Working Methods.

3.2. First Indicators of the Validity of the SElf-ST Instrument

Table 2 shows Spearman's correlations between the factors of self-efficacy beliefs of science teaching (factors 1–10), the lessons taught (in practical training) in school(s) (factor 11), and the factors of the srCK in science (factors 12–14).

The correlations between the factors of self-efficacy beliefs of science teaching range from 0.20 to 0.56 ($p < 0.05$). Regarding the first hypothesis (H1), nine of the ten factors of self-efficacy beliefs showed a low to medium ($r = 0.20$–0.37, $p < 0.05$) correlation with the number of lessons taught. Four central science-specific factors—Differentiated Fostering of Scientific Inquiry and Communication in Science, Using Subject-specific Materials in Science, Applying Natural Scientific Working Methods, and Surveying and Fostering Natural Scientific Content Knowledge—showed a medium correlation ($r = 0.33$–0.37, $p < 0.01$). Three science-specific factors—Surveying Dimensions of Scientific Literacy, Teaching Ethically Relevant Issues of Applied Science, and Considering Learning Difficulties and Needs of Students in Science—showed a low ($r = 0.20$–0.24, $p < 0.05$) correlation with the lessons taught. One generic factor—Applying Media—and one partly generic factor—Including Science-specific and General Instructional Strategies—showed low correlations ($r = 0.25$–0.26, $p < 0.01$). The generic factor Applying Methods of Evaluation had no significant correlation.

Regarding the second hypothesis (H2), excluding one correlation—Using Subject-specific Materials in Science with Self-rated Content Knowledge in Physics ($r = 0.22$, $p < 0.05$)—the analyses showed medium correlations ($r = 0.31$–0.45, $p < 0.01$) of three central science-specific factors—Using Subject-specific Materials in Science, Applying Natural Scientific Working Methods, and Surveying and Fostering Natural Scientific Content Knowledge—with the three factors of srCK. Two generic factors of the self-efficacy beliefs of science teaching—Applying Media and Applying Methods of Evaluation—had no significant correlation with the factors of the srCK in science. However, every (even only partial) science-specific factor of the self-efficacy beliefs of science teaching showed a low (or medium) positive correlation with the srCK in one or more of the subjects of biology, chemistry, or physics (H2; $r = 0.19$–0.49, $p < 0.05$).

Table 2. Spearman's correlations between the (weighted) factors of the self-efficacy beliefs of interdisciplinary science teaching (SElf-ST) instrument (final version; factors 1–10), the lessons taught (in practical training) in schools (factor 11) and the (weighted) factors of the self-rated content knowledge in science (factors 12–14). Adjusted p-values with Benjamini–Hochberg method due to multiple significance testing.

Factor	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. Surveying Dimensions of Scientific Literacy	1													
2. Applying Media	0.26 **	1												
3. Teaching Ethically Relevant Issues of Applied Science	0.48 **	0.31 **	1											
4. Differentiated Fostering of Scientific Inquiry and Communication in Science	0.37 **	0.30 **	0.52 **	1										
5. Using Subject-specific Materials in Science	0.37 **	0.25 **	0.38 **	0.48 **	1									
6. Applying Natural Scientific Working Methods	0.51 **	0.46 **	0.41 **	0.37 **	0.30 **	1								
7. Applying Methods of Evaluation	0.36 **	0.28 **	0.46 **	0.36 **	0.33 **	0.33 **	1							
8. Considering Learning Difficulties and Needs of Students in Science	0.45 **	0.20 *	0.51 **	0.39 **	0.38 **	0.38 **	0.37 **	1						
9. Including Science-specific and General Instructional Strategies	0.48 **	0.38 **	0.46 **	0.43 **	0.25 **	0.50 **	0.42 **	0.45 **	1					
10. Surveying and Fostering Natural Scientific Content Knowledge	0.56 **	0.30 **	0.42 **	0.48 **	0.50 **	0.52 **	0.41 **	0.45 **	0.49 **	1				
11. Taught lessons in school(s)	0.20 *	0.26 **	0.22 *	0.36 **	0.37 **	0.33 **	0.10	0.24 **	0.25 **	0.36 **	1			
12. Self-rated Content Knowledge in Biology	0.07	0.14	0.14	0.21 *	0.40 **	0.31 **	0.05	0.04	0.14	0.39 **	0.29 **	1		
13. Self-rated Content Knowledge in Chemistry	0.34 **	0.03	0.10	0.14	0.31 **	0.33 **	0.04	0.21 *	0.19 *	0.45 **	0.38 **	0.36 **	1	
14. Self-rated Content Knowledge in Physics	0.49 **	0.05	0.19 *	0.11	0.22 *	0.33 **	0.13	0.38 **	0.33 **	0.36 **	0.13	−0.18 *	0.44 **	1

$* p < 0.05$; $** p < 0.01$.

4. Discussion

4.1. The Self-Efficacy Beliefs of Interdisciplinary Science Teaching (SElf-ST) Instrument

This draft for a new measurement instrument of self-efficacy beliefs of science teaching is aimed at teaching interdisciplinary science at grammar or comprehensive schools. Many existing measurement instruments for science subjects refer to the primary school (Table A1). Overall, the measurement instruments reflect different focuses and aims (Section 1.1). Thus, they are often limited to one empirical factor [8,29] or maximally contain four content-related empirical factors, differentiated by planning and performance [30]. The presented draft for a measurement instrument with ten theory-based (PCK) factors could enable differentiated analyses of learning capabilities and education successes, thus, providing starting points for an evidence-based advancement of teacher education for science teaching. The EFA provides initial indicators for the differentiation of these ten different factors for science teaching (more evidence, see: Section 4.2).

To date, a theory-based operationalization of self-efficacy beliefs of science teaching has been rather neglected in research (see Section 1.1). For example, the measurement instrument of Gibson and Dembo [78] was developed to STEBI-A and -B by modifying and adding further items, as well as by a subsequent expert survey [8,79]. The STEBI was a milestone in measuring self-efficacy beliefs of science teaching. To date, this measurement instrument has deficits in regard to current teachers' professional competence, standards, and core curricula. In addition, measurement instruments like the STEBI have been modified based on a literature review or further developed by new items, which introduces new focuses [23,24]. However, they remained rather global. Regardless of the STEBI, for example, there is one measurement instrument that is simultaneously based on a literature review and the (Swedish) curricula for pre-school and compulsory school [25]. It focuses on being time-saving [25]. Rarely have self-efficacy instruments in this area been developed based on a model as a guiding principle, as attempted by Vidwans [28], with culturally responsive pedagogy (for criticism, see Table A2 and Section 1.1). Our draft for a new measurement instrument could be the first to meet these three requirements—literature-based, theory-based, and curricular-valid. The integration of current research findings, the integration of the curricula, as well as indicators for curricular validity (see below), and the result of the EFA, according to the model, support this claim. In addition, our draft could be internationally compatible due to the international relevant PCK model [38]. This underlying model could enable specific conclusions about the learning outcomes of teacher education as well.

Apart from lacking a theory-based or a specific multi-faceted measurement, another reason for developing a new measurement instrument is the limited integration of obstacles thus far. Our new draft for a measurement instrument, in contrast to many previous measurement instruments (Section 1.1), would meet the construction rule of an obstacle [9,11]. Items not integrating an obstacle could violate the construction rules and be too easy [9]. To avoid the issue of systematic variation, we used one content-related standardized obstacle (three slightly different wordings) for all items—interdisciplinary science teaching. In contrast to unsystematically varying obstacles, this standardization could be an innovative and parsimonious approach for an unambiguous interpretation of the content-related challenges of each item in teaching the natural sciences.

Besides the subcategories of the PCK model from Park and Chen [38], Teaching Ethically Relevant Issues of Applied Science emerged as an additional factor. It contains, among other things, horizontal and vertical fostering of socio-scientific decision making, which is curricular-demanded for science [56–58]. The consideration of ethically relevant issues is an evident gap in Park and Chen's [38] model. The new factor could meet this requirement.

The items regarding the curriculum are aggregated as one factor for content knowledge, one for scientific inquiry and communication, and one for ethically relevant issues, including socio-scientific decision making. Against the background of content- and process-related competencies, and trying to develop a curricular-valid measurement instrument (cf. [56–59]), this arrangement seems to be plausible and provides an indicator for curricular validity (i.e., a type of content validity [20]). These

three factors are complemented with varying numbers of items concerning the consideration of different learning conditions by differentiation (Table A3, item c_4), surveying scientific literacy (Table A3, items e_1 and e_7), and considering difficulties with ethically complex issues (Table A3, item b_4).

In accordance with the theory, the items of Learning Difficulties and Needs (Table A3, items b_1–3 and c_1–2) load on the joint empirical factor Considering Learning Difficulties and Needs of Students in Science, as they belong to the same category of the PCK model [38]. The applying models item (Table A3, item i_1), of the subcategory Representations, is now part of Applying Natural Scientific Working Methods. The underlying subcategory is Activities. Both subcategories are part of the same category—Topic-specific Strategies (Section 1.2). Therefore, the arrangement is comprehensible. In sum, despite adapting to current challenges, like curricula and the state of research, the operationalized theoretical (sub)categories could be empirically found to a great extent as first hypotheses about the factor structure.

The test participants showed observed means on the factors from 2.88 to 3.42 (theoretical mean = 2.5) and, an overall observed mean of 3.07 (SD = 0.42). Exceeding the theoretical mean rather indicates positive self-efficacy beliefs of science teaching. In comparison to other studies about self-efficacy beliefs, this finding proves to be according to expectations (Table 3). Studies with pre-service teachers [8,33], as well as studies with in-service teachers [28,61], show similarly high values, which are comparable to the sample of predominantly pre-service teachers in this study.

Table 3. Summary of diverse studies about self-efficacy beliefs of teaching. *M* (Factors) = published means of factors, *M* (Scale) = published mean overall, and Theoretical *M* = theoretical mean with the scale.

Authors	Scope	*M* (Factors)	*M* (Scale)	Theoretical *M* (Scale)
Enochs and Riggs [8]	Science	-	3.62	3 (1 to 5)
Schwarzer and Schmitz [61]	General teaching	-	2.87	2.5 (1 to 4)
Rabe et al. [33]	Physics	1.50 to 1.92	1.70	1.5 (0 to 3)
Vidwans [28]	Science/General	-	7.20	5 (0 to 10)

We want to emphasize two content-related reasons that potentially elicit the rather positive self-efficacy beliefs—the misjudgments of the pre-service teachers and the similarity of the didactics. One reason could be that (prospective) teachers overestimate their abilities [80], underestimate the challenges [81], or do not know which skills are required [22]. These possible misjudgments could result in (too) high self-efficacy beliefs. Another potential reason is that the measured PCK in science consists of the PCK in biology, chemistry, and physics. The curricula of these three subjects show various similarities and overlaps [7]. All the subjects have the same three spheres of procedural competence and, e.g., share the common topics experiments, using models, and appropriate argumentation (cf. [59]). Thus, someone studying biology, chemistry, or physics seems to be unlikely to state, "Is not right"(1) and rather seems to state "Is a little right"(2), or "Is rather right"(3) as the PCK of at least one subject is well-known. Having predominantly biology, chemistry, and physics pre-service teachers in the sample, could cause this rather positive self-efficacy beliefs. This conclusion is supported by 34 of the 41 items most frequently answered with "Is rather right" (3). This finding rather contradicts questions regarding extreme ceiling effects, as "Is exactly right" (4) was most frequently chosen only six times ((3) and (4) were equivalently chosen one time).

In the present study, the factors Applying Media (*M* = 3.42, SD = 0.59) and Considering Learning Difficulties and Needs of Students in Science (*M* = 3.24, SD = 0.50) showed very positive values. Both contain basic requirements for every teacher, which are addressed early in teacher education (e.g., students' conceptions, typical mistakes during experimenting). Lecturers of biology education support this claim, as they highlight the importance of fostering competence in regard to methods

and media [82]. In addition, well-founded knowledge of students' attributes is one aim of teacher education at the university level [83].

Less positive, in comparison, are the means of the factors Teaching Ethically Relevant Issues of Applied Science (*M* = 2.97, SD = 0.67), Applying Methods of Evaluation (*M* = 2.93, SD = 0.61), and Differentiated Fostering of Scientific Inquiry and Communication in Science (*M* = 2.88, SD = 0.51). The slightly lower means in the area of ethics and socio-scientific decision making can be explained by the means of the interview studies, wherein teachers expressed difficulties with diagnosing [84] and fostering [85] socio-scientific decision making. Whereas the internship of trainee teachers should lead to mastering the evaluation of performance, the university should teach the basics of performance evaluation [83]. Thus, mastering methods of evaluation is addressed at a later stage after university. In this study, we surveyed the subjective mastering of these methods of students, especially. The observed mean is in accordance with the requirements for teacher education [83] and could suggest curricular validity. Regarding Differentiated Fostering of Scientific Inquiry and Communication in Science, surveying biology pre-service teachers showed that they believe that, among other things, communication competence is least fostered in teacher education [82]. According to lecturers in biology education, differentiation is rather less fostered as well [82]. These findings could be starting points for explaining the lower means of the three factors mentioned above.

4.2. First Indicators of the Validity of the SElf-ST Instrument

Besides the indicators on curricular validity stated before, this study could produce the first indicators of reliability and different types of validity for the SElf-ST instrument. The factors of self-efficacy beliefs of science teaching show at least acceptable values for Cronbach's α and good values for the PFA parameters (e.g., KMO value = 0.858). The relatively low intercorrelations of the factors for the self-efficacy beliefs of science teaching rather support the divergent validity of ten separated factors as they are correlated, but low, as they are distinct constructs. These findings address the research question in regard to identifying empirical evidence for proving the subcategories based on the PCK model for self-efficacy beliefs. The low and medium correlations of teaching experience with the factors of self-efficacy beliefs of science teaching (H1; except generic factor Applying Methods of Evaluation) support, according to concurrent validity, the theoretical assumption that direct teaching experience positively correlates with self-efficacy beliefs [9].

Against the background of Bandura claiming direct experience to be the most important source [9], higher correlations could have been expected. Due to the different contexts between the most surveyed teaching experiences—biology, chemistry, and physics—and the interdisciplinary self-efficacy beliefs of science teaching, it seems comprehensible that the correlations with teaching experience are not strong, according to context specificity [10]. Examining the other sources—vicarious experience, verbal persuasion, and physiological and affective states [9]—could reveal the most effective source of self-efficacy beliefs in our context. Science methods courses (e.g., [13]) could be aligned with the most effective source. As a first finding, this paper reveals direct experience in teaching biology, chemistry, physics, or science to be correlated with self-efficacy beliefs as an indicator for concurrent validity and as one possible starting-point for improving professional development in teacher education, including self-efficacy beliefs.

The low (or medium) correlations of every (even only partial) science-specific factor of the self-efficacy beliefs of science teaching with the srCK in no less than one of the subjects (H2) support the theoretical assumption that srCK positively relates to self-efficacy beliefs [45,50]. At the same time, the low correlations prove them as separated constructs, which argues for divergent validity.

Generally speaking, central science-specific factors of the self-efficacy beliefs of science teaching indicate primarily medium correlations with practical teaching experience (H1) and low or medium correlations with the srCK in science (H2). In contrast, (rather) generic factors show low(er) or no correlations. According to the context specificity of self-efficacy beliefs [10], for both hypotheses, (rather) generic factors show lesser correlations with natural sciences teaching experience, or srCK in

science, than science-specific factors, due to the different context. This finding is a further indicator of divergent validity, as generic factors differ from the science-specific context.

Against the background of the context specificity, dividing srCK into biology, chemistry, and physics could result in some science-specific factors of the self-efficacy beliefs of science teaching being correlated with only one or two factors of srCK in science. This assumption is supported by the contexts, or subjects, not being completely identical. The finding that every (even only partial) science-specific factor of the self-efficacy beliefs of science teaching is still correlated with at least one factor of the srCK in science (H2) [45,50] can be interpreted as an indicator of divergent validity.

4.3. Limitations

Due to the "Qualitätsoffensive Lehrerbildung" having numerous simultaneous surveys, test participants from different universities could not be included in the sample as planned. This limitation should be addressed in a subsequent study. In addition, the sample only contains a small number of trainee teachers.

Dividing srCK among the three natural science subjects displays the reality of teacher education. However, an interdisciplinary construct was compared with disciplinary constructs. This information should be considered by interpreting the correlations. Moreover, the composition and size of the sample possibly resulted in not all of the science-specific factors correlating significantly with all the srCK factors. In this study, srCK, not CK was measured. This limitation was unavoidable due to time restrictions, as three subject area surveys (biology, chemistry, and physics) would need to be conducted.

The financial reward potentially impacted the sample by motivating participants with low self-efficacy beliefs as well. The incentives could be a reason why there are few missing responses on the item level. The incentives could influence the mood of the test participants. This influence could have an impact on self-efficacy belief results in comparison with unrewarded test participants.

4.4. Future Research

This paper presented a draft for a new measurement instrument of self-efficacy beliefs of interdisciplinary science teaching (SElf-ST) to meet current demands, especially in German teacher education and schools. The aim was to develop a new measurement instrument, obtain first hypotheses about the factor structure (subsequently needing to be proven with an independent sample), and provide initial indicators on its reliability and (concurrent, divergent, and curricular) validity. In the first step, the PCK model [38] required adjustment towards our aim of surveying self-efficacy beliefs of science teaching. The latter was empirically performed with an EFA. Regarding construct validity, the first step was completed by identifying hypotheses about the factor structure with the EFA. In a further study, this hypothetical factor structure needs verification by means of a confirmatory factor analysis that includes a larger and independent sample [66,69]. In addition, researchers will need to perform further validation, for example, with correlations with the STEBI [8] or the multidimensional scale of teacher self-efficacy beliefs [62]. It could be further desirable to investigate the other sources of self-efficacy beliefs of interdisciplinary science teaching as starting-points for improving seminars and practical training.

Following the main study, including confirming the factors with a CFA and further validation, the SElf-ST instrument can be used for different purposes, e.g., evidence-based teacher education for interdisciplinary science teaching in Germany—and beyond (see below). It could identify the effects of interventions, as science methods, or content courses, as a pre-post-measurement (cf. [13,14]). Interventions with different sources of self-efficacy beliefs could also be tested. The SElf-ST instrument could help to reflect one's own abilities in science teaching. After confirming the model with a CFA, the SElf-ST instrument could be used to examine the development of self-efficacy beliefs in longitudinal studies.

The theory-based approach could be adapted to other interdisciplinary subjects, such as education for sustainable development. The SElf-ST instrument could be used in other countries than Germany to

examine teacher education as well. Perhaps only a slight adaptation for the country-specific curricula would be necessary for nations without interdisciplinary science teacher education like Germany. The SElf-ST instrument can be applied to inspect learning outcomes of teacher education. In countries with interdisciplinary teacher education in science, an adaption to the country-specific curricula, as well as an adjustment of the obstacle, could be necessary. The obstacle (interdisciplinary science teaching) may be less effective in countries with interdisciplinary science teacher education.

This paper presented a new theory-based approach, developed according to Bandura's guidelines [9], for the SElf-ST instrument, which can serve as a starting point for various adaptions to promote (research regarding) self-efficacy beliefs in science education and beyond.

SPONSORED BY THE

Federal Ministry
of Education
and Research

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Appendix A

Table A1. Chronological selection of published measurement instruments regarding self-efficacy beliefs for teaching natural science subjects in primary education. To manage the amount of literature and as we investigated the education of teachers, we focused on the Science Teaching Efficacy Belief Instrument-B (STEBI-B) for prospective teachers and made no clear distinction between versions A and B, as they are relatively similar. In addition, outcome expectation is not considered at any point. We considered only published measurement instruments. Measurement instruments that were used for secondary education but did not show an obvious adaptation at that stage were counted as primary education (e.g., [26]). EFA = exploratory factor analysis (this may include principal component analysis), CFA = confirmatory factor analysis. All measurement instruments include negative and positive five-point items, unless a different measurement is explicitly noted under "Characteristics". All measurement instruments do not include obstacles.

Authors/Measurement Instrument	Scope	Characteristics	Assessment from a Contemporary Point of View
[Riggs and Enochs [79]/STEBI-A] Enochs and Riggs [8]/STEBI-B	Science	• $N = 212$ • Analysis: CFA • $\alpha = 0.90$ for 13 items • One factor: • Personal Science Teaching Efficacy Belief	• Current demands (standards, professional competence of teachers, core curricula, etc.) did not exist; thus, not considered • Groundwork is a measurement instrument of Gibson and Dembo [78], which is based on interviews with teachers and literature • Future tense does not fit the guidelines of Bandura ("capabilities as of now, not their potential capabilities or their expected future capabilities", [9] (p. 44))
Ritter [23]/SEBEST	Derived from STEBI; special focus on equality	• $N = 226/102 + 23$ • Analysis: EFA • $\alpha = 0.82/0.83 + 0.81$ for 17 items • Four factors: (a) Ethnicity (b) Language Minority (c) Gender (d) Socioeconomic Status	• STEBI as the base; thus, with similar deficits in regard to current teachers' professional competence, standards, and core curricula • Categories based on a literature review • Strength is the focus on equality
Roberts and Henson [24]/SETAKIST	Modified STEBI including natural scientific content knowledge	• $N = 247$ • Analysis: CFA • CFI = 0.937, RMSEA = 0.057 for 16 items • Two factors: (a) Teaching Efficacy (b) Knowledge Efficacy	• STEBI as the base • Literature-based adding of items/one factor (natural scientific content knowledge) • Reasonable addition of content knowledge as a part of professional competence, but no concrete topics of content knowledge are listed

Table A1. *Cont.*

Authors/Measurement Instrument	Scope	Characteristics	Assessment from a Contemporary Point of View
Savran and Çakiroğlu [27]/BTEBI	STEBI referred to biology	• $N = 29$ • Analysis: *t*-Test • $\alpha = 0.83$ for 12 items • One subscale: Personal Biology Teaching Efficacy (Beliefs)	• For 12 of 13 STEBI-items, only "science" was replaced by "biology" (not considered: "Even if I try very hard, I will not teach science as well as I will most subjects.", [8] (p. 28)) • Very small sample for quantitative analysis, response categories collapsed • It remains vague, to what extent the measurement instrument for primary education (STEBI) is appropriate for surveying regarding secondary schools
Riese [26]	STEBI/Bleicher [86] referred to physics and reduced	• $N = 271$ • Analysis: CFA • $\alpha = 0.74$ for 6 items • CFA: RMSEA = 0.034, $p < 0.05$ • One factor: Self-Efficacy Particular Referred to Learning Physics	• Nearly identical on the content level with the STEBI and therefore with deficits in regard to current teachers' professional competence, standards, and core curricula as well • Reduced version of STEBI referred to physics provides fewer facets • Tested with teachers of secondary education, but no adjustment of the measurement instrument on secondary education is apparent
Mavrikaki and Athanasiou [22]/BioTSEB	STEBI referred to biology and complemented	• $N = 202$ • Analysis: EFA • $\alpha = 0.60–0.96$ for 44 items • Four factors: (a) Self-Efficacy in Plants, Ecology, and Human Biology Concepts (b) Pedagogical Content Knowledge, In-depth Understanding and Willingness to Teach Biology (c) Self-Efficacy in Evolution, Molecular Biology, and Microbiology (d) Motivation and Engagement	• Adding newer findings of research regarding teacher education • Partly literature-based (relating to motivation, pedagogical content knowledge, innovative methods), subject concepts based on curriculum (Greece) • For example, on the level of instructional and evaluation strategies, the survey is almost only general and little concretized ("I am continually finding better ways to teach the biology concepts that are included in the curriculum.", [22] (p. 213)) • Exemplary addressing of pedagogical content knowledge; limited analysis of differentiated learning capabilities possible • Factors (a) and (c) can be rather considered as self-rated content knowledge than self-efficacy beliefs as they ask for the familiarity with biology concepts (to feel confident to teach them)
Walan and Chang Rundgren [25]	Independent of STEBI and for science	• $N = 71$ • Analysis: Analysis of Variance • $\alpha = 0.75$ for 13 items • Three subscales: (a) Scientific Literacy (b) Curriculum (c) Learning Environment • Measurement: three-point response scale, positive items	• Literature-based and curricular valid for Sweden • Applied only three of the nine theoretical identified subscales (basis: Swedish curricula) due to time restrictions; thus, the breadth is limited • Relatively little differentiation of the measurement instrument with three response categories • Insufficient empirical factor-analytic testing

Table A2. Chronological selection of published measurement instruments regarding self-efficacy beliefs for teaching natural science subjects in secondary education. Primary education is partly included, for example, if in the United States of America, it is referred to as K-12. Outcome expectation is not considered at any point. EFA = exploratory factor analysis (this may include principal component analysis), CFA = confirmatory factor analysis. All measurement instruments do not include obstacles, unless it is explicitly noted under "Assessment from a Contemporary Point of View".

Authors/Measurement Instrument	Scope	Characteristics	Assessment from a Contemporary Point of View
Barros et al. [29]	(Among other things) modified Science Teaching Efficacy Belief Instrument (STEBI) for physics	• $N = 136$ • Analysis: EFA • $\alpha = 0.79$ for 9 items • One factor: Personal Efficacy Belief of Physics Teachers • Measurement: five-point response scale, positive and negative items	• Deficits in regard to current teachers' professional competence, standards, and core curricula like the STEBI: current demands of practice and research findings (professional competence, core curricula, …) are little integrated • Relatively little differentiated measurement instrument (9 items)
Pruski et al. [31]/SETAKIST-R	Modified SETAKIST [24] (changing the wording of items 6, 9, and 12) and adjusted for primary and secondary education	• $N = 334$ • Analysis: CFA, IRT (Partial Credit) • CFI = 0.915, RMSEA = 0.063, $p < 0.05$ for 16 Items • Two factors: (a) Teaching Efficacy (b) Knowledge Efficacy • Measurement: five-point response scale, positive and negative items	• Problems with the item wordings and/or the scaling • Possible reasons: among other things, the negative items and reversed scaling (Strongly Agree 1, to Strongly Disagree 5) • Based on these arguments, the authors conclude, regarding their measurement instrument: "we do not recommend its use", [31] (p. 1152), as it is not robust
Vidwans [28]	Modified Culturally Responsive Teaching Self-Efficacy Scale (CRTSE) [87], considers cultural and linguistic background when teaching science	• $N = 76$ • Mixed-methods approach • Analysis: *t*-Test, Correlation (, EFA) • $\alpha = 0.95$ for 40 items • One scale • Measurement: eleven-point answering scale (0 to 10), positive items	• (Rather) theory-based: general pedagogy with five categories of culturally responsive pedagogy [88]: a) Developing a Cultural Diversity Knowledge Base, b) Designing Culturally Relevant Curricula, c) Demonstrating a Cultural Caring and Building a Learning Community, d) Cross-cultural Communications and e) Cultural Congruity in Classroom Instruction • Literature-based (e.g., critical pedagogy) • Adjusting the survey without successful in-depth statistical testing (consequently no subscales or factors regarding culturally responsive pedagogy, as "Conducting a principal components analysis to create underlying sub-scales within the survey did not yield successful results.", [28] (p. 87)) • Not every level of each item was chosen by the test participants (for example, answers only between 4 and 10 instead of 0 and 10) • Despite surveying science teachers, the items are mostly generic • No successful empirical testing of possible factorial structures to prove it as theory-based • Does not meet the rules for construction according to Bandura [9] (for example "I can … " or similar wording)
Rabe et al. [33] Meinhardt et al. [89] Meinhardt et al. [30]	Independent, based on a literature review, generated for physics	• $N = 931$ • Analysis: CFA, IRT • $\alpha = 0.77$–0.86 for overall 59 items • CFA: from good to very good model fit • Four factors: (a) Simplification (b) Experimenting (c) Dealing with Students' Conceptions (d) Dealing with Tasks with two dimensions each (planning and performance) • Measurement: six-point response scale, positive items	• Generated fields of action in physics education, literature-based • Various (pilot) studies to refine the scale • Unsystematically varying obstacles could complicate the interpretation of items (Is the task or the obstacle difficult?) • Separation of the dimensions planning and performance is questionable due to partly found very strong correlations [32] • Depth of content in single aspects (for example experimenting), but partly missing breadth (for example more instructional strategies)

Table A3. Pattern matrix of the sixth rerun of the principal factor analysis (PFA). Loadings ≥0.3 are displayed. Bold indicates the chosen loadings. The items are based on the following initial operationalized subcategories of the PCK model [38]: a: Curriculum Materials, b: Learning Difficulties, c: Needs (Misconceptions, Motivation, Interest, Need), d_1–5: Vertical Curriculum, d_6–10: Horizontal Curriculum, e: Dimensions of Science Learning to Assess, f: Methods of Assessing Science Learning, g: Subject/Science-specific Strategies, h: Activities, i: Representations. The items are numbered after the order in the used questionnaire. Thus, the numbers of items excluded from the PFA are missing.

Item/Factor	1	2	3	4	5	6	7	8	9	10
e_2	**0.697**									
e_6	**0.626**									
e_3	**0.620**									
e_5	**0.600**									
e_4	**0.588**									
e_1	0.466									−0.447
e_7	0.382		**0.336**							
i_3		**0.804**								
i_2		**0.725**								
i_5		**0.639**								
i_4		**0.614**								
d_8			**0.672**							
d_3			**0.573**							
b_4			**0.516**							
d_2				**0.722**						
d_6				**0.582**						
d_7				**0.563**						
d_1				**0.560**						
c_4				**0.385**						
a_3					**0.701**					
a_1					**0.637**					
a_4					**0.558**					
a_2					**0.512**					
h_6						**0.726**				
h_7						**0.701**				
h_5						**0.610**				
h_4						**0.587**				
i_1						**0.444**				
f_2							**−0.617**			
f_1							**−0.541**			
f_3							**−0.536**			
b_2								**−0.694**		
c_1								**−0.583**		
c_2								**−0.577**		
b_3								**−0.480**		
b_1								**−0.459**		
g_1									**0.608**	
g_2									**0.576**	
g_3									**0.480**	
d_9										**−0.600**
d_4										**−0.452**

References

1. Carlson, J.; Daehler, K.R. The Refined Consensus Model of Pedagogical Content Knowledge in Science Education. In *Repositioning Pedagogical Content Knowledge in Teachers' Knowledge for Teaching Science*; Hume, A., Cooper, R., Borowski, A., Eds.; Springer Singapore: Singapore, 2019; pp. 77–92. ISBN 978-981-13-5897-5.

2. Shulman, L.S. PCK: Its genesis and exodus. In *Re-Examining Pedagogical Content Knowledge in Science Education*, 1st ed.; Teaching and Learning in Science Series; Berry, A., Friedrichsen, P., Loughran, J., Eds.; Routledge: New York, NY, USA, 2015; pp. 3–13. ISBN 9781315735665.

3. Forsthuber, B.; Horvath, A.; Almeida Coutinho, A.S.D.; Motiejūnaitė, A.; Baïdak, N. *Science Education in Europe: National Policies, Practices and Research*; Education, Audiovisual and Culture Executive Agency: Brussels, Belgium, 2011; ISBN 978-92-9201-218-2.

4. Niedersächsisches Kultusministerium. Kerncurriculum für die integrierte Gesamtschule Schuljahrgänge 5–10: Naturwissenschaften. Available online: http://db2.nibis.de/1db/cuvo/datei/kc_2012_igs_nws_i.pdf (accessed on 12 October 2017).

5. Statistisches Bundesamt (Destatis). Bildung und Kultur: Allgemeinbildende Schulen. Schuljahr 2017/18. Available online: https://www.destatis.de/DE/Publikationen/Thematisch/BildungForschungKultur/Schulen/ AllgemeinbildendeSchulen2110100187004.pdf?__blob=publicationFile (accessed on 5 November 2018).

6. Statistisches Bundesamt (Destatis). Bildung und Kultur: Allgemeinbildende Schulen. Schuljahr 2008/09. Available online: https://www.destatis.de/GPStatistik/servlets/MCRFileNodeServlet/DEHeft_derivate_ 00006815/2110100097004.pdf (accessed on 6 March 2018).

7. Labudde, P. Fächerübergreifender naturwissenschaftlicher Unterricht—Mythen, Definitionen, Fakten. *Z. Didakt. Naturwiss.* **2014**, *20*, 11–19. [CrossRef]

8. Enochs, L.G.; Riggs, I.M. Further Development of an Elementary Science Teaching Efficacy Belief Instrument: A Preservice Elementary Scale. Presented at the Annual Meeting of the National Association of Research in Science Teaching, Atlanta, GA, USA, 8–11 April 1990.

9. Bandura, A. *Self-Efficacy: The Exercise of Control*; W.H. Freeman and Company: New York, NY, USA, 1997; ISBN 9780716728504.

10. Tschannen-Moran, M.; Woolfolk Hoy, A.; Hoy, W.K. Teacher Efficacy: Its Meaning and Measure. *Rev. Educ. Res.* **1998**, *68*, 202–248. [CrossRef]

11. Bandura, A. Guide for Constructing Self-Efficacy Scales. In *Self-Efficacy Beliefs of Adolescents*; Pajares, F., Urdan, T., Eds.; Information Age Publishing: Greenwich, CT, USA, 2006; pp. 307–337.

12. Baumert, J.; Kunter, M. The COACTIV Model of Teachers' Professional Competence. In *Cognitive Activation in the Mathematics Classroom and Professional Competence of Teachers: Results from the COACTIV Project*; Mathematics Teacher Education 8; Kunter, M., Baumert, J., Blum, W., Klusmann, U., Krauss, S., Neubrand, M., Eds.; Springer: New York, NY, USA, 2013; pp. 25–48. ISBN 978-1-4614-5148-8.

13. Gunning, A.M.; Mensah, F.M. Preservice Elementary Teachers' Development of Self-Efficacy and Confidence to Teach Science: A Case Study. *J. Sci. Teach. Educ.* **2011**, *22*, 171–185. [CrossRef]

14. Palmer, D.; Dixon, J.; Archer, J. Changes in Science Teaching Self-efficacy among Primary Teacher Education Students. *Aust. J. Teach. Educ.* **2015**, *40*, 27–40. [CrossRef]

15. Ramey-Gassert, L.; Shroyer, M.G.; Staver, J.R. A Qualitative Study of Factors Influencing Science Teaching Self-Efficacy of Elementary Level Teachers. *Sci. Ed.* **1996**, *80*, 283–315. [CrossRef]

16. de Laat, J.; Watters, J.J. Science Teaching Self-Efficacy in a Primary School: A Case Study. *Res. Sci. Educ.* **1995**, *25*, 453–464. [CrossRef]

17. Appleton, K.; Kindt, I. Beginning Elementary Teachers' Development as Teachers of Science. *J. Sci. Teach. Educ.* **2002**, *13*, 43–61. [CrossRef]

18. Lumpe, A.; Czerniak, C.; Haney, J.; Beltyukova, S. Beliefs about Teaching Science: The relationship between elementary teachers' participation in professional development and student achievement. *Int. J. Sci. Educ.* **2012**, *34*, 153–166. [CrossRef]

19. Hinterholz, C.W.; Nitz, S. Selbstwirksamkeitserwartungen von angehenden und ausgebildeten Biologielehrkräften: Pilotierung eines neu entwickelten Instruments. In Proceedings of the 21 Internationale Tagung der Fachsektion Didaktik der Biologie (FDdB) im VBIO, Halle (Saale), Germany, 11–14 September 2017; pp. 192–194.

20. Moosbrugger, H.; Kelava, A. *Testtheorie und Fragebogenkonstruktion*; Springer: Berlin, Germany, 2012; ISBN 978-3642200717.

21. Dira Smolleck, L.; Zembal-Saul, C.; Yoder, E.P. The Development and Validation of an Instrument to Measure Preservice Teachers' Self-Efficacy in Regard to The Teaching of Science as Inquiry. *J. Sci. Teach. Educ.* **2006**, *17*, 137–163. [CrossRef]

22. Mavrikaki, E.; Athanasiou, K. Development and Application of an Instrument to Measure Greek Primary Education Teachers' Biology Teaching Self-efficacy Beliefs. *Eurasia J. Math. Sci. Technol. Educ.* **2011**, *7*, 203–213. [CrossRef]

23. Ritter, J.M. The Development and Validation of the Self-Efficacy Beliefs about Equitable Science Teaching and Learning Instrument for Prospective Elementary Teachers. Ph.D. Thesis, Pennsylvania State University, State College, PA, USA, 1999.

24. Roberts, J.K.; Henson, R.K. Self-Efficacy Teaching and Knowledge Instrument for Science Teachers (SETAKIST): A Proposal for a New Efficacy Instrument. Presented at the Annual Meeting of the Mid-South Educational Research Association, Bowling Green, KY, USA, 17–19 November 2000.

25. Walan, S.; Chang Rundgren, S.-N. Investigating Preschool and Primary School Teachers' Self-Efficacy and Needs in Teaching Science: A Pilot Study. *Ceps J.* **2014**, *4*, 51–67.

26. Riese, J. *Professionelles Wissen und professionelle Handlungskompetenz von (angehenden) Physiklehrkräften*; Studien zum Physik- und Chemielernen 97; Logos: Berlin, Germany, 2009.

27. Savran, A.; Çakiroğlu, J. Preservice biology teachers' perceived efficacy beliefs in teaching biology. *Hacet. Üniv. Eğitim Fakültesi Derg.* **2001**, *21*, 105–112.

28. Vidwans, M. Exploring Science Teachers' Self-Efficacy Perceptions to Teach in Ontario's Diverse Classrooms: A Mixed-Methods Investigation. Ph.D. Thesis, University of Western Ontario, London, UK, 2016.

29. Barros, M.A.; Laburú, C.E.; da Silva, F.R. An instrument for measuring self-efficacy beliefs of secondary school physics teachers. *Procedia Soc. Behav. Sci.* **2010**, *2*, 3129–3133. [CrossRef]

30. Meinhardt, C.; Rabe, T.; Krey, O. Formulierung eines evidenzbasierten Validitätsarguments am Beispiel der Erfassung physikdidaktischer Selbstwirksamkeitserwartungen mit einem neu entwickelten Instrument. *Z. Didakt. Naturwiss.* **2018**, *24*, 131–150. [CrossRef]

31. Pruski, L.A.; Blanco, S.L.; Riggs, R.A.; Grimes, K.K.; Fordtran, C.W.; Barbola, G.M.; Cornell, J.E.; Lichtenstein, M.J. Construct Validation of the Self-Efficacy Teaching and Knowledge Instrument for Science Teachers-Revised (SETAKIST-R): Lessons Learned. *J. Sci. Teach. Educ.* **2013**, *24*, 1133–1156. [CrossRef]

32. Meinhardt, C.; Rabe, T.; Krey, O. Quantitative Validierung eines Testinstruments zu Selbstwirksamkeitserwartungen in physikdidaktischen Handlungsfeldern—Erste Ergebnisse. In *Heterogenität und Diversität—Vielfalt der Voraussetzungen im Naturwissenschaftlichen Unterricht*; Sascha, B., Ed.; IPN: Kiel, Germany, 2014; pp. 283–285.

33. Rabe, T.; Meinhardt, C.; Krey, O. Entwicklung eines Instruments zur Erhebung von Selbstwirksamkeitserwartungen in physikdidaktischen Handlungsfeldern. *Z. Didakt. Naturwiss.* **2012**, *18*, 293–315.

34. Neumann, K.; Kind, V.; Harms, U. Probing the amalgam: The relationship between science teachers' content, pedagogical and pedagogical content knowledge. *Int. J. Sci. Educ.* **2019**, *41*, 847–861. [CrossRef]

35. Shulman, L.S. Those Who Understand: Knowledge Growth in Teaching. *Educ. Res.* **1986**, *15*, 4–14. [CrossRef]

36. Shulman, L.S. Knowledge and Teaching: Foundations of the New Reform. *Harv. Educ. Rev.* **1987**, *57*, 1–22. [CrossRef]

37. Gess-Newsome, J. A Model of Teacher Professional Knowledge and Skill Including PCK: Results of the thinking from the PCK Summit. In *Re-Examining Pedagogical Content Knowledge in Science Education*; Teaching and Learning in Science Series; Berry, A., Friedrichsen, P., Loughran, J., Eds.; Routledge: New York, NY, USA, 2015; pp. 28–42. ISBN 9781315735665.

38. Park, S.; Chen, Y.-C. Mapping Out the Integration of the Components of Pedagogical Content Knowledge (PCK): Examples From High School Biology Classrooms. *J. Res. Sci. Teach.* **2012**, *49*, 922–941. [CrossRef]

39. Magnusson, S.; Krajcik, J.; Borko, H. Nature, Sources, and Development of Pedagogical Content Knowledge for Science Teaching. In *Examining Pedagogical Content Knowledge: The Construct and Its Implications for Science Education*; Gess-Newsome, J., Lederman, N.G., Eds.; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1999; pp. 95–132.

40. Park, S. A Study of PCK of Science Teachers for Gifted Secondary Students Going through the National Board Certification Process. Ph.D. Thesis, University of Georgia, Athens, GA, USA, 2005.

41. Park, S.; Oliver, J.S. Revisiting the Conceptualisation of Pedagogical Content Knowledge (PCK): PCK as a Conceptual Tool to Understand Teachers as Professionals. *Res. Sci. Educ.* **2008**, *38*, 261–284. [CrossRef]

42. Park, S.; Oliver, J.S. National Board Certification (NBC) as a Catalyst for Teachers' Learning about Teaching: The Effects of the NBC Process on Candidate Teachers' PCK Development. *J. Res. Sci. Teach.* **2008**, *45*, 812–834. [CrossRef]

43. Grossman, P.L. *The Making of a Teacher: Teacher Knowledge and Teacher Education*; Teachers College Press: New York, NY, USA, 1990; ISBN 0-8077-3047-5.

44. Tamir, P. Subject Matter and Related Pedagogical Knowledge in Teacher Education. *Teach. Teach. Educ.* **1988**, *4*, 99–110. [CrossRef]

45. Velthuis, C.; Fisser, P.; Pieters, J. Teacher Training and Pre-service Primary Teachers' Self-Efficacy for Science Teaching. *J. Sci. Teach. Educ.* **2014**, *25*, 445–464. [CrossRef]

46. Mahler, D.; Großschedl, J.; Harms, U. Opportunities to Learn for Teachers' Self-Efficacy and Enthusiasm. *Educ. Res. Int.* **2017**, *2017*, 4698371. [CrossRef]

47. Schmitz, G. Entwicklung der Selbstwirksamkeitserwartungen von Lehrern. *Unterrichtswissenschaft* **1998**, *26*, 140–157.

48. Yilmaz, H.; Çavaş, P.H. The Effect of the Teaching Practice on Pre-service Elementary Teachers' Science Teaching Efficacy and Classroom Management Beliefs. *Eurasia J. Math. Sci. Technol. Educ.* **2008**, *4*, 45–54. [CrossRef]

49. Schoon, K.J.; Boone, W.J. Self-Efficacy and Alternative Conceptions of Science of Preservice Elementary Teachers. *Sci. Ed.* **1998**, *82*, 553–568. [CrossRef]

50. Yangin, S.; Sidekli, S. Self-Efficacy for Science Teaching Scale Development: Construct Validation with Elementary School Teachers. *J. Educ. Train. Stud.* **2016**, *4*, 54–69. [CrossRef]

51. Park, S. Reconciliation Between the Refined Consensus Model of PCK and Extant PCK Models for Advancing PCK Research in Science. In *Repositioning Pedagogical Content Knowledge in Teachers' Knowledge for Teaching Science*; Hume, A., Cooper, R., Borowski, A., Eds.; Springer Singapore: Singapore, 2019; pp. 117–128. ISBN 978-981-13-5897-5.

52. Schwarzer, R.; Jerusalem, M. Das Konzept der Selbstwirksamkeit. *Z. Pädagogik Beih.* **2002**, *44*, 28–53.

53. Kultusministerkonferenz. Ländergemeinsame Inhaltliche Anforderungen für die Fachwissenschaften und Fachdidaktiken in der Lehrerbildung: Beschluss der Kultusministerkonferenz vom 16.10.2008 i.d.F. vom 06.10.2016. No Longer Available Online.

54. Labudde, P. (Ed.) *Fachdidaktik Naturwissenschaft: 1.-9. Schuljahr*; Haupt: Bern, Switzerland, 2010.

55. Nerdel, C. *Grundlagen der Naturwissenschaftsdidaktik: Kompetenzorientiert und Aufgabenbasiert für Schule und Hochschule*; Springer: Berlin, Germany, 2017.

56. Kultusministerkonferenz. Bildungsstandards im Fach Biologie für den Mittleren Schulabschluss: Beschluss vom 16.12.2004. Available online: https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/ 2004/2004_12_16-Bildungsstandards-Biologie.pdf (accessed on 12 October 2017).

57. Kultusministerkonferenz. Bildungsstandards im Fach Chemie für den Mittleren Schulabschluss: Beschluss vom 16.12.2004. Available online: http://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/ 2004/2004_12_16-Bildungsstandards-Chemie.pdf (accessed on 12 October 2017).

58. Kultusministerkonferenz. Bildungsstandards im Fach Physik für den Mittleren Schulabschluss: Beschluss vom 16.12.2004. Available online: http://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/ 2004/2004_12_16-Bildungsstandards-Physik-Mittleren-SA.pdf (accessed on 12 October 2017).

59. Niedersächsisches Kultusministerium. Kerncurriculum für das Gymnasium Schuljahrgänge 5–10: Naturwissenschaften. Available online: http://db2.nibis.de/1db/cuvo/datei/nw_gym_si_kc_druck.pdf (accessed on 12 October 2017).

60. Grossman, P.; Thompson, C. Learning from curriculum materials: Scaffolds for new teachers? *Teach. Teach. Educ.* **2008**, *24*, 2014–2026. [CrossRef]

61. Schwarzer, R.; Schmitz, G. Skala zur Lehrer-Selbstwirksamkeitserwartung (WIRKLEHR). In *Skalen zur Erfassung von Lehrer- und Schülermerkmalen: Dokumentation der Psychometrischen Verfahren im Rahmen der Wissenschaftlichen Begleitung des Modellversuchs Selbstwirksame Schulen*; Schwarzer, R., Jerusalem, M., Eds.; Freie Universität Berlin: Berlin, Germany, 1999; pp. 60–61.

62. Schulte, K.; Watermann, R.; Bögeholz, S. Überprüfung der faktoriellen Validität einer multidimensionalen Skala der Lehrer-Selbstwirksamkeitserwartung. *Empir. Pädagogik* **2011**, *25*, 232–256.

63. Handtke, K.; Oberle, M.; Bögeholz, S. *Selbstwirksamkeitserwartungen zum Unterrichten von Naturwissenschaften*; Unpublished Measurement Instrument; Georg-August-Universität Göttingen: Göttingen, Germany, 2017.

64. Handtke, K.; Oberle, M.; Bögeholz, S. *Subjektive Einschätzung des Fachwissens in den Naturwissenschaften*; Unpublished Measurement Instrument; Georg-August-Universität Göttingen: Göttingen, Germany, 2017.

65. Henson, R.K.; Roberts, J.K. Use of Exploratory Factor Analysis in Published Research: Common Errors and Some Comment on Improved Practice. *Educ. Psychol. Meas.* **2006**, *66*, 393–416. [CrossRef]

66. Fabrigar, L.R.; Wegener, D.T.; MacCallum, R.C.; Strahan, E.J. Evaluating the Use of Exploratory Factor Analysis in Psychological Research. *Psychol. Methods* **1999**, *4*, 272–299. [CrossRef]

67. Bühner, M. *Einführung in die Test- und Fragebogenkonstruktion*, 3rd ed.; Pearson Studium: München, Germany, 2011.

68. Field, A. *Discovering Statistics Using IBM SPSS Statistics: And Sex and Drugs and Rock 'n' Roll*, 4th ed.; SAGE: Los Angeles, CA, USA, 2013.

69. Conway, J.M.; Huffcutt, A.I. A Review and Evaluation of Exploratory Factor Analysis Practices in Organizational Research. *Organ. Res. Methods* **2003**, *6*, 147–168. [CrossRef]

70. Pospeschill, M. *Testtheorie, Testkonstruktion, Testevaluation: Mit 77 Fragen zur Wiederholung*; utb.de-Bachelor-Bibliothek 3431; Ernst Reinhardt: München, Germany, 2010; ISBN 978-3825234317.

71. Guttman, L. Some necessary conditions for common factor analysis. *Psychometrika* **1954**, *19*, 149–161. [CrossRef]

72. Kaiser, H.F.; Dickman, K. Analytic determination of common factors. *Am. Psychol.* **1959**, *14*, 425–439.

73. Cattell, R.B. The Scree Test for the Number of Factors. *Multivar. Behav. Res.* **1966**, *1*, 245–276. [CrossRef] [PubMed]

74. Bortz, J.; Döring, N. *Forschungsmethoden und Evaluation: Für Human- und Sozialwissenschaftler*; Springer: Berlin, Germany, 2015.

75. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*; Erlbaum: Hillsdale, NJ, USA, 1988.

76. Benjamini, Y.; Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc.* **1995**, *57*, 289–300. [CrossRef]

77. Kaiser, H.F.; Rice, J. Little Jiffy, Mark IV. *Educ. Psychol. Meas.* **1974**, *34*, 111–117. [CrossRef]

78. Gibson, S.; Dembo, M.H. Teacher Efficacy: A Construct Validation. *J. Educ. Psychol.* **1984**, *76*, 569–582. [CrossRef]

79. Riggs, I.M.; Enochs, L.G. Toward the Development of an Elementary Teacher's Science Teaching Efficacy Belief Instrument. *Sci. Ed.* **1990**, *74*, 625–637. [CrossRef]

80. Kazempour, M. The interrelationship of science experiences, beliefs, attitudes, and self-efficacy: A case study of a pre-service teacher with positive science attitude and high science teaching self-efficacy. *Eur. J. Sci. Math. Educ.* **2013**, *1*, 106–124. [CrossRef]

81. Weinstein, C.S. Preservice teachers' expectations about first year of teaching. *Teach. Teach. Educ.* **1988**, *4*, 31–40. [CrossRef]

82. Krüger, D.; Kloss, L.; Cuadros, I. Was macht "gute" Biologielehrkräfte aus?: Befragungen von Lehrenden in der Didaktik der Biologie und Biologie-Lehramtsstudierenden an deutschen Hochschulen. *Z. Didakt. Biol.* **2009**, *17*, 63–88.

83. Kultusministerkonferenz. Ländergemeinsame inhaltliche Anforderungen für die Fachwissenschaften und Fachdidaktiken in der Lehrerbildung: Beschluss der Kultusministerkonferenz vom 16.10.2008 i.d.F. vom 16.05.2019. Available online: https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2008/2008_10_16-Fachprofile-Lehrerbildung.pdf (accessed on 3 June 2019).

84. Steffen, B.; Hößle, C. Diagnose von Bewertungskompetenz durch Biologielehrkräfte–Negieren eigener Fähigkeiten oder Bewältigen einer Herausforderung? *Z. Didakt. Naturwiss.* **2015**, *21*, 155–172. [CrossRef]

85. Alfs, N.; Heusinger von Waldegge, K.; Hößle, C. Bewertungsprozesse verstehen und diagnostizieren. *Z. Interpret. Schul-Und Unterrichtsforsch.* **2012**, *1*, 83–112.

86. Bleicher, R.E. Revisiting the STEBI-B: Measuring Self-Efficacy in Preservice Elementary Teachers. *Sch. Sci. Math.* **2004**, *104*, 383–391. [CrossRef]

87. Siwatu, K.O. Preservice teachers' culturally responsive teaching self-efficacy and outcome expectancy beliefs. *Teach. Teach. Educ.* **2007**, *23*, 1086–1101. [CrossRef]

88. Gay, G. Preparing for Culturally Responsive Teaching. *J. Teach. Educ.* **2002**, *53*, 106–116. [CrossRef]

89. Meinhardt, C.; Rabe, T.; Krey, O. Selbstwirksamkeitserwartungen in Physikdidaktischen Handlungsfeldern. Skalendokumentation. Version 1.0 (Februar 2016). Available online: http://www.pedocs.de/volltexte/2016/11818/pdf/Meinhardt_2016_Selbstwirksamkeitserwartungen.pdf (accessed on 4 March 2019).