

Impact of an Experiment-Based Intervention on Pre-Service Primary School Teachers' Experiment-Related and Science Teaching-Related Self-Concepts

Melanie Marita Beudels^{1*}, Angelika Preisfeld¹, Karsten Damerau²

¹ University of Wuppertal, GERMANY

² European University of Flensburg, GERMANY

*Corresponding Author: melanie.beudels@uni-wuppertal.de

Citation: Beudels, M. M., Preisfeld, A., & Damerau, K. (2022). Impact of an experiment-based intervention on pre-service primary school teachers' experiment-related and science teaching-related self-concepts. *Interdisciplinary Journal of Environmental and Science Education*, 18(1), e2258. <https://doi.org/10.21601/ijese/11323>

ARTICLE INFO

ABSTRACT

Received:
30 August 2021

Accepted:
7 October 2021

Teachers' academic self-concept is considered an important factor influencing their professional competence. Regarding primary science education, positive science (teaching) related self-concepts might encourage teachers to plan and teach 'minds on' experiment-based science lessons leading to deep learning processes. However, research on pre- and in-service primary teachers' self-concepts and influencing factors, such as previous experimental experience, is scarce. Thus, this study investigates the impact of an experiment-based intervention on pre-service primary school teachers' experiment-related self-concept and self-concepts on planning and teaching experiment-based lessons. The evaluation followed a quasi-experimental, longitudinal (pre-post) design with an experimental group of N = 158 pre-service primary teachers and a baseline group (N = 44), not attending the course. According to the results, pre-service teachers gained little to moderate experimental experience in school and studying at university. Besides, the pre-service teachers with a science major gained significantly more experimental experience than those with other majors during their time at the university. Significant, positive correlations were found between previous experimental experiences and the self-concepts examined in this study. While self-concepts did not change in the baseline group, they increased significantly in the experimental group. One reason for this could be the perception of competence, as the findings reveal positive correlations between changes in self-concepts and perceived experimental competence during the intervention. Regarding the impact of the variable 'course format' on reinforcing the self-concepts, participants of the intensive block format seem to have a slight advantage compared to pre-service teachers attending the traditional, weekly course format. Furthermore, the results indicate that the course is equally beneficial for pre-service teachers with and without a science major.

Keywords: academic self-concept, experimental experience, experimental competencies, professionalization of pre-service primary school teachers

INTRODUCTION

Primary school teachers need specific professional competencies to plan, teach and reflect experiment-based science lessons that encourage pupils to 'minds-on' scientific investigations (Minner et al., 2010; National Academies of Sciences, Engineering, and Medicine, 2015). Competent science teachers depend on syntactic knowledge, e.g. knowledge of the

control-of-variables strategy (Haslbeck, 2019; Schwichow et al., 2016) and knowledge of the steps of experimentation. Pedagogical content knowledge – including knowledge about appropriate instructional strategies for experiment-based lessons and (mis)conceptions as well as possible learning difficulties of pupils – is also required (Baumert &

Kunter, 2013; Franken et al., 2020; KMK, 2008/2019). However, professional knowledge alone is not sufficient for planning and teaching inquiry-based science lessons. Primary school teachers should also be able to generate and test hypotheses themselves (KMK, 2008/2019). According to Capps et al. (2012), inquiry professional development benefits from experimental experiences. Experiences (of competence) can, in turn, influence teachers' academic self-concept (Bong & Skaalvik, 2003; Dickhäuser, 2006; Shavelson et al., 1976), which is an essential predictor of teaching effectiveness and teaching behavior (Guskey, 1988; Yeung et al., 2014). It is defined as "self-evaluated perception of [teachers'] professional knowledge" (Paulick et al., 2016, p. 174), abilities and performance (Dickhäuser, 2006) and is considered a multidimensional, hierarchical construct (Marsh et al., 1988; Paulick et al., 2016; Shavelson et al., 1976).

Perceptions of competencies regarding scientific inquiry and planning/teaching inquiry-based science lessons can be located on lower hierarchical levels of teachers' academic self-concept (Atzert et al., 2020; Franken et al., 2020; Paulick et al., 2016). The experiment-related self-concept is a person's perception of her/his abilities and skills in experimentation (Buse et al., 2018; Damerau, 2012; Franken et al., 2020). Based on the model of experimentation competency by Schreiber et al. (2009), latter authors differentiate between the dimensions of planning (hypothesize, develop an experimental setup), conducting (handling of equipment, measurement, record results), and interpreting (analyze data, draw conclusions, verify/falsify hypotheses) experiments. However, in other studies, the two subscales 'planning & interpreting' and 'conducting' experiments (Atzert et al., under review; Rautenstrauch & Busker, 2020) or an overall scale (Peschel, 2018) are used. Within the field of competence in teaching, a distinction can be made between self-concepts on planning, teaching, and reflecting experiment-based lessons referring to the work of Bosch (2006) and Merkens (2010). The phase of planning lessons includes, among other things, determining learning objectives and activities, whereas teaching comprises interacting with the pupils or dealing with deviations from the plan. The phase of reflecting covers activities such as drawing consequences from the evaluation of the lesson (Bosch, 2006; Merkens, 2010).

According to Yeung et al. (2014), teachers with a positive self-concept are more likely to motivate their pupils and initiate deep learning processes than teachers with a negative self-concept. Low domain-specific self-concepts can lead teachers to avoid certain topics or activities in class (Appleton, 2007;

Damerau, 2012). Besides, teachers' self-concept was found to predict the usage of new instructional practices (Guskey, 1988) and indicate teachers' performance in professional knowledge (Paulick et al., 2016). Regarding these findings, it is to be assumed that primary school teachers with a high experiment-related self-concept and self-concept on planning/teaching experiment-based lessons are more willing and able to arrange challenging experiment-based lessons than teachers with low self-concepts (Franken et al., 2020). Furthermore, these self-concepts can be used as an indicator for factual experimental competencies and syntactic knowledge (Paulick et al., 2016; Schreiber et al., 2016).

To support primary school teachers in planning and teaching 'minds on' experiment-based science lessons, teacher education programs should aim at enhancing pre-service teachers' corresponding self-concepts as early as possible. However, there is a lack of research on (the development of) pre- and in-service teachers' academic self-concepts and influencing factors (Franken, 2020; Paulick et al., 2016; Sorge et al., 2019).

Regarding the experiment-related self-concept, it was found that both secondary school students (Buse, 2017; Damerau, 2012; Rautenstrauch & Busker, 2020) and pre-service biology, chemistry (secondary school), and primary science and social studies teachers within the master program (Franken et al., 2020) assess their skills in 'conducting experiments' higher than in the dimensions of planning and interpreting experiments. This could be due to few opportunities for experimentation in school – especially regarding the steps of planning and interpreting – and dominating 'recipe-type' activities, where learners follow prescribed procedures (Seidel et al., 2007; Yip, 2001). Thus, many pupils and university students show deficits in planning and evaluating experiments (Fleischer et al., 2020; Hilfert-Rüppell et al., 2013). Among other things, this applies to formulating usable research questions (Hofstein et al., 2005), understanding and adopting the control-of-variables strategy (Emereole, 2009; Hilfert-Rüppell et al., 2013), recording data, and attending to the hypothesis to draw conclusions (Germann & Aram, 1996). So far, however, it has not been explicitly investigated whether there is a positive correlation between teachers' past experimental experiences and their experiment-related self-concepts and self-concepts on planning and teaching experiment-based lessons.

Longitudinal studies concerning these self-concepts are also scarce. It has been shown that experiment-based educational settings can reinforce

secondary school students' (Buse et al., 2018; Damerau, 2012) and pre-service primary science and social studies teachers' (Peschel, 2018) experiment-related self-concept. A similar development was found for pre-service primary teachers' self-concept on the control-of-variables strategy in the study by Haslbeck (2019). In contrast, the experiment-related self-concept of pre-service biology, chemistry, and primary science and social studies teachers did not change significantly during their practical semester in the master program (Franken, 2020).

To date, it has rarely been explored which variables and factors in the course of study exactly play a role in forming teachers' self-concepts. In addition to experiences of competence (Dickhäuser, 2006; Shavelson et al., 1976), individual feedback on performance and a supportive learning environment could be of relevance (Bong & Skaalvik, 2003; Lüdtke et al., 2005; Möller & Trautwein, 2015). According to the results of Paulick et al. (2017) and Sorge et al. (2019), pre-service teachers' self-concepts are also formed using external and internal frames of reference (Marsh et al., 1988). Due to the findings of Franken et al. (2020) and Nadelson et al. (2013), the major field of study can also be assumed to be an important variable influencing science (teaching) related self-concepts. This factor is of great relevance as many primary school teachers have to teach science without appropriate training in this area (Porsch & Wendt, 2016). However, it has not yet been investigated whether pre-service primary teachers who do not have a science major have lower experiment-related and science teaching-related self-concepts than those with this major. Besides, it is not known whether the academic self-concepts of these two groups of students change differently as a result of university courses focusing on experimentation. Especially with a view to pre-service primary teachers without a science major, who might be interested in further training in the area of science (education), intensive block course formats could be attractive. Compared to 'traditional' courses taking place once or twice a week, their "main characteristic [...] appears to be that an equal number of class hours is delivered in more concentrated bursts" (Burton & Nesbit, 2008, p. 5). The question arises as to whether the course format influences the development of pre-service teachers' academic self-concepts. Here, too, research is minimal. Schaal & Randler (2004) and Hilkenmeier & Sommer (2014) revealed higher scores in perceived competence for pre-service teachers' attending a block course format. Since experiencing competence is described as a determinant forming the self-concept (Bong & Skaalvik, 2003; Shavelson et al., 1976), these findings could suggest that block courses are more likely to

reinforce pre-service teachers' self-concepts than weekly course formats.

Aim of the Study and Research Questions

The following research questions were derived from the lack of empirical research on the status quo and the impact of university courses on pre-service primary teachers' experiment-related self-concepts and self-concepts on planning and teaching experiment-based lessons described above.

Since academic self-concepts can be fed by experiences with the environment and experienced competence in the past (Bong & Skaalvik, 2003; Shavelson et al., 1976), the first research questions are:

RQ1 (*Experimental experience*)

RQ1.1: How much experimental experience have pre-service primary-teachers gained in primary school, in secondary school, and while studying at university? Are there any differences between students with a science major and those with a different major?

RQ1.2: Is there a positive correlation between previous experimental experiences and the self-concepts examined in this study before starting with the intervention?

The next step was to investigate whether a university course (intervention) that focuses on experimentation impacts pre-service primary teachers' *experiment-related self-concept* and *self-concepts on planning and teaching experiment-based lessons*.

RQ2 (*Effects of the intervention*)

RQ2.1: Does participation in the intervention lead to a change in self-concepts compared to non-participation?

RQ2.2: Do participants of the intervention show differences in developing these self-concepts depending on the course format (weekly/traditional or block/intensive)?

RQ2.3: Do participants of the intervention show differences in developing these self-concepts depending on their major field of study (science major or not)?

Assuming that changes in these self-concepts result from the experience of competence during the intervention (Bong & Skaalvik, 2003; Dickhäuser, 2006), the third research question is:

RQ3 (*Perceived experimental competencies during the intervention*)

Is there a positive correlation between perceived competence in planning, conducting, and interpreting experiments during the intervention and a change of self-concepts from pre- to posttest?

METHOD

Study Design and Data Collection

The present study examined two different university course formats (intervention) with a quasi-experimental longitudinal design with two points of data collection and two main study groups of pre-service primary teachers (see Figure 1). The experimental group (EG) took part in the intervention offered either in a 'traditional' course format weekly during the lecture period or as an intensive, four-day block course during semester break. The control group, hereinafter referred to as the baseline group (BG), did not attend the course. Its purpose was to determine the effects of other factors on the dependent variables studied (e.g. the impact of participating in other courses or the practice semester). To ensure comparability of the results, the BG was also divided into two subgroups: Like participants in the weekly course format, one subgroup took the survey (see chapter Measurement Instrument) at the beginning and end of the lecture period. The other BG subgroup filled in the questionnaire as if they were taking a block course. All participants were asked to take two online questionnaires (tool: SoSci Survey; Leiner, 2006) via the news forum of a digital learning room. The pretest took place within ten days prior to the intervention and a posttest within ten days after completing the last course session.

As many factors influence an intervention, it was always held in the same classroom and taught by the same lecturer and student assistant. For this reason,

the influence of these factors is negligible for the evaluation.

Educational Concept and its Curricular Framework

This is a summary of the curricular framework and educational concept, focusing on reinforcing pre-service primary teachers' experimental competencies in experimentation and planning as well as teaching experiment-based lessons. For a detailed description of this newly developed intervention, refer to Beudels, Schilling, and Preisfeld (under review).

The intervention can be attended by pre-service primary school teachers in the bachelor's and master's programs. Due to the examination regulations, most of these students can only take the course voluntarily, i.e. not earn any credit points. This applies for example to students of science and social studies in their master's program and students focusing on special education, music, or theology. The reason to make this intervention available to all pre-service primary teachers is that many of them have to teach science (within the subject of science and social studies) without an adequate university education (Porsch & Wendt, 2016). As part of a pilot project, bachelor students having a major field of study in science (biology, chemistry, physics) and technology were able to attend the course as a replacement for a mandatory module component.

The intervention consists of twelve 100-minute sessions each for both course format types (see Figure 1). To guarantee a good and similar supervisory relationship, there was a maximum of 30 students per course. Based on previous research findings (Kleickmann et al., 2006; Schwichow et al., 2016), a moderate-constructivist learning environment with tutorial support was chosen: Accompanied by the lecturer and phases of reflection,

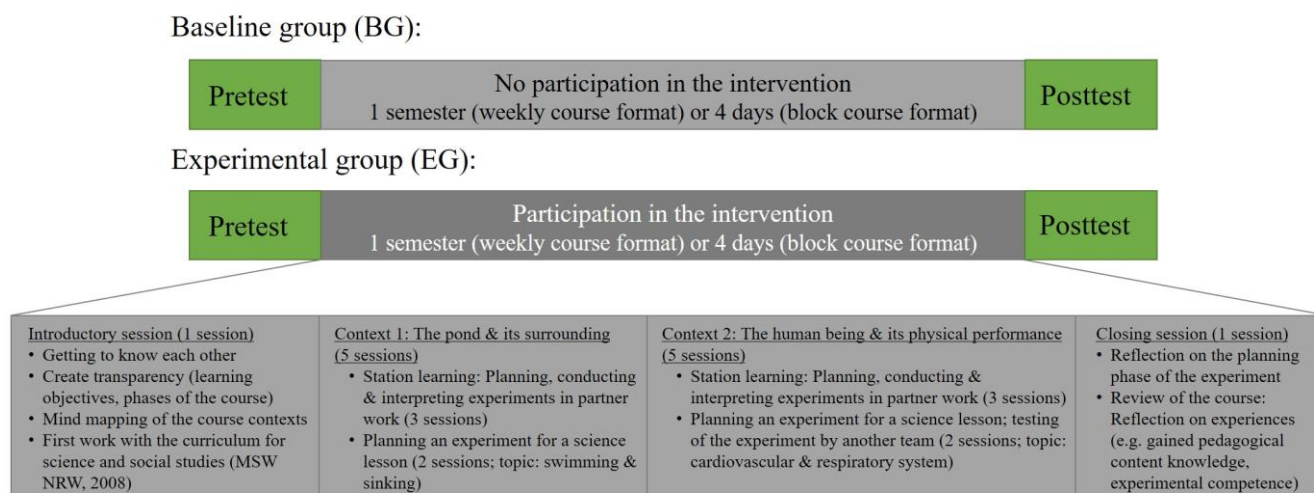


Figure 1. Overview of the study design (top) and the sequence as well as the contents of the course sessions (bottom)

many opportunities are given to check previous knowledge, to further develop competencies and to exchange ideas with fellow students. Framed by an introductory and a closing session, two five-session-long contexts from children's living environment, 'the pond and its surrounding' and 'the human being and its physical performance', form the framework of the activities described below. Each context block was divided into station learning and planning an experiment for a science lesson (see Figure 1).

1. *Experimentation at stations (three sessions each)*: Questions such as 'How can water striders walk on water?' or 'Can I recognize food by taste if I cannot see and smell it?' form the starting point of station learning in partner work. There are two to four stations per session with several sets of experimental materials. In the following, a 'pedagogical biplane' (Tolsdorf & Markic, 2018; Wahl, 2013) is applied: Being in the role of learners, the pre-service teachers can learn science content, learn about scientific inquiry (e.g. principles of systematic observations), and learn to do inquiry (e.g. skills to handle measuring devices; Gyllenpalm & Wickman, 2011) through experimentation. From a teacher's perspective, experimentation at stations can simultaneously be used as a model for a possible learning setting in science lessons (learning to teach science, Gyllenpalm & Wickman, 2011).

Worksheets located at the stations are designed so that the participants undergo all three phases of experiments: They are asked to formulate hypotheses on a research question, plan a suitable experimental setup, interpret the results of the experiments and draw conclusions. This design intends to strengthen the experimental competencies, especially regarding the phases of planning and interpreting experiments based on the identified shortcomings in these phases described above. To avoid excessive demands on the participants – e.g. due to a lack of routine (Girwidz, 2020) – strongly guided experimentation (prescribed procedures) is applied at the beginning of the course. Later, more open, 'minds-on' experimental formats, e.g. including the phase of planning an experiment autonomously, are also used.

2. *Debriefing and reflection regarding professional competencies and teaching practice*: After each experimental phase, the findings are collected and reflected on in the plenum. Guided by the lecturer, the participants are asked to briefly summarize the process of gaining scientific knowledge: What was the assumption? What was done? What was observed? What can be conducted from this? Difficulties encountered in experimentation are collected to reflect possible solutions to similar problems in the classroom (e.g.

uncertainties regarding the handling of experimental equipment; Kurth & Wodzinski, 2020). The lecturer also uses this phase to discuss different forms of classroom experiments or provide theoretical input regarding syntactic knowledge (e.g. control-of-variables strategy; Schwichow et al., 2016).

3. *Planning an experiment for a science lesson (two sessions each)*: Once in each context, the teams plan a classroom experiment. They are faced with the challenge of applying their previous experimental experiences and the findings from the reflection phases in a situation typical of their profession (Kirsch, 2020). Since planning an experimental lesson is a complex process (Nerdel, 2017), only particular aspects of planning are taken up in the course. The task is to design a worksheet or researcher's protocol for an experiment to be used in a science lesson at primary school. Grade level and the type of experiment can be freely selected; typical steps of gaining scientific knowledge via experimentation (see above) should be integrated. In parallel, teaching-related considerations, e.g. learning objectives and dealing with emerging difficulties, are recorded on a poster. Planning aids include a list of operators for formulating learning objectives and excerpts from teaching materials.

4. *Reflection and feedback*: These planning phases are again reflected in the plenum. In addition, all groups receive written feedback from the lecturer on their teaching-related considerations, the experiment, and the worksheet. After finishing the planning activities within context two, each group also conducts the experiment of another team and gives short written feedback using predetermined feedback rules.

Sample

The study was carried out from winter semester 2017/2018 until the end of summer semester 2019 at a university in North Rhine-Westphalia (Germany). 238 pre-service primary school teachers participated. 191 of them formed the EG, participating in the intervention, whereas 47 persons belonged to the BG. The course took place four times weekly and four times in a block format during the survey period. Data records of participants who answered the survey at none or one of the two measurement times were excluded. Hence, data analysis (see below) was performed with a total sample size of $N_{\text{total}} = 202$ ($N_{\text{EG}} = 158$; $N_{\text{BG}} = 44$). 79.2% of the participants were enrolled in bachelor's degree programs and 20.8% were master students. When the posttest started, the mean age was 22.65 years ($SD = 2.93$ years). The gender distribution of 90.1% female participants represents the high percentage of female primary school teachers in

Germany (89.4% female teachers in the school year 2019/2020; Federal Statistical Office, 2020). 55.7% ($N_{EG, weekly} = 88$) of the participants in the EG completed the weekly course format, 44.3% ($N_{EG, block} = 70$) attended the block format. To ensure comparability, BG participants were also divided in a 'weekly' ($N_{BG, weekly} = 21$; i.e. 47.7%) and 'block' subgroup ($N_{BG, block} = 23$; i.e. 52.3%). While 66.5% of the participants in the EG stated that they study primary school teaching with a science major (SciMaj), the other 33.5% were enrolled with other majors (non-SciMaj), e.g. social or religious studies. In the BG, 61.4% had a science major, 38.6% studied other majors. Almost half of the participants (e.g. all students without a science major) could only take part in the intervention voluntarily ($N_{EG, voluntary} = 72$; i.e. 45.6%). As a pilot project, the other half ($N_{EG, mandatory} = 86$; i.e. 54.4%) participated as part of a mandatory module (see above).

Measurement Instrument

Experimental experience, experiment-related self-concepts, self-concepts on planning and teaching experiment-based lessons, and perceived experimental competencies during the intervention were investigated using five-point Likert-type scales from 1 = *strongly disagree* to 5 = *strongly agree*. Examples of items associated with these scales are shown in Table A1 (Appendix). As the test language was German, the items are listed in their original form and translated versions. Exploratory factor analyses (principal axis factor analysis with varimax rotation; Bühner, 2021) were performed to examine the construct validity of the subscales presented below (see also Table A1) since the items were self-constructed or adapted from scales used in a survey with another study group (Damerau, 2012).

Via Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy (Kaiser & Rice, 1974) and Bartlett's test of sphericity (Bartlett, 1951), adequacy of the data for factor analysis was checked previously. The number of extracted factors was determined regarding theoretical considerations (Bühner, 2021) and the total variance of the items explained by the factors (Janssen & Laatz, 2017). Then the items were assigned to the factor on which they load the most, whereby factor loading values $\lambda \leq .50$ were not considered (Backhaus et al., 2018). Items with cross-loadings $\geq .40$ (Noorman, 2017) were also excluded. Factor loadings $\lambda < 0.3$ are not reported (Fromm, 2012).

In the pretest, *experimental experience* in primary school, in secondary school, and while studying at university (RQ1) was measured using two items each (see Table A1; self-construction). The factor analysis results support this division into three subscales (see

Table A2; three-factor-solution explaining 79.36 % of total variance). As shown in Table A1, the reliability of the constructs in the form of internal consistency is good (Cronbach's $\alpha \geq .8$) or excellent ($\alpha \geq .9$; George & Mallery, 2003) respectively.

At both test times, *experiment-related self-concepts* were quantified using two subscales with items adapted from Damerau (2012). The original scale – consisting of three subscales ('planning', 'conducting' and 'interpreting') – could not be confirmed by the factor analysis with two factors being extracted (see Table A3). Based on the factor analysis results and in accordance with Atzert et al. (under review), the subscales 'planning' and 'interpreting' were combined into one subscale, the self-concept on planning & interpreting experiments (six items; see Table A1). It focuses on self-assessed abilities associated with theoretical considerations such as making predictions, developing suitable setups for experiments, and analyzing experimental observations (Rautenstrauch & Busker, 2020; Schreiber et al., 2009). In the case of the self-concept on conducting experiments (three items), the focus is on psycho-motor skills required in the phase of conducting an experiment (Schreiber et al., 2009). With Cronbach's α values $\geq .8$ in pre- and posttest, the internal consistency for both subscales is satisfactory.

Self-concepts on planning (three items; self-constructed) and *teaching* (three items; self-constructed) *experiment-based lessons* were also queried at both measurement times. A subdivision into the dimensions of planning and teaching is based on the work of Bosch (2006), Merkens (2010), and the standards of educational sciences determined by the Standing Conference of the Ministers of Education and Cultural Affairs in Germany (KMK, 2004). As shown in Table A4, the factor analysis results also support this division into two subscales (two-factor-solution that explains 78.59 % of total variance). The reliability of both subscales is good (Cronbach's $\alpha \geq .8$) or excellent ($\alpha \geq .9$; see Table A1) respectively.

To answer RQ3, the EG was also asked in the posttest for *perceived experimental competencies* during the intervention. All items of this scale were adapted from Damerau (2012), who operationalized the sub-competencies 'planning', 'conducting' and 'interpreting' experiments in the model of experimental competence by Schreiber et al. (2009). As with the construct of experiment-related self-concept, a two-factor solution was chosen based on the factor analytical results (see Table A5), resulting in the two subscales *perceived competence in planning & interpreting experiments* (six items) and *perceived competence in conducting experiments* (three items). Three items had to be removed

because the factor loading was too low ($\lambda \leq .50$). The reliability of the first subscale is good ($\alpha = .816$), that of the second is acceptable ($\alpha = .707$; George & Mallery, 2003).

Data Analysis

All data were analyzed using SPSS Statistics (version 27). Since the function of reminding participants to submit an answer for each item was set in the survey tool, no missing values had to be replaced in the records. With tight constructs, as is the case here, item discriminability values r_{it} between .3-.5 are classified as medium, values of $r_{it} > .5$ as high (Döring & Bortz, 2016). Therefore, items that had a $r_{it} < .3$ in the pre- or posttest were removed before further analysis steps were carried out. Internal consistency was measured with Cronbach's α (Döring & Bortz, 2016). In all subscales, Cronbach's α -value is satisfactory with $\alpha \geq .7$ (see above; George & Mallery, 2003).

Since all total scores – normed to a maximum of 5 – show a Gaussian distribution, parametric methods could be used for the cross-sectional and longitudinal comparisons now described. Pretest mean values of the scales examined were first considered to be able to exclude floor and ceiling effects (Döring & Bortz, 2016).

Based on Damerau (2012), the following evaluation procedure was carried out for longitudinal comparisons between two groups: First, independent samples t -tests (Janssen & Laatz, 2017) were used to check data of two groups for pretest differences.

In case of no significant pretest differences, two-way (RQ2.1 and 2.3) or rather three-way (RQ2.2) repeated measures analyses of variance (ANOVAs; Rasch et al., 2014b) were conducted to detect significant interaction group/treatment*reference time (factor 1 (within-subject): reference time; factor 2: between-subject factor defining the main comparison group/treatment; factor 3 (RQ2.2): further between-subject factor the impact of which should be factored out). The influence of the latter factor ('type of participation' (voluntary/mandatory) or 'major' (SciMaj/non-SciMaj)) was factored out to be able to trace back possible changes in self-concepts to the impact of the course format. Those between-subject factors – nominally scaled – could not be used as covariates since metrically scaled covariates are required for ANCOVAs (Backhaus et al., 2018). In addition, it was impossible to include both factored-out factors simultaneously as students without a science major were not obliged to participate, resulting in a subgroup sample size of $N_{\text{non-SciMaj} + \text{mandatory participation}} = 0$. The interaction effects – representing the combined effects of factors on the dependent variable (Rasch et al., 2014b) – show

whether the self-concepts of the observed groups develop significantly differently over time.

In the case of pretest differences, analyses of covariance (ANCOVAs; Backhaus et al., 2018) with the posttest sum scores as the dependent variable, the pretest sum scores as the covariate, and the independent variable as fixed (between-subject) factor were conducted (one fixed factor to answer RQ2.1 (treatment) and 2.3 (major); two fixed factors (course format and factored-out variable: 'type of participation' or 'major', RQ2.2)). By integrating the covariate and the second fixed factor, their impact on the dependent variable was factored out. The so-called main effect of the first fixed factor (RQ2.1: treatment, RQ2.2: course format, RQ2.3: major) was examined to detect differences in the development of the self-concepts between the groups.

Paired samples t -tests were then used to determine whether pre- and posttest means differ significantly within a group (Janssen & Laatz, 2017). To be able to trace back possible changes in self-concepts to the impact of the course format (RQ2.2), the influence of the variables 'type of participation' and 'major' was factored out as follows: Instead of paired samples t -tests, two-way repeated measures ANOVAs with 'time' as within-subject factor and 'type of participation' or 'major' as between-subject factor were employed. The main effect was examined to assess the impact of the factor 'time' on the dependent variable (Rasch et al., 2014b).

Before carrying out the independent samples t -tests and ANCOVAs, Levene's tests were performed to check the equality of variances of the two populations. In case of inequality of variances, Welch's tests were conducted (Janssen & Laatz, 2017). Partial eta squared (η_p^2) is given as the measure of effect size for paired samples t -tests and the analyses of variance (Rasch et al., 2014a, b). Values of $.01 \leq \eta_p^2 < .06$ can be interpreted as small effects, $.06 \leq \eta_p^2 < .14$ as medium, and $\eta_p^2 \geq .14$ as large effects (Cohen, 1988). For independent samples t -tests omega squared (ω^2) was calculated to assess the effect size ($.01 \leq \omega^2 < .06$: small effect; $.06 \leq \omega^2 < .15$: medium effect; $\omega^2 \geq .15$: large effect; Albert & Koster, 2002). Effect sizes are only reported in the text if the results are significant.

To answer RQ3, self-concept changes were first calculated by determining the differences between pre- and posttest for each participant. Pearson r correlation was then used to examine the strength and direction of the relationship (Janssen & Laatz, 2017) between perceived competence in planning, conducting, and interpreting experiments during the intervention (posttest) and change in self-concepts from pre- to posttest. A Pearson correlation

Table 1. Mean values (M) and standard deviations (SD) of the participants' experimental experience. p -values and effect sizes ω^2 are given for the comparison of participants with a science major (SciMaj) and with other majors (non-SciMaj)

	Total sample		SciMaj group		non-SciMaj group		p	ω^2
	M	SD	M	SD	M	SD		
Primary school	2.26	.93	2.30	1.00	2.21	.79	.492	.000
Secondary school	3.18	.96	3.31	.95	2.94	.94	.010**	.028
University	2.59	1.26	3.08	1.07	1.69	1.06	.001***	.274

Significance levels: $p \leq .01$ very significant (**), $p \leq .001$ highly significant (***) (Bühl, 2019); effect size: $.01 \leq \omega^2 < .06$: small effect, $.06 \leq \omega^2 < .15$: medium effect, $\omega^2 \geq .15$: large effect (Albert & Koster, 2002)

coefficient of $|r| \geq .1$ is regarded as small/weak, of $|r| \geq .3$ as medium/moderate, and of $|r| \geq .5$ as large/strong correlation (Cohen, 1988). If $0 < r \leq 1$, the linear relationship is positive, if $-1 \leq r < 0$, the linear relationship is negative (Kuckartz et al., 2013).

RESULTS

Considering the results of the total sample size ($N = 202$) the pre-service primary school teachers gained little to moderate experimental experience in their past (RQ1.1). The highest but only moderate experimental experience was reported for the period of secondary school (see Table 1). Less experience was gained while studying at university and primary school (see Table 1). While the SciMaj pre-service teachers' experimental experience in primary school does not differ from the non-SciMaj group ($t(170.272) = .688$, $p = .492$), differences can be observed during the secondary school and university period. The SciMaj group gained significantly higher experimental experience than the non-SciMaj group both in secondary school ($t(200) = 2.594$, $p = .010$, $\omega^2 = .028$) and university ($t(200) = 8.799$, $p \leq .001$, $\omega^2 = .274$; see Table 1).

Moderate correlations (RQ1.2) were found between experimental experiences in primary school ($r(200) = .200$; $p = .004$), secondary school

($r(200) = .206$, $p = .003$) as well as while studying at university ($r(200) = .185$, $p = .009$) and the self-concept on planning & interpreting experiments (see Table 2). Moderate correlations were also found between the experimental experiences in primary school ($r(200) = .170$, $p = .016$), secondary school ($r(200) = .289$, $p \leq .001$), and while studying at university ($r(200) = .166$, $p = .018$) and the self-concept on conducting experiments. No significant correlations were observed between the experimental experience in primary or secondary school and the self-concept on planning experiment-based lessons. However, this self-concept and the experimental experience while studying at university were highly significantly correlated ($r(200) = .346$; $p \leq .001$). Further moderate correlations were identified between the experimental experience in primary school ($r(200) = .202$, $p = .004$), secondary school ($r(200) = .214$, $p = .002$) and while studying at university ($r(200) = .327$, $p \leq .001$) and the self-concept on teaching experiment-based lessons.

The pretest values of both the self-concept on planning & interpreting experiments just as on planning experiment-based lessons are of moderate height (see Table 3). The self-concepts on conducting experiments and on teaching experiment-based lessons are also only slightly higher.

Table 2. Correlations between experimental experience and self-concepts (pretest): Pearson correlation coefficient r (each in the top line) and p -value (each in the lower line) ($N = 202$)

	Self-concept on planning & interpreting experiments	Self-concept on conducting experiments	Self-concept on planning experiment-based lessons	Self-concept on teaching experiment-based lessons
Experimental experience in primary school	.200 .004**	.170 .016*	.093 .189	.202 .004**
Experimental experience in secondary school	.206 .003**	.289 .000***	.064 .364	.214 .002**
Experimental experience while studying at university	.185 .009**	.166 .018*	.346 .000***	.327 .000***

Significance levels: $p \leq .05$ significant (*), $p \leq .01$ very significant (**), $p \leq .001$ highly significant (***) (Bühl, 2019)

Table 3. Mean values (M) and standard deviations (SD) of experiment-related self-concepts and self-concepts on planning and teaching experiment-based lessons at both measurement times in comparison of EG (N = 158) and BG (N = 44). *p*-values and effect sizes η_p^2 given for inner- and intergroup comparison

Scale	Group	Ref. time	M	SD	<i>p</i>	η_p^2	<i>p</i> _{groups}	$\eta_{p, groups}^2$
Self-concept on planning & interpreting experiments	EG	pre	3.15	.63	.000***	.533	.000***	.204
		post	3.77	.56				
	BG	pre	3.24	.51	.334	.022		
		post	3.16	.57				
Self-concept on conducting experiments	EG	pre	3.83	.71	.000***	.537	.000***	.202
		post	4.50	.50				
	BG	pre	3.92	.70	.331	.022		
		post	3.82	.83				
Self-concept on planning experiment-based lessons	EG	pre	2.56	.84	.000***	.782	.000***	.228
		post	4.12	.64				
	BG	pre	3.30	.90	.008**	.153		
		post	3.57	.93				
Self-concept on teaching experiment-based lessons	EG	pre	3.62	1.08	.000***	.438	.000***	.089
		post	4.53	.55				
	BG	pre	3.80	.71	.065	.077		
		post	3.98	.87				

Significance levels: $p \leq .05$ significant (*), $p \leq .01$ very significant (**), $p \leq .001$ highly significant (***) (Bühl, 2019); effect size: $.01 \leq \eta_p^2 < .06$: small effect, $.06 \leq \eta_p^2 < .14$: medium effect, $\eta_p^2 \geq .14$: large effect (Cohen, 1988)

Therefore, floor and ceiling effects (Döring & Bortz, 2016) can be largely neglected. The independent samples *t*-test revealed no differences between the EG and BG at the pretest time in any scales except for the self-concept on planning experiment-based lessons ($t(200) = -5.07, p \leq .001, \omega^2 = .109$).

When comparing the effect of the intervention on EG and BG (RQ2.1) with a two-way repeated measures ANOVA, statistically significant interactions between time and group (EG/BG) were found in the self-concepts on planning & interpreting experiments ($F(1,200) = 51.118, p \leq .001, \eta_p^2 = .204$), conducting experiments ($F(1,200) = 50.709, p \leq .001, \eta_p^2 = .202$) and teaching experiment-based lessons ($F(1,200) = 19.650, p \leq .001, \eta_p^2 = .089$). Due to the pretest differences in the self-concept on planning experiment-based lessons reported above, an ANCOVA with the pretest values as covariate and the posttest values as the dependent variable was carried out. It revealed that this self-concept developed significantly differently in both groups ($F(1,199) = 58.789, p \leq .001, \eta_p^2 = .228$).

Dependent samples *t*-tests were conducted to situate the observed differences within both groups. The experiment-related self-concepts (planning & interpreting experiments: $t = -13.380, p \leq .001, \eta_p^2 = .533, n = 158$; conducting experiments: $t = -13.492, p \leq .001, \eta_p^2 = .537, n = 44$) as well as both science teaching-related self-concepts (planning experiment-based lessons: $t = -23.731, p \leq .001, \eta_p^2 = .782, n = 158$; teaching experiment-based lessons: $t = -11.071, p \leq .001, \eta_p^2 = .438, n = 158$) increase significantly from pre- to posttest in the EG.

Except for a slight increase in the self-concept on planning experiment-based lessons ($t = -2.791, p = .008, \eta_p^2 = .153, n = 44$), there are no significant pretest-posttest changes in the BG (see Table 3).

To analyze the effect of the course format on EG's self-concepts (RQ2.2), the effects of the type of participation (voluntary/mandatory) and major (SciMaj/non-SciMaj) were factored out as described above. Since there were no pretest differences in all constructs, three-way repeated measures ANOVAs were conducted (1. factor: time, 2. factor: course format, 3. factor: the factored-out variable, that means the type of participation or major respectively). This results in the need to report two statistical characteristics per scale and group. Therefore, these are only reported in Table 4 to ensure the reading flow.

The academic self-concept increased significantly in each subscale, regardless of the course format and the type of factored-out variable (Table 4). There are no significant interactions between course format and reference time in the self-concepts on conducting experiments and teaching experiment-based lessons. A slightly but significantly higher increase in the self-concepts on planning & interpreting experiments and planning experiment-based lessons is observed in the block group, albeit with small effect sizes (see Table 4). There are no significant interactions between time, course format, and type of participation (factored-out variable 1) or rather time, course format, and major (factored-out variable 2). Therefore, no relevant distortion of the results by the type of participation and the major is to be expected.

Despite the non-significant interactions between time, course format, and the factored-out variable ‘major’ (see results for RQ2.2), two-way repeated ANOVAs (in case of pretest equivalency) and ANCOVAs with the pretest values as covariate and the posttest values as the dependent variable (in case of non-existent pretest equivalency) were conducted to test the impact of major separated into the course formats (RQ2.3; see Table 5). The effect of the type of participation (voluntary/mandatory) could not be factored out, as no cases of non-SciMaj participants mandatorily attending the intervention are available due to examination regulations. Particularly striking are the pretest differences between both subgroups. Significantly higher pretest values in the SciMaj group compared to the non-SciMaj group were found in the self-concept on planning & interpreting experiments (weekly course format: $t(86) = 3.000, p = .004, \omega^2 = .083$), as well as in the self-concepts on conducting experiments (weekly: $t(86) = 3.645, p \leq .001, \omega^2 = .123$; block: $t(33.699) = 2.367, p = .024, \omega^2 = .064$), on planning experiment-based lessons (weekly: $t(86) = 4.650, p \leq .001, \omega^2 = .190$), and on teaching experiment-based lessons (weekly: $t(37.982) = 5.380, p \leq .001, \omega^2 = .241$; block: $t(68) = 3.350, p \leq .001, \omega^2 = .132$). The SciMaj and the non-SciMaj group showed a significant increase in all self-concepts regardless of the course format. Only in terms of planning experiment-based lessons, the changes in self-concept between pretest and posttest

differed for the two groups. The two-way repeated measures ANOVAs showed a significant interaction between time and major for the students attending the block format ($F(1,68) = 7.242, p = .009, \eta_p^2 = .096$). Nevertheless, both the SciMaj and the non-SciMaj group benefited from the course.

To investigate the relationship between the changes in self-concepts (posttest-pretest) and perceived experimental competencies related to the course (RQ3), bivariate correlations including all participants of the EG were calculated (Table 6). The more the participants experienced themselves as competent in planning & interpreting experiments, the more their self-concept on planning and interpreting experiments increased ($r(156) = .228; p = .004$). Another small but significant correlation was found between the perceived competence in conducting experiments and the increase in self-concept on conducting experiments ($r(156) = .215; p = .007$). Furthermore, a small but significant negative correlation between the perceived competence in planning & interpreting experiments and the self-concept on teaching experiment-based lessons can be observed ($r(156) = -.158; p = .047$).

Table 4. Mean values (M) and standard deviations (SD) of experiment-related self-concepts and self-concepts on planning and teaching experiment-based lessons at both measurement times in comparison of the group of weekly (N = 88) and block format (N = 70) participants. *p*-values and effect sizes η_p^2 given for inner- and intergroup comparison

Scale	Group	Ref. time	M	SD	<i>p</i>	η_p^2	<i>p</i> _{groups}	η_p^2 _{groups}
Self-concept on planning & interpreting experiments	weekly	pre	3.16	.66	(.000***)/	(.479)/[.452]	(.021*)/[.035*]	(.034)/[.029]
		post	3.70	.59	[.000***)]			
	block	pre	3.14	.60	(.000***)/	(.621)/[.609]		
		post	3.86	.50	[.000***)]			
Self-concept on conducting experiments	weekly	pre	3.80	.70	(.000***)/	(.541)/[.563]	(.760)/[.856]	(.001)/[.000]
		post	4.48	.50	[.000***)]			
	block	pre	3.86	.72	(.000***)/	(.564)/[.568]		
		post	4.53	.50	[.000***)]			
Self-concept on planning experiment-based lessons	weekly	pre	2.62	.86	(.000***)/	(.765)/[.776]	(.012*)/[.061]	(.040)/[.023]
		post	4.06	.67	[.000***)]			
	block	pre	2.49	.80	(.000***)/	(.830)/[.826]		
		post	4.20	.60	[.000***)]			
Self-concept on teaching experiment-based lessons	weekly	pre	3.63	1.08	(.000***)/	(.467)/[.554]	(.263)/[.964]	(.008)/[.000]
		post	4.51	.56	[.000***)]			
	block	pre	3.61	1.08	(.000***)/	(.533)/[.533]		
		post	4.55	.54	[.000***)]			

round brackets = ANOVA with further between-subject factor ‘type of participation’; square brackets = ANOVA with further between-subject factor ‘major’; significance levels: $p \leq .05$ significant (*), $p \leq .01$ very significant (**), $p \leq .001$ highly significant (***) (Bühl, 2019); effect size: $.01 \leq \eta_p^2 < .06$: small effect, $.06 \leq \eta_p^2 < .14$: medium effect, $\eta_p^2 \geq .14$: large effect (Cohen, 1988)

Table 5. Mean values (M) and standard deviations (SD) of experiment-related self-concepts and self-concepts on planning and teaching experiment-based lessons at both measurement times in comparison of weekly participants with a science major (SciMaj, N = 60) and with other majors (non-SciMaj, N = 28) and participants of the block format (SciMaj, N = 45/non-SciMaj, N = 25). *p*-values and effect sizes η_p^2 given for inner- and intergroup comparison

Scale	Group	Subgroup	Ref. time	M	SD	<i>p</i>	η_p^2	<i>p</i> _{groups}	$\eta_{p, groups}^2$
Self-concept on planning & interpreting experiments	weekly	SciMaj	pre	3.30	.61	.000***	.451	.152	.024
			post	3.82	.58				
	non-SciMaj	pre	2.87	.69	.000***	.542			
		post	3.43	.54					
	block	SciMaj	pre	3.20	.54	.000***	.570		
			post	3.84	.49				
non-SciMaj	pre	3.03	.69	.000***	.666				
	post	3.88	.52						
Self-concept on conducting experiments	weekly	SciMaj	pre	3.98	.67	.000***	.447	.863	.000
			post	4.54	.47				
	non-SciMaj	pre	3.43	.63	.000***	.737			
		post	4.33	.54					
	block	SciMaj	pre	4.03	.54	.000***	.522		
			post	4.57	.45				
non-SciMaj	pre	3.56	.91	.000***	.605				
	post	4.45	.58						
Self-concept on planning experiment-based lessons	weekly	SciMaj	pre	2.88	.76	.000***	.751	.862	.000
			post	4.17	.65				
	non-SciMaj	pre	2.06	.80	.000***	.823			
		post	3.85	.67					
	block	SciMaj	pre	2.61	.75	.000***	.850		
			post	4.13	.55				
non-SciMaj	pre	2.25	.86	.000***	.809				
	post	4.31	.67						
Self-concept on teaching experiment-based lessons	weekly	SciMaj	pre	4.04	.76	.000***	.387	.739	.001
			post	4.61	.50				
	non-SciMaj	pre	2.75	1.16	.000***	.634			
		post	4.31	.63					
	block	SciMaj	pre	3.91	.93	.000***	.312		
			post	4.53	.48				
non-SciMaj	pre	3.07	1.14	.000***	.684				
	post	4.59	.64						

Significance levels: $p \leq .05$ significant (*), $p \leq .01$ very significant (**), $p \leq .001$ highly significant (***) (Bühl, 2019); effect size: $.01 \leq \eta_p^2 < .06$: small effect, $.06 \leq \eta_p^2 < .14$: medium effect, $\eta_p^2 \geq .14$: large effect (Cohen, 1988)

DISCUSSION

Experimental Experiences and Correlations with Academic Self-Concepts

This study contributes to filling the research gap regarding the factors that influence primary teachers' experiment-related and science teaching-related self-concepts. Similar to previous reports on experimentation in school (Seidel et al., 2007), the pretest revealed that the participants gained relatively little experimental experience in primary school and moderate experience in secondary school. There have also been few opportunities for

experimentation whilst studying at university, especially for pre-service teachers without a science major. Given the following results, this fact is worrying: While no significant correlations were found between experimental experience in primary and secondary school and the self-concept on planning experiment-based lessons, there was a highly significant correlation between this self-concept and experimental experience while studying at university. These findings emphasize the relevance of experimental experience at university to develop professional competence, mentioned in the literature

Table 6. Correlations between perceived experimental competences (posttest) and change of self-concepts (pre-post-difference): Pearson correlation coefficient r (each in the top line) and p -value (each in the lower line) (N = 158)

	Self-concept on planning & interpreting experiments	Self-concept on conducting experiments	Self-concept on planning experiment-based lessons	Self-concept on teaching experiment-based lessons
Perceived competence in planning & interpreting experiments	.228 .004**	.058 .466	.020 .805	-.158 .047*
Perceived competence in conducting experiments	.103 .197	.215 .007**	.096 .228	-.118 .139

Significance levels: $p \leq .05$ significant (*), $p \leq .01$ very significant (**), $p \leq .001$ highly significant (***) (Bühl, 2019)

(Capps et al., 2012). The other moderate correlations between experimental experiences at school and university and experiment-related as well as teaching-related self-concepts indicate that experiences with experimentation dating back far in the past can have an influence on these self-concepts, too (Bong & Skaalvik, 2003; Dickhäuser, 2006; Shavelson et al., 1976).

As suggested by Franken et al. (2020), experiences in the dimensions of planning, conducting, and interpreting experiments for each of the three educational institutions could be correlated with the subscales of self-concepts and needs to be investigated in future studies to get a deeper insight into possible factors influencing these domain-specific self-concepts. As subjective interpretations of previous experiences also impact self-concepts (Wigfield & Eccles, 1992), it might be necessary to record not only the level of experimental experience but also individual perception.

The Dimensionality of the Experiment-Related Self-Concept

While Damerau (2012) and Franken et al. (2020) distinguish between three subscales (planning, conducting, and interpreting experiments) based on factor analytical findings, participants of this study differentiate between their abilities in planning and interpreting experiments and their abilities in conducting experiments according to our factor analytical results. As with Rautenstrauch & Busker (2020) and Atzert et al. (under review), the first subscale includes phases of theoretical considerations in experimentation, whilst the second comprises the phase of hands-on activities (Schreiber et al., 2009). For the context of school, it was shown that pupils differentiate between highly interrelated sub-competencies even if no corresponding grading occurs (e.g. Arens & Jansen, 2016). In the case of this study, the relatively limited experimental experience (see results on RQ1.1) and a lack of opportunities to independently plan and interpret experiments at

school and university (Seidel et al., 2007; Schulz, Wirtz, & Starauschek, 2012; Tesch & Duit, 2004) could have made it difficult for the pre-service teachers to differentiate between their abilities in planning and evaluating experiments. Subsequent studies should take a closer look at the structural stability (Möller & Trautwein, 2015) of experiment-related self-concept subscales. It needs to be investigated whether participating in several experiment-based interventions leads to a differentiation between the perception of abilities in planning, conducting, and interpreting experiments.

Looking at the level of the experiment-related self-concept, values in the subscale 'conducting' are higher than those in the subscale 'planning & interpreting' at both measurement times and in all (sub-)groups. This finding is consistent with the results of Buse (2017), Damerau (2012), and Rautenstrauch & Busker (2020) for secondary school students and master's pre-service biology, chemistry, and science and social studies teachers (Franken et al., 2020). Due to the previously discussed difficulties and deficits concerning planning and interpreting experiments of school students (Germann & Aram, 1996; Hofstein et al., 2005) and pre-/in-service teachers (Emereole, 2009; Fleischer et al., 2020; Hilfert-Rüppell et al., 2013), such a result was to be expected. Similarly, Schulz et al. (2012) postulate that both school and university students are most experienced in conducting experiments, while they lack experience in the phases of theoretical considerations. These differences in the perception of experiment-related abilities emphasize the relevance of courses like the one presented here, in which planning and interpreting experiments can also be practiced.

Effects of the Intervention and the Role of Perceived Experimental Competences

The present study revealed a positive impact of the intervention on pre-service primary teachers' experiment-related self-concepts and self-concepts on planning and teaching experiment-based lessons

(RQ2.1). Regarding experiment-related self-concepts, the results of Peschel (2018), Buse et al. (2018), and Damerau (2012) – showing that experiment-based educational settings can lead to a significant increase in university and school students' self-concepts on planning, conducting, and interpreting experiments – are confirmed. The finding that the BG's self-concepts do not change significantly over time is consistent with the results on RQ1.1: There seem to be only a few possibilities at university to gain experience in experimentation and/or teaching experiment-based lessons. Only the BG's self-concept on planning experiment-based lessons also increased significantly from pre- to posttest. Those experiences in planning experiment-based lessons could have been gained in advanced courses (bachelor program) or during the practical semester.

Several possible factors could have led to a significant increase in self-concepts: EG's responses to the open-ended questions presented in Beudels et al. (under review) point to *positive experiences* regarding experimentation and planning experiment-based lessons. The course activities offer many opportunities to test own abilities, make mistakes and reflect on them (see Educational Concept). Regarding RQ3, partial results are consistent with the thesis that self-concept is influenced by *competence experiences* (Bong & Skaalvik, 2003; Dickhäuser, 2006; Shavelson et al. 1976). The more the participants experienced themselves as competent in planning, conducting, and interpreting experiments during the intervention, the higher the experiment-related self-concept was stated. The small but significant negative correlation between the perceived competence in planning & interpreting experiments and the self-concept on teaching experiment-based lessons indicates that other factors also played a role in the positive development of the teaching-related self-concepts. The EG also lists individualized, positive *feedback* on performance and a *supportive learning environment* as reasons for recommending the course (Beudels et al., under review). These two factors have already been mentioned as determinants influencing the self-concept in literature (Bong & Skaalvik, 2003; Lüdtke et al., 2005; Möller & Trautwein, 2015).

As shown by Paulick et al. (2017) and Sorge et al. (2019), pre-service teachers' self-concepts are also formed with internal and external frames of reference (Marsh et al., 1988). Participants may have compared their performance at the end of the course to their previous performances in school or the beginning of the intervention (*temporal comparisons*; Marsh et al., 2015). In some open

answers (see Beudels et al., under review), students stated that planning an experiment for a science lesson was easier for them the second time than the first time. *Dimensional comparisons* (Wolff et al., 2018), e.g. comparing one's abilities in the competence areas of experimentation, may also have played a role. Successfully handling criterial requirements (*criterial reference norms*; Möller & Trautwein, 2015) could have led to the self-concepts being assessed more positively in the posttest. According to the results of Atzert et al. (2020), criterial reference norms, in particular, positively influence pupils' experiment-related self-concept. In our study, such criterial reference norms could have been learning objectives, verbal and written tasks for experimentation at the stations and planning experiment-based lessons, and competence expectations for primary school teachers (KMK, 2008/2019) discussed. An educational setting with a very heterogeneous learning group, where partner work predominates, and other teams can easily be observed in their actions, has the potential to provoke external, *social comparisons* (Marsh et al., 1988; 2015; Möller & Trautwein, 2015). The following elements were used to avoid adverse effects of social comparisons: Students did not receive grades for their planning products during the course; written feedback was given based on criteria; written feedback was anonymized so that the participants did not know which people were behind which planning product.

Impact of the Course Format

There is a lack of research on the effect of the course format on the development of teachers' self-concepts (RQ2.2). Under control of the variables 'type of participation' and 'major', it was shown that the examined self-concepts increased significantly with high effect size both when participating in the weekly and the block format. However, block course participants seem to have a slight advantage in terms of the reinforcement of self-concepts on planning & interpreting experiments and planning experiment-based lessons. These findings support the results of Schaal & Randler (2004) and Hilkenmeier & Sommer (2014), indicating that pre-service teachers' self-concepts develop more positively through block scheduling than through traditional scheduling as the students showed higher scores regarding perceived competence, a factor influencing the self-concept (Bong & Skaalvik, 2003; Dickhäuser, 2006; Shavelson et al., 1976). According to Dixon & O'Gorman (2019), block formats give students a faster sense of achievement as tasks are completed in a short time. Without being interrupted by other courses and thus having a continuous learning experience (Daniel,

2000), they can get an overview of (the development of) their skills and performance. Such 'concentrated' competence experiences and feedback combined with a positive learning environment, which is found more often with block formats than weekly formats (Samarawickrema & Cleary, 2021), could lead to a significantly higher increase in self-concepts. Weekly participants may also compare their achievements in the course with those in other courses they attend in parallel during the semester. This dimensional frame of reference (Wolff et al., 2018) does not apply to students of the block course. All these factors could have led to the block course participants overestimating their abilities. However, the small effect sizes indicate that the course format did not play a major role in changing self-concepts in this study.

Impact of the Major Field of Study

Due to the facts mentioned in the introduction, one main objective of the intervention was to positively develop the self-concepts of those pre-service teachers without a science major. With one exception (block format, self-concept on planning experiment-based lessons, medium effect), self-concepts of pre-service teachers with and without a science major developed equally positively. Despite the finding that SciMaj students' self-concepts were almost consistently significantly higher in the pretest than those of the non-SciMajs, both groups seem to have encountered factors that led to a positive development of the self-concepts (see discussion on RQ2.1). Concerning the social reference norm (Marsh et al., 1988; 2015), non-SciMaj participants could have either benefited from the attempt not to focus on social comparison processes or – using social comparisons – they might have realized that their skills in terms of experimentation and planning experiment-based lessons are equal to those of SciMajs. At the same time, all participants may have perceived gains in skills and performance through temporal comparisons (Möller & Trautwein, 2015). The application of a questionnaire similar to that of Atzert et al. (2020), gathering students' self-concepts influenced by different reference norms, is recommendable to empirically substantiate these assumptions. The findings on RQ2.3 are also important for designing university curricula insofar as the course concept seems to be suitable for a very heterogeneous audience, which can be an advantage concerning the limited capacity of lecturers.

Study Limitations and Implications for Future Studies

Regarding the sample of this study, a positive selection can be assumed, as many of the pre-service teachers attended the course voluntarily (see

Curricular Framework). The same applies to the choice of the course format: Students could choose the format they prefer. Thus, a participant who thinks that the block format better suits her/his learning style will likely choose this format and vice versa (Burton & Nesbit, 2008). Daniel (2000) suggested that this situation could be countered with a random assignment of the pre-service teachers to the (sub-) groups. However, a suitable curricular framework is necessary for this. As long as course participation is counted towards a study module in some degree programs and in others not, it will be difficult to justify a randomized allocation of participants (Burton & Nesbit, 2008).

Due to the existing study regulations, too, it was not possible to include the two factored-out variables in the calculations simultaneously. Integrating this in the study design requires students without a science major to participate mandatorily in the intervention. Since self-concept was found to be changeable through interventions (e.g. Damerau, 2012; Peschel, 2018), but is also described as relatively stable (Shavelson et al., 1976), future studies could investigate the long-term effects of the intervention.

Operationalization of the constructs with more and more heterogeneous items would be appropriate to capture the self-concepts in more detail. For example, the scale 'self-concept on planning an experiment-based lesson' items could be developed on the six sub-areas of planning primary science and social studies lessons, mentioned by Tänzer (2010). However, the questionnaire was also designed to record other course effects (see Beudels et al, under review). Therefore, the number of items per subscale was minimized to maintain test efficiency and for motivational reasons.

In this study, the subjective perception of abilities was assessed. Even though a medium, positive correlation between self-concepts and factual skills and competencies is to be expected (Marsh, 1992; Paulick et al., 2016), the self-concept alone provides limited information about actual competencies (Festner et al., 2018). As mentioned above, it is questionable whether *all* pre-service teachers can realistically assess their experimental competencies (Schreiber et al., 2016). Similar to Paulick et al. (2016) and Sorge et al. (2019) and as suggested by Festner et al. (2018), in addition to self-concepts, professional knowledge (e.g. syntactic knowledge concerning experimentation (Haslbeck, 2019); pedagogical content knowledge for teaching experiment-based lessons) might need to be investigated to ensure the validity of the scales. It might also be an option to evaluate individual experimentation skills and abilities for planning experimental lessons through a category-based

evaluation of video recordings – as in Tesch & Duit (2004) – or the analysis of planning products. However, a more complex study design and assessment strategy might limit the sample size (Schreiber et al., 2009).

CONCLUSIONS

This study contributes to closing the gap regarding research into pre-service primary teachers' experiment-related and science teaching-related self-concepts. The results illustrate the relevance of experimental experiences at university to develop prospective teachers' professional competence. Besides, this study could demonstrate that the course format influences the development of self-concepts. The impact of this variable is not only to be considered in future studies but also in designing study curricula or training programs for in-service teachers. The intervention has the potential to strengthen the self-concepts of pre-service teachers with very different experimental experiences and motivations equally. Especially the positive impact on self-concepts of pre-service primary teachers without a science major is desirable given the academic self-concept being a predictor of teachers' performance in primary science lessons.

FUNDING

This work was supported by the Ministry of Education and Research within the framework of the Gemeinsame Qualitätsoffensive Lehrerbildung [joint quality offensive for teacher training] of the Federal Government of Germany and the Länder [grant number: OJA1507].

REFERENCES

- Albert, R., & Koster, C. J. (2002). *Empirie in Linguistik und Sprachlehrforschung. Ein methodologisches Arbeitsbuch* [Empirical research in linguistics and language teaching. A methodological workbook]. Tübingen, Germany: Narr.
- Appleton, K. (2007). Elementary Science Teaching. In S. K. Abell & N. G. Lederman (Eds.), *Handbook of Research on Science Education* (pp. 493–535). Mahwah, NJ: Lawrence Erlbaum.
- Arens, A. K., & Jansen, M. (2016). Self-concepts in reading, writing, listening, and speaking: A multidimensional and hierarchical structure and its generalizability across native and foreign languages. *Journal of Educational Psychology, 108*(5), 646–664. <https://doi.org/10.1037/edu0000081>
- Atzert, R., John, R., Preisfeld, A., & Damerau, K. (2020). Der Einfluss kriterialer, sozialer und individueller Bezugsnormen auf das experimentbezogene Fähigkeitsselbstkonzept [The influence of the criterial, social, and temporal reference norm on students' experiment-related self-concept]. *Zeitschrift für Didaktik der Naturwissenschaften, 26*, 89–102. <https://doi.org/10.1007/s40573-020-00114-x>
- Atzert, R., Tilli, J., Preisfeld, A., & Damerau, K. (under review). Experimentation-related self-concept of German upper secondary school students. *RISTAL*.
- Backhaus, K., Erichson, B., Wulff, P., & Weiber, R. (2018). *Multivariate Analysemethoden* [Multivariate methods of analysis] (15th ed.). Berlin/Heidelberg, Germany: Springer Gabler. <https://doi.org/10.1007/978-3-662-56655-8>
- Bartlett, M. S. (1951). The effect of standardization on a chi square approximation in factor analysis. *Biometrika, 38*(3/4), 337–344. <https://doi.org/10.1093/biomet/38.3-4.337>
- Baumert, J., & Kunter, M. (2013). The COACTIV model of teachers' professional competence. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), *Cognitive activation in the mathematics classroom and professional competence of teachers. Mathematics teacher education, vol. 8* (pp. 25–48). Boston, MA: Springer. https://doi.org/10.1007/978-1-4614-5149-5_2
- Beudels, M., Schilling, Y., & Preisfeld, A. (under review). Mit Experimenten zu Wasserläufer & Co Kohärenz erleben - Potenziale eines interdisziplinären, experimentellen Kurses zur Professionalisierung angehender Sachunterrichtslehrkräfte [Experience coherence with experiments on water striders & Co - Potentials of an interdisciplinary, experimental course for professionalizing pre-service General Studies teachers]. DiMawe - Zeitschrift für Konzepte und Arbeitsmaterialien für Lehrer*innenbildung und Unterricht.
- Bong, M., & Skaalvik, E. M. (2003). Academic self-concept and self-efficacy: How different are they really? *Educational Psychology Review, 15*(1), 1–40. <https://doi.org/10.1023/A:1021302408382>
- Bosch, K. (2006). *Planning classroom management. A five-step process to creating a positive learning environment* (2nd ed.). Thousand Oaks, CA: Corwin Press. <https://doi.org/10.4135/9781483329451>
- Bühl, A. (2019). *Einführung in die moderne Datenanalyse ab SPSS 25* [Introduction to modern data analysis from SPSS 25] (16th ed.). Hallbergmoos, Germany: Pearson.
- Bühner, M. (2021). *Einführung in die Test- und Fragebogenkonstruktion* [Introduction to test and questionnaire construction] (4th ed.). München, Germany: Pearson.
- Burton, S., & Nesbit, P. L. (2008). Block or traditional? An analysis of student choice of teaching format. *Journal of Management & Organization, 14*(1), 4–19. <https://doi.org/10.1017/S1833367200003448>
- Buse, M. (2017). *Bilinguale englische experimentelle Lehr-Lernarrangements im Fach Biologie. Konzeption, Durchführung und Evaluation der kognitiven und affektiven Wirksamkeit* [Bilingual

- teaching in biology. Its cognitive and affective-motivational impact]. Dissertation, University of Wuppertal, Germany. Retrieved from: <http://elpub.bib.uni-wuppertal.de/edocs/dokumente/fbc/biologie/diss2017/buse/dc1722.pdf>.
- Buse, M., Damerau, K., & Preisfeld, A. (2018). A scientific out-of-school programme on neurobiology employing CLIL. Its impact on the cognitive acquisition and experimentation-related ability self-concepts. *International Journal of Environmental & Science Education*, 13(8), 647-660.
- Capps, D. K., Crawford, B. A., & Conostas, M. A. (2012). A review of empirical literature on inquiry professional development: alignment with best practices and a critique of the findings. *Journal of Science Teacher Education*, 23(3), 291-318. <https://doi.org/10.1007/s10972-012-9275-2>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York, NY: Erlbaum.
- Damerau, K. (2012). *Molekulare und Zell-Biologie im Schülerlabor* [Molecular and cell biology in an out-of-school laboratory]. Dissertation, University of Wuppertal, Germany. Retrieved from: <http://elpub.bib.uni-wuppertal.de/edocs/dokumente/fbc/biologie/diss2012/damerau/dc1231.pdf>.
- Daniel, E. L. (2000). A review of time-shortened courses across disciplines. *College Student Journal*, 34, 298-308.
- Dickhäuser, O. (2006). Fähigkeitsselbstkonzepte – Entstehung, Auswirkung, Förderung [Academic self-concepts – Formation, effect, enhancement]. *Zeitschrift für Pädagogische Psychologie*, 20(1/2), 5-8. <https://doi.org/10.1024/1010-0652.20.12.5>
- Dixon, L. & O’Gorman, V. (2019). ‘Block teaching’ – exploring lecturers’ perceptions of intensive modes of delivery in the context of undergraduate education. *Journal of Further and Higher Education*, 44(5), 583-595. <https://doi.org/10.1080/0309877X.2018.1564024>
- Döring, N., & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* [Research methods and evaluation in social sciences and humanities] (5th ed.). Berlin/Heidelberg, Germany: Springer. <https://doi.org/10.1007/978-3-642-41089-5>
- Emereole, H. U. (2009). Learners’ and teachers’ conceptual knowledge of science processes: The case of Botswana. *International Journal of Science and Mathematics Education*, 7(5), 1033-1056. <https://doi.org/10.1007/s10763-008-9137-8>
- Federal Statistical Office (2020). *Lehrkräfte nach Schularten und Beschäftigung. Schuljahr 2019/20. Stand: 20. Oktober 2020* [Teachers by type of school and employment. The academic year 2019/20. Status as of 20. October 2020]. Retrieved from: <https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bildung-Forschung-Kultur/Schulen/Tabellen/allgemeinbildende-beruflicheschulen-lehrkraefte.html>.
- Festner, D., Schaper, N., & Gröschner, A. (2018). Einschätzung der Unterrichtskompetenz und -qualität im Praxissemester [Assessment of teaching skills and quality in the practical semester]. In J. König, M. Rothland, N. Schaper N. (Eds.), *Learning to Practice, Learning to Reflect?* (pp. 163-194). Wiesbaden, Germany: Springer VS. https://doi.org/10.1007/978-3-658-19536-6_6
- Fleischer, T., Virtbauer, L., & Strahl, A. (2020). Experimente im NAWI-Unterricht. Kompetenzen angehender Lehrkräfte [Experiments in science lessons. Pre-service teachers’ competencies]. In S. Habig (Ed.), *Naturwissenschaftliche Kompetenzen in der Gesellschaft von morgen* (pp. 848-851). Duisburg/Essen, Germany: University of Duisburg-Essen.
- Franken, N. (2020). *Kognitive und affektiv-motivationale Kompetenzen von Lehramtsstudierenden der Fächer Biologie, Chemie und Sachunterricht im Kontext des Praxissemesters* [Cognitive and affective-motivational competencies of pre-service biology, chemistry and science and social studies teachers in the context of the practical semester]. Dissertation, University of Wuppertal, Germany. Retrieved from: <http://elpub.bib.uni-wuppertal.de/servlets/DerivateServlet/Derivate-10821/dc2024.pdf>.
- Franken, N., Damerau, K., & Preisfeld, A. (2020). „Experimentieren kann ich gut!“ – Experimentbezogene Fähigkeitsselbstkonzepte von Lehramtsstudierenden der Fächer Biologie, Chemie und Sachunterricht [“My experimental skills are excellent!” – Academic self-concepts of experimental skills of pre-service teachers with the school subjects biology, chemistry and elementary science]. *Zeitschrift für die Didaktik der Biologie (ZDB) – Biologie Lehren und Lernen*, 24, 48-66. <https://doi.org/10.4119/zdb-1733>
- Fromm, S. (2012). Faktoren- und Reliabilitätsanalyse [Factor and reliability analysis]. In S. Fromm (Ed.), *Datenanalyse mit SPSS für Fortgeschrittene 2: Multivariate Verfahren für Querschnittsdaten* (pp. 53-82). Wiesbaden, Germany: Springer VS. https://doi.org/10.1007/978-3-531-18794-5_3
- George, D., & Mallery, P. (2003). *SPSS for Windows Step by Step: A Simple Guide and Reference. 11.0 update* (4th ed.). Boston, MA: Allyn & Bacon.
- Germann, P. J., & Aram, R. J. (1996). Student performances on the science processes of recording data, analyzing data, drawing conclusions, and providing evidence. *Journal of Research in Science Teaching*, 33(7), 773-798. [https://doi.org/10.1002/\(SICI\)1098-2736\(199609\)33:7<773::AID-TEA5>3.0.CO;2-K](https://doi.org/10.1002/(SICI)1098-2736(199609)33:7<773::AID-TEA5>3.0.CO;2-K)
- Girwidz, R. (2020). Experimentieren im Physikunterricht [Experimentation in physics

- lessons]. In E. Kircher, R. Girdwidz & H.E. Fischer (Eds.), *Physikdidaktik Grundlagen* (4th ed.) (pp. 263–291). Berlin, Germany: Springer Spektrum. https://doi.org/10.1007/978-3-662-59490-2_7
- Guskey, T. R. (1988). Teacher efficacy, self-concept, and attitudes toward the implementation of instructional innovation. *Teaching and Teacher Education*, 4(1), 63–69. [https://doi.org/10.1016/0742-051X\(88\)90025-X](https://doi.org/10.1016/0742-051X(88)90025-X)
- Gyllenpalm, J., & Wickman, P.-O. (2011). “Experiments” and the inquiry emphasis conflation in science teacher education. *Science Teacher Education*, 95(5), 908–926. <https://doi.org/10.1002/sce.20446>
- Haslbeck, H. (2019). Die Variablenkontrollstrategie in der Grundschule [The control-of-variables strategy in primary school]. Dissertation, Technical University of Munich, Germany. Retrieved from: <https://mediatum.ub.tum.de/doc/1487608/1487608.pdf>
- Hilfert-Rüppell, D., Looß, M., Klingenberg, K., Eghtessad, A., Höner, K., Müller, R., Strahl, A., & Pietzner, V. (2013). Scientific reasoning of prospective science teachers in designing a biological experiment. *Lehrerbildung auf dem Prüfstand*, 6(2), 135–154.
- Hilkenmeier, J., & Sommer, S. (2014). Praxisnahe Fallarbeit – Block versus wöchentliches Seminar. Ein Vergleich zweier Veranstaltungsformate in der Lehrerinnen- und Lehrerbildung [Case-based learning – Block scheduling versus traditional scheduling. A comparison of two course concepts in teacher education]. *Beiträge zur Lehrerinnen- und Lehrerbildung*, 32(1), 88–100.
- Hofstein, A., Navon, O., Kipnis, M., & Mamlok-Naaman, R. (2005). Developing students' ability to ask more and better questions resulting from inquiry-type chemistry laboratories. *Journal of Research in Science Teaching*, 42(7), 791–806. <https://doi.org/10.1002/tea.20072>
- Janssen, J., & Laatz, W. (2017). *Statistische Datenanalyse mit SPSS* [Statistical data analysis with SPSS] (9th ed.). Berlin, Germany: Springer Gabler. <https://doi.org/10.1007/978-3-662-53477-9>
- Kaiser, H. F., & Rice, J. (1974). Little Jiffy – Mark IV. *Educational and Psychological Measurement*, 34(1), 111–117. <https://doi.org/10.1177/001316447403400115>
- Kirsch, A. (2020). Qualitätsstandards für die Unterrichtsplanung im Fach Sachunterricht [Quality standards for planning in the subject of General Studies]. *HLZ – Herausforderung Lehrer*innenbildung*, 3(1), 406–422. <https://doi.org/10.4119/hlz-2557>
- Kleckmann, T., Möller, K., & Jonen, A. (2006). Die Wirksamkeit von Fortbildungen und die Bedeutung tutorieller Unterstützung [Effectiveness of trainings and importance of tutorial support]. In R. Hinz & T. Pütz (Eds.), *Professionelles Handeln in der Grundschule. Entwicklungslinien und Forschungsbefunde* (pp. 121–128). Hohengehren, Germany: Schneider.
- Kuckartz, U., Rädiker, S., Ebert, T., & Schehl, J. (2013). *Statistik. Eine verständliche Einführung* [Statistics. An understandable introduction] (2nd ed.). Wiesbaden, Germany: Springer VS. <https://doi.org/10.1007/978-3-531-19890-3>
- Kurth, C., & Wodzinski, R. (2020). Schwierigkeiten beim eigenständigen Experimentieren am Beispiel Hebel [Difficulties in autonomous experimentation using the example of levers]. In S. Habig (Ed.), *Naturwissenschaftliche Kompetenzen in der Gesellschaft von morgen* (pp. 262–265). Duisburg/Essen, Germany: University of Duisburg-Essen.
- Leiner, D. (2006). SoSci Survey (Computer Software). Retrieved from: <https://www.sosicisurvey.de/>.
- Lüttke, O., Köller, O., Marsh, H. W., & Trautwein, U. (2005). Teacher frame of reference and the big-fish-little-pond effect. *Contemporary Educational Psychology*, 30(3), 263–285. <https://doi.org/10.1016/j.cedpsych.2004.10.002>
- Marsh, H. W. (1992). Content specificity of relations between academic achievement and academic self-concept. *Journal of Educational Psychology*, 84(1), 35–42. <https://doi.org/10.1037/0022-0663.84.1.35>
- Marsh, H. W., Byrne, B. M., & Shavelson, R. J. (1988). A multifaceted academic self-concept: Its hierarchical structure and its relation to academic achievement. *Journal of Educational Psychology*, 80(3), 366–380. <https://doi.org/10.1037/0022-0663.80.3.366>
- Marsh, H. W., Salah Abduljabbar, A., Parker, P. D., Abdelfattah, F., Nagengast, B., Möller, J., & Abu-Hilal, M. M. (2015). The internal/external frame of reference model of self-concept and achievement relations: Age-cohort and cross-cultural differences. *American Educational Research Journal*, 52(1), 168–202. <https://doi.org/10.3102/0002831214549453>
- Merkens, H. (2010). *Unterricht: Eine Einführung* [Teaching: An introduction]. Wiesbaden, Germany: VS Verlag.
- Minner, D. D., Levy, A. J., & Century, J. (2010). Inquiry-based science instruction – what is it and does it matter? Results from a research synthesis years 1984 to 2002. *Journal of Research in Science Teaching*, 47(4), 474–496. <https://doi.org/10.1002/tea.20347>
- Möller, J., & Trautwein, U. (2015). Selbstkonzept [Self-concept]. In E. Wild & J. Möller (Eds.), *Pädagogische Psychologie* (2nd ed.) (pp. 177–199). Heidelberg, Germany: Springer. https://doi.org/10.1007/978-3-642-41291-2_8
- Nadelson, L. S., Callahan, J., Pyke, P., Hay, A., Dance, M., & Pfiester, J. (2013). Teacher STEM perception and preparation: Inquiry-based STEM professional development for elementary teachers. *The Journal*

- of *Educational Research*, 106(2), 157-168.
<https://doi.org/10.1080/00220671.2012.667014>
- National Academies of Sciences, Engineering, and Medicine (2015). *Science teachers' learning. Enhancing opportunities, creating supportive contexts*. Washington, D.C.: National Academies Press. <https://doi.org/10.17226/21836>
- Nerdel, C. (2017). Grundlagen der Naturwissenschaftsdidaktik [Basics of (natural) sciences didactics]. Berlin, Germany: Springer Spektrum. <https://doi.org/10.1007/978-3-662-53158-7>
- Noormann, P. (2017). *Mehrstufige Eigenmarken – Eine empirische Analyse von Zielen, Erfolgsdeterminanten und Grenzen* [Multilevel own brands – An empirical analysis of goals, success factors and limits]. Wiesbaden, Germany: Springer Gabler. <https://doi.org/10.1007/978-3-658-16005-0>
- Paulick, I., Großschedl, J., Harms, U., & Möller, J. (2016). Preservice teachers' professional knowledge and its relation to academic self-concept. *Journal of Teacher Education*, 67(3), 173–182.
<https://doi.org/10.1177/0022487116639263>
- Paulick, I., Großschedl, J., Harms, U., & Möller, J. (2017). How teachers perceive their expertise: The role of dimensional and social comparisons. *Contemporary Educational Psychology*, 51, 114–122.
<https://doi.org/10.1016/j.cedpsych.2017.06.007>
- Peschel, M. (2018). SelfPro: Entwicklung von Professionsverständnissen und Selbstkonzepten angehender Lehrkräfte beim Offenen Experimentieren [SelfPro: Development of pre-service teachers' professional understanding and self-concepts in open experimentation]. In S. Miller et al. (Eds.), *Profession und Disziplin. Jahrbuch Grundschulforschung vol. 22* (pp. 191–196). Wiesbaden, Germany: Springer VS.
https://doi.org/10.1007/978-3-658-13502-7_21
- Porsch, R., & Wendt, H. (2016). Aus- und Fortbildung von Mathematik- und Sachunterrichtslehrkräften [Education and training of maths and science and social studies teachers]. In H. Wendt et al. (Eds.), *TIMSS 2015. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (pp. 189–204). Münster, Germany/New York, NY: Waxmann.
- Rasch, B., Friese, M., Hofmann, W., & Naumann, E. (2014a). Quantitative Methoden 1 [Quantitative methods 1] (4th ed.). Berlin/Heidelberg, Germany: Springer. <https://doi.org/10.1007/978-3-662-43524-3>
- Rasch, B., Friese, M., Hofmann, W., & Naumann, E. (2014b). Quantitative Methoden 2 [Quantitative methods 2] (4th ed.). Berlin/Heidelberg, Germany: Springer. <https://doi.org/10.1007/978-3-662-43548-9>
- Rautenstrauch, H., & Busker, M. (2020). Experimentieren im Naturwissenschafts- und Chemieunterricht [Experimentation in science and chemistry lessons]. In S. Habig (Ed.), *Naturwissenschaftliche Kompetenzen in der Gesellschaft von morgen* (pp. 345–348). Duisburg/Essen, Germany: University of Duisburg-Essen.
- Samarawickrema, G., & Cleary, K. (2021). Block mode study: Opportunities and challenges for a new generation of learners in an Australian university. *Student Success*, 12(1), 13–23.
<https://doi.org/10.5204/ssj.1579>
- Schaal, S., & Randler, S. (2004). Konzeption und Evaluation eines computergestützten kooperativen Blockseminars zur Systematik der Blütenpflanzen [Development and evaluation of a computer-supported cooperative compact course in botanical systematics]. *Zeitschrift für Hochschuldidaktik*, 2(6), 1–18. <https://doi.org/10.3217/zfhd02/03>
- Schreiber, N., Theyßen, H., & Schecker, H. (2009). Experimentelle Kompetenz messen?! [Evaluating experimental competencies]. *Physik und Didaktik in Schule und Hochschule*, 3(8), 92–101.
- Schreiber, N., Theyßen, H., & Dickmann, M. (2016). Wie genau beurteilen Schülerinnen und Schüler ihre eigenen experimentellen Fähigkeiten? Ein Ansatz zur praktikablen Diagnostik experimenteller Fähigkeiten im Unterrichtsalltag. [How accurate can students assess their own experimental skills? An approach to feasibly assess experimental skills in the classroom]. *Physik und Didaktik in Schule und Hochschule*, 1(15), 49–63.
- Schulz, A., Wirtz, M., & Starauschek, E. (2012). Das Experiment in den Naturwissenschaften [The experiment in (natural) sciences]. In W. Rieß, M. Wirtz, B. Barzel & A. Schulz (Eds.), *Experimentieren im mathematisch-naturwissenschaftlichen Unterricht. Schüler lernen wissenschaftlich denken und arbeiten* (pp. 15–38). Münster, Germany: Waxmann.
- Schwichow, M. G., Croker, S., Zimmerman, C., Höffler, T. N., & Härtig, H. (2016). Teaching the control-of-variables strategy: A meta-analysis. *Developmental Review*, 39(1), 37–63.
<http://dx.doi.org/10.1016/j.dr.2015.12.001>
- Seidel, T., Prenzel, M., Wittwer, J., & Schwindt, K. (2007). Unterricht in den Naturwissenschaften [Science classes]. In M. Prenzel et al. (Eds.), *PISA 2006. Die Ergebnisse der dritten internationalen Vergleichsstudie* (pp. 147–180). Münster: Waxmann.
- Shavelson, R. J., Hubner, J. J., & Stanton, G. C. (1976). Self-concept: Validation of construct interpretations. *Review of Educational Research*, 46(3), 407–441.
<https://doi.org/10.3102/00346543046003407>
- Sorge, S., Keller, M. M., Neumann, K., & Möller, J. (2019). Investigating the relationship between pre-service

- physics teachers' professional knowledge, self-concept, and interest. *Journal of Research in Science Teaching*, 56(7), 937–955.
<https://doi.org/10.1002/tea.21534>
- Standing Conference of the Ministers of Education and Cultural Affairs (KMK) (2004). *Standards für die Lehrerbildung: Bildungswissenschaften* [Teacher education standards: Educational sciences]. Retrieved from:
http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2004/2004_12_16-Standards-Lehrerbildung.pdf.
- Standing Conference of the Ministers of Education and Cultural Affairs (KMK) (2008/2019). *Ländergemeinsame inhaltliche Anforderungen für die Fachwissenschaften und Fachdidaktiken in der Lehrerbildung* [Common standards for specialist science and subject didactics in teacher education]. Retrieved from:
http://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2008/2008_10_16-Fachprofile-Lehrerbildung.pdf.
- Tänzer, S. (2010). Bedingungen und Voraussetzungen in der Lehrperson [Conditions and requirements in a teacher]. In S. Tänzer & R. Lauterbach (Eds.), *Sachunterricht begründet planen* (pp. 64–76). Bad Heilbrunn, Germany: Klinkhardt.
- Tesch, M., & Duit, R. (2004). Experimentieren im Physikunterricht – Ergebnisse einer Videostudie [Experimentation in physics lessons – Results of a video study]. *Zeitschrift für Didaktik der Naturwissenschaften*, 10, 51–69.
- Tolsdorf, Y., & Markic, S. (2018). Development and changes in student teachers' knowledge concerning diagnostic in chemistry teaching – A longitudinal case study. *EURASIA Journal of Mathematics, Science and Technology Education*, 14(12), em1613. <https://doi.org/10.29333/ejmste/94232>
- Wahl, D. (2013). *Lernumgebungen erfolgreich gestalten. Vom trägen Wissen zum kompetenten Handeln* [Successfully designing learning environments. From dull knowledge to competent action] (3rd ed.). Bad Heilbrunn, Germany: Klinkhardt.
- Wigfield, A., & Eccles, J. S. (1992). The development of achievement task values: A theoretical analysis. *Developmental Review*, 12(3), 265–310.
[https://doi.org/10.1016/0273-2297\(92\)90011-P](https://doi.org/10.1016/0273-2297(92)90011-P)
- Wolff, F., Helm, F., Zimmermann, F., Nagy, G., & Möller, J. (2018). On the effects of social, temporal, and dimensional comparisons on academic self-concept. *Journal of Educational Psychology*, 110(7), 1005–1025. <https://doi.org/10.1037/edu0000248>
- Yeung, A. S., Craven, R. G., & Kaur, G. (2014). Teachers' self-concept and valuing of learning: Relations with teaching approaches and beliefs about students. *Asia-Pacific Journal of Teacher Education*, 42(3), 305–320.
<https://doi.org/10.1080/1359866X.2014.905670>
- Yip, D. Y. (2001). Assessing and developing the concept of assumptions in science teachers. *Journal of Science Education and Technology*, 10(2), 173–179.
<https://doi.org/10.1023/A:1009473215126>

APPENDIX

Table A1. Operationalization of the constructs in the questionnaire. Scale names including the number of items, translations of the original items, item abbreviations, and internal consistency (Cronbach's α)

Scale	Number of items	Example of an original item	Item translation	Item abbreviation (pre/post)	Cronbach's α (pre/post)
Experimental experience	2	In meiner Grundschulzeit haben wir im Sachunterricht viel experimentiert.	When I was in primary school, we did a lot of experiments in science and social studies lessons.	TU24_01/- TU24_05/-	.886/-
	2	Experimente waren in meiner Schulzeit auf der weiterführenden Schule nicht oft Bestandteil der naturwissenschaftlichen Fächer.*	During secondary school, experiments were not frequently part of science lessons.*	TU24_02/- TU24_04/-	.838/-
	2	In meiner (bisherigen) Studienzzeit haben wir in den Kursen viel experimentiert.	During my (previous) time at the university we have done a lot of experiments in my courses.	TU24_03/- TU24_07/-	.919/-
Academic self-concept	6	Es fällt mir leicht, ein Experiment zur Aufklärung einer vorgegebenen Fragestellung zu entwickeln.	It is easy for me to develop an experiment to clarify a given question.	TU29_01/NT17_04 TU29_05/NT17_07 TU29_07/NT17_01 TU29_03/NT17_08 TU29_06/NT17_03 TU29_09/NT17_06	.822/.842
	3	Für die Durchführung von Experimenten habe ich kein Händchen.*	Conducting experiments doesn't come naturally to me.*	TU29_02/NT17_05 TU29_04/NT17_02 TU29_08/NT17_09	.808/.804
	3	Ich bin gut darin, eine Sachunterrichtsstunde, in der experimentiert werden soll, zu planen.	I am good at planning an experiment-based science and social studies lesson.	TU31_07/NT19_19 TU31_12/NT19_07 TU31_19/NT19_02	923/.908
	3	Ich traue mir zu, Experimente mit den Schülern im Sachunterricht durchzuführen.	I have the confidence to conduct experiments in science and social studies lessons with students.	TU30_03/NT18_05 TU30_07/NT18_02 TU30_15/NT18_14	.901/.867
Perceived experimental competence	6 (8)	Es ist mir schwer gefallen, Vermutungen zum Ausgang der Kurseexperimente aufzustellen.* ^o	It was difficult for me to make assumptions about the outcome of the course experiments.* ^o	-/NT13_03 ^o -/NT13_12 -/NT13_19 -/NT13_20 ^o -/NT13_06 -/NT13_10 -/NT13_14 -/NT13_18	-.816 [#]
	3 (4)	Der Umgang mit den zur Verfügung gestellten Geräten und Materialien für die Experimente ist mir leicht gefallen.	It was easy for me to use the devices and materials provided for the experiments.	-/NT13_02 -/NT13_07 ^o -/NT13_11 -/NT13_17	-.707 [#]

* = inversely formulated items that have been recoded; ^o = items that were excluded after factor analysis; [#] = Cronbach's α after removing the marked items

Table A2. Assignment of the items (construct: experimental experience; pretest) to the factors (rotated factor matrix). Loadings leading to the assignment of the variables to the factor/subscale are highlighted in grey

Item abbreviation	Factor 1	Factor 2	Factor 3
TU24_01	.892	-	-
TU24_05	.879	-	-
TU24_02	-	.832	-
TU24_04	-	.844	-
TU24_03	-	-	.919
TU24_07	-	-	.923

Table A3. Assignment of the items (construct: experiment-related self-concept; pretest) to the factors (rotated factor matrix). Loadings leading to the assignment of the variables to the factor/subscale are highlighted in grey

Item abbreviation	Factor 1	Factor 2
TU29_01	.511	.394
TU29_05	.723	-
TU29_07	.511	.394
TU29_03	.626	-
TU29_06	.624	-
TU29_09	.646	-
TU29_02	-	.728
TU29_04	.387	.644
TU29_08	-	.755

Table A4. Assignment of the items (construct: self-concepts on planning and teaching experiment-based lessons; pretest) to the factors (rotated factor matrix). Loadings leading to the assignment of the variables to the factor/subscale are highlighted in grey

Item abbreviation	Factor 1	Factor 2
TU31_07	.784	.331
TU31_12	.875	-
TU31_19	.865	-
TU30_03	-	.801
TU30_07	-	.901
TU30_15	.365	.746

Table A5. Assignment of the items (construct: perceived experimental competencies; posttest) to the factors (rotated factor matrix). Loadings leading to the assignment of the variables to the factor/subscale are highlighted in grey

Item abbreviation	Factor 1	Factor 2
NT13_03*	.420	-
NT13_12	.552	-
NT13_19	.520	-
NT13_20*	.377	.461 ⁺
NT13_06	.590	.353
NT13_10	.792	-
NT13_14	.602	.329
NT13_18	.561	.363
NT13_02	-	.516
NT13_07*	-	.419
NT13_11	-	.738
NT13_17	-	.699

* = items were excluded from further analysis as factor loadings were $\lambda \leq .50$ (Backhaus et al., 2018) or/and items had cross-loadings (marked with a plus sign) $\lambda \geq .40$ (Noorman, 2017)