Preparing Teacher Training Students for Evidence-Based Practice Promoting Students' Research Competencies in Research-Learning Projects

Franziska Böttcher-Oschmann[1]*[†], Jana Groß Ophoff[2†] and Felicitas Thiel[1]

[1] *Department of Education and Psychology, Freie Universität Berlin, Berlin, Germany,* [2] *Institute for Secondary Education and Subject Didactics, University College of Teacher Education Vorarlberg, Feldkirch, FK, Austria*

Research-learning projects (RLP) enable teacher training students to acquire competencies for evidence-based practice (EBP) in the context of their university studies. The aim of this longitudinal study is to develop, implement, and evaluate a RLP format to promote competencies for EBP in teacher training students. These competencies can be broken down into the categories of using research, which involves reflection on and use of evidence to solve problems in teaching practice, and establishing research, which involves investigating a research question independently by applying research methods. In a longitudinal study we evaluate the increase in competencies based on a self-assessment of competencies (indirect measurement) focusing on establishing research, and a competence test (direct measurement) focusing on using research. We also add a retrospective pre-assessment version (quasi-indirect measurement) to consider response shift and over- or underestimation in self-assessments. Our findings show that teacher training students can be prepared for EBP through RLP. Further development potential for the RLP format is being discussed.

Keywords: evidence-based practice, research-learning projects, teacher training students, longitudinal study, research competencies

INTRODUCTION

Teachers in practice encounter findings from (inter)national student performance studies and comparative assessments (e.g., Programme for International Student Assessment—PISA, Trends in International Mathematics and Science Study—TIMMS, or written comparison tests in Germany named "VERgleichsArbeiten"—VERA) on a daily basis in their work. Evaluations of teaching quality and evidence-based approaches to school development are also having a growing impact on professional teaching practice (Humpert et al., 2006; Mandinach and Gummer, 2016; Kippers et al., 2018). As a result, the ability to make use of scientific evidence is becoming increasingly crucial to teachers' practice in classroom. The need for evidence-based practice (EBP) among teachers has been the subject of international discussion for some time (e.g., Hargreaves, 1996, 1997; Hammersley, 1997, Hammersley, 2007).

The concept of EBP originated in medicine and was applied to the teaching context in the early 1990s (Hargreaves, 1996, 1997; Sackett et al., 1996; Hammersley, 1997, Hammersley, 2007; Moon et al., 2000). According to Davies (1999), "Evidence-based education means integrating individual teaching and learning expertise with the best available external evidence from systematic research" (p. 117). In Germany, the curricular prerequisites were created in the 2000s with when EBP was anchored in the *National Standards for Teacher Education* (Kultusministerkonferenz, 2004, 2014), which define competencies that teacher training students should have acquired by the end of their study programs. For universities, this implies that teacher training programs should be designed to prepare teacher training students for their role as evidence-based experts (Mandinach and Gummer, 2016) of their own professional teaching practice.

The majority of teacher training programs in Germany are divided into two phases at university: bachelor's and master's degree programs. In both degree programs, teacher training students study their future two subjects of teaching as well as educational science. After this phase, they become trainee teachers at school. At the end, after final exams, they can work as teachers in schools. In 2016, an internship semester, a 6-month period of practical training in schools, was introduced to Berlin universities in the master's degree programs. Thereby, EBP was anchored in the teacher training curriculum. This means that students in teacher training should already acquire EBP competencies during the course of their studies and hence, before they become trainee teachers at schools. While concepts of teaching and learning that can be used to support the development of these competencies have been tested and evaluated in Anglo-American contexts (Reeves and Honig, 2015; Green et al., 2016; Reeves and Chiang, 2017; van der Scheer et al., 2017; Kippers et al., 2018), only a few such studies have been conducted in Germany to date (Groß Ophoff et al., 2017; Thoren et al., 2020).

This article presents the concept for a course designed to teach teacher training students EBP competencies, as well as the findings from our evaluation study. We begin by discussing the competencies required for EBP. We then look in detail at several important aspects that should be considered in planning a course. The introduction concludes with the course concept.

Evidence-based practice competencies can be divided, according to Davies (1999), into the categories of *using research* and *establishing research*, whereas Borg (2010) divides EBP into the categories of *engagement with research* and *engagement in research*. *Using research*, or *engagement with research*, comprises reflection on and use of evidence to solve problems in teaching practice. It means *knowing* where and how to find systematic and comprehensible evidence (for instance, in educational databases), and how to use this evidence to answer questions. *Establishing research*, or *engagement in research*, is based on the aforementioned competencies and requires additional competencies in investigating a research question *independently* through the use of basic research methods. This includes, for example, the planning and preparation of a study, data collection, and analysis. In contrast to another form of

teacher research, action research, which requires going through various investigative cycles (Borg, 2010), *establishing research* or *engagement in research* is defined by the logic of the research process. In the following, we use the terms *using research* and *establishing research* proposed by Davies (1999; see **Figure 1**), but they stand equally for the terms used by Borg (2010).

Establishing research requires competencies in research, comprising (a) Skills in Reviewing the State of Research, (b) Methodological Skills, (c) Skills in Reflecting on Research Findings, (d) Communication Skills, and (e) Content Knowledge (Thiel and Böttcher, 2014; Böttcher and Thiel, 2018; Böttcher-Oschmann et al., 2019). Teachers should be able to

(a) frame questions, search for relevant literature, and evaluate its quality (Ghali et al., 2000; Valter and Akerlind, 2010; Green et al., 2016; Mandinach and Gummer, 2016);
(b) plan and prepare the research process as well as collect and analyze data (Holbrook and Devonshire, 2005; Valter and Akerlind, 2010; Reeves and Honig, 2015; Green et al., 2016; Mandinach and Gummer, 2016; van der Scheer et al., 2017; Kippers et al., 2018);
(c) interpret evidence and draw conclusions for teaching practice (Ghali et al., 2000; Holbrook and Devonshire, 2005; Valter and Akerlind, 2010; Reeves and Honig, 2015; Green et al., 2016; Mandinach and Gummer, 2016; Reeves and Chiang, 2017; Kippers et al., 2018);
(d) communicate about this evidence (Valter and Akerlind, 2010; Mandinach and Gummer, 2016).
(e) To do this, teachers need knowledge about research, for example, about how to develop a research design (Green et al., 2016).

At Freie Universität Berlin, we developed a learning environment to foster teacher training students' acquisition of these competencies: the research-learning projects (RLP). When designing the RLP learning environment to promote the development of competencies, we included features that foster students' cognitive processing and motivation, as well as social interaction among students. At the same time, we wanted the environment to be as authentic and realistic as possible. In the following, we will explain these assumptions in more detail.

To foster *cognitive processing* of competencies that are not directly observable, instructors can use methods from
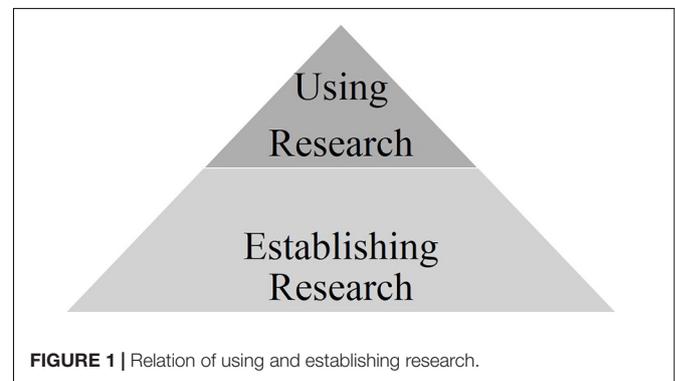

FIGURE 1 | Relation of using and establishing research.

the cognitive apprenticeship approach (Collins et al., 1989): cognitive modeling, coaching, and scaffolding. Through *cognitive modeling*, students can create a mental representation of a target strategy, such as strategies for literature review, enabling them to externalize instructors' internally available strategies, for instance, through prompting. When students adopt the instructors' strategy and begin using it on their own, instructors can provide *coaching* to support them in this process. This means that when students begin their own literature review, instructors' coaching can help students apply strategies they have already learned. When students reach the limits of their abilities, instructors can help by providing *scaffolding*, in which specific hints (prompts) are given. This could be, for example, hints such as the use of alternative search words or other databases for literature review.

Students' learning motivation to perform a particular learning activity may depend on both intrinsic and extrinsic *motivation* (Deci and Ryan, 2000). Factors that influence these types of learning motivation should therefore be integrated into the learning environment. Such factors may include needs for competence (e.g., feeling satisfied because of a perceive progress in a specific competence), autonomy (e.g., the ability to self-organize and self-regulate the learning process), and relatedness (e.g., relatedness to a group; Deci and Ryan, 2000).

At the level of *social interaction*, elements can be integrated that, on the one hand, support cognitive processing and, on the other hand, also foster sustained learning motivation by creating a feeling of social inclusion (Wecker and Fischer, 2014): instructors' supervision of students and group work. Instructors' supervision of students represents the interaction between members of a community of practice with different levels of expertise: Instructors are the experienced community members that promote students' competence development (Reinmann and Mandl, 2006). Moreover, this interaction allows students to engage with the thought patterns, attitudes, and normative standards of this community of practice (Reinmann and Mandl, 2006; Nückles and Wittwer, 2014). Group work enables the interaction between members of a community of practice with similar levels of expertise, such as teacher training students in RLP. Group members have the opportunity to learn and develop their own understanding through social activities that promote learning, such as explaining and thus externalizing their own thoughts (Wecker and Fischer, 2014) to, in this case, their fellow students.

Research-oriented teaching (Thiel and Böttcher, 2014; Böttcher and Thiel, 2018) formats, such as the RLP format, are suitable when designing learning opportunities around *authentic* and *realistic* situations: Students are enabled to *establishing research* by going through either individual phases of the research process or the entire process independently. They are empowered to meet the demands of this process by developing the necessary routines. Instructors support and advise students throughout these phases. In addition, students also regularly switch out of this active role and assume a receptive role. In these phases, the instructors impart knowledge about specific phases needed to go through the research process. An authentic and realistic learning environment is guaranteed by the fact that

students either participate actively in research, for example, in the framework of a research internship, or go through the entire research process themselves as part of a RLP.

The RLP course concept is based on the aforementioned assumptions about the design of learning environments and research-oriented teaching formats. The aim is that teacher training students are able to evaluate their own teaching quality on the basis of acquired competencies for EBP. To reduce complexity of the topic teaching quality, students could choose from the subtopics instructional quality, learning motivation or classroom management. The course is divided into three phases: (1) preparation, (2) fieldwork, and (3) presentation (see **Figure 2**). During the first phase, preparation, instructors impart knowledge about the phases of the research process: literature review, educational research methods[1] (planning of the research process, research design, data collection and measurement, etc.), and analytical procedures (e.g., calculation of descriptive statistics) as well as the presentation of findings. Students get together in groups, then choose one of the three subtopics and start to work on a literature review, formulate their research question, and develop an appropriate research design. In this phase, modeling promotes the development of Content Knowledge, Skills in Reviewing the State of Research, and Methodological Skills.

In the second phase, fieldwork, students prepare their studies and carry out data analysis and interpretation. They come to the university for two consultation sessions in which instructors provide coaching and scaffolding to support their process. In this phase, instructors focus on the development of Methodological Skills and Skills in Reflecting on Research Findings.

In the third phase, presentation, students create a scientific poster and present their findings. In this phase, the emphasis is on the development of Communication Skills.

The development of EBP competencies in RLP is fostered through the following aspects: The students go through the research process independently in an authentic and realistic field of action: here, the field of research. They receive advice (prompts) from instructors when they have reached the limits of their abilities. This enables students to experience the feeling of competence. The feeling of autonomy is strengthened through the large degree of freedom students have in the research process (to independently formulate their research question, develop a research design, and select instruments). Relatedness is encouraged by having students go through the research process in groups with support from instructors. In this social context, students have the opportunity to be socialized into the *scientific* community of practice: first, through interaction with instructors representing experienced researchers and, second, through the interaction with their fellow students.

With the introduction of an internship semester in 2016, EBP was anchored in the teacher training curriculum at Berlin universities. This was to implement the directives of the *National Standards for Teacher Education* (Kultusministerkonferenz, 2004, 2014) that teacher trainings students should be prepared for an EBP during the course of their studies. Hence, the RLP format was introduced at Berlin universities. At Freie Universität

---

[1] The focus was primarily on quantitative research methods.

FIGURE 2 | The three phases of a research-learning project (RLP).

Berlin, we then defined competencies needed for EBP as well as developed and implemented a RLP learning envornment to foster their acqisition. In this study, we want to evaluate if teacher training students can be enabled to acquire EBP competencies in the newly developed RLP format. The study design and findings are presented below.

MATERIALS AND METHODS

Study Design and Sampling

In a longitudinal study, data were collected at Freie Universität Berlin in the winter semester of 2016/17: at the beginning of the semester, in October 2016, and at the end of the semester, in February 2017. From September 2016 to February 2017 teacher training students completed their internship[2]. The goal of this internship is for master's teacher training students to gain their first teaching experience before becoming trainee teachers with greater responsibility in schools. During the winter semester, teacher training students attended university courses in parallel. These included the RLP format described above.

Ninety-seven teacher training students participated at the first point of assessment and 78 at the second point. These students were distributed among nine seminars with five instructors. For the longitudinal analysis, students were selected who had taken part in both waves of the survey ($n = 36$; see **Table 1**). During the internship semester, most RLP students were in the third semester of their master's studies ($M = 2.94$, SD $= 0.33$).

It should be noted that the sample of teacher training students is subject to a certain degree of pre-selection bias because students could choose to register for the RLP course on their own. A maximum of 15 students were distributed, first, on the basis of preferences and second, in case of overbooking, randomly among the remaining seminars. Hence, this is a non-randomized sample.

Ethics

The studies involving human participants were reviewed and approved by the *Ethikkommission der Freien Universität Berlin Fachbereich Erziehungswissenschaft und Psychologie* (the Ethics Committee of Freie Universität Berlin Department of Education and Psychology, own translation). Written informed consent from the participants was not required to participate in this study in accordance with the national legislation and the institutional requirements.

---

[2]This is the second practical experience that accompanies the course of study. The first internship at the beginning of the bachelor's degree program is a much shorter internship to get to know the teaching profession better.

TABLE 1 | Characteristics of participants in the longitudinal sample.

N	36
Gender	
Male	13
Female	22
Not reported	1
First/second teaching subject	
Languages	12/5
Social sciences	5/7
Natural sciences	14/14
Not reported	5
Semester	
Mean (Standard deviation)	2.94 (0.33)

Measures and Data Collection

To evaluate the competence development, we used a combination of self-assessment of competencies and competence testing. This combination was chosen to counteract the weaknesses of each survey method: on the one hand, the possibility of misjudgments or tendencies toward bias in self-assessments (Lucas and Baird, 2006; Chevalier et al., 2009), and on the other hand, the tendency of competence tests to focus solely on selected areas of knowledge due to time limitations (Cramer, 2010; Mertens and Gräsel, 2018). We used the two methods together, as recommended by Lucas and Baird (2006), to compensate for the specific weaknesses of each.

Self-Assessment of Competencies

Participants self-assessed their competencies with the instrument for assessing student research competencies (R-Comp; Böttcher and Thiel, 2016, 2018). The R-Comp consists of 32 items (with a five-point response scale ranging from "1 – strongly disagree" to "5 – strongly agree") on five scales: Skills in Reviewing the State of Research (four items; $\alpha = 0.87$; e.g., "I am able to systematically review the state of research regarding a specific topic."), Methodological Skills (eight items; $\alpha = 0.88$; e.g., "I am able to decide which data/sources/materials I need to address my research question."), Skills in Reflecting on Research Findings (six items; $\alpha = 0.92$; e.g., "I am able to critically reflect on methodological limitations of my own research findings."), Communication Skills (five items; $\alpha = 0.89$; e.g., "I am able to write a publication in accordance with the standards of my discipline."), and Content Knowledge [nine items; $\alpha = 0.88$; e.g., "I am informed about the main (current) theories in my discipline."]. The R-Comp thus measures the competencies that are necessary for the entire research process and thus focuses on *establishing research* (Davies, 1999). However, the R-Comp is a cross-disciplinary instrument (Böttcher and Thiel, 2016, 2018). To specifically address the research process in the field of educational research, the R-Comp included instructions asking students to answer questions specifically for their studies in education. This was intended to ensure that students self-assessed their research competencies in the specific area of educational research.

The self-assessment of competencies brings with it certain problems. These include overestimation and underestimation of competencies (Böttcher-Oschmann et al., 2019) and the phenomenon of response shift ("a change in the meaning of one's self-evaluation of a target construct", Schwartz and Sprangers, 1999, p. 1532; Schwartz and Sprangers, 2010; Piwowar and Thiel, 2014). Therefore, a retrospective pre-assessment version of R-Comp (Schwartz and Sprangers, 1999, 2010) was used additionally at the second point of assessment. In this version, the answers in R-Comp were reformulated asking students to indicate how they had assessed their skills and knowledge "before the RLP." On the one hand, this ensured that the same internal standards were used when answering the items at the second point of assessment (Schwartz and Sprangers, 2010), making it possible to prevent response shift from biasing the results. Furthermore, the retrospective pre-assessment version offered students the possibility to reflect on their own increase in competencies (Hill and Betz, 2005).

Competence Testing

We used the test instrument for assessing Educational Research Literacy (ERL; Groß Ophoff et al., 2014, 2017) to measure student competencies. This test was developed especially for teacher training students and we were therefore able to use it without making any changes. It should be noted that the test only measures *using research* (Davies, 1999; Groß Ophoff et al., 2017). Two test booklets were used, consisting of 18 items in the pretest version (Information Literacy: 7 items; Statistical Literacy: 7 items; Evidence-Based Reasoning: 4 items) and 17 items in the posttest version (Information Literacy: 7 items; Statistical Literacy: 5 items; Evidence-Based Reasoning: 5 items). These included items on the use of research strategies for Information Literacy and statistical/numerical tasks for Statistical Literacy, both mainly in multiple-choice formats. Items in the field of Evidence-Based Reasoning included, for example, two abstracts that had to be evaluated regarding the admissibility of several statements. Test items were selected from a large pool of 193 items that were standardized in a large-scale study ($N_i = 1360$, cf. Groß Ophoff et al., 2014). In the selection of items, care was taken to ensure a broad spectrum of competencies and sufficient discriminatory power of selected items ($M(r_{it}) = 0.31$).

Procedure

Self-assessment and competence testing were carried out before (t_1) and after (t_2) the RLP in a paper-and-pencil survey. At t_2, self-assessment was additionally used in the retrospective pre-assessment version (t_1^{retro}). Prior to both surveys, students were informed about the study's aims and the voluntary nature of the survey, and their anonymity was guaranteed. At the end of each survey, students provided some personal data (gender, first and second teaching subject[3], and number of semesters).

Analyses

Statistical Characteristics

R-Comp

For each version of the R-Comp (first point of assessment: $_{self}t_1$, second point of assessment: $_{self}t_2$, retrospective pre-assessment version: $_{self}t_1^{retro}$), five scale scores were calculated according to the five dimensions of competence in the RMRC-K model. The hierarchical structure of the empirical model (Böttcher and Thiel, 2018) was thus taken into account. Internal consistency for all R-Comp scales was evaluated using Cronbach's α.

Test instrument for assessing Educational Research Literacy

Scale scores were computed using the three-dimensional ERL model (*Information Literacy, Statistical Literacy*, and *Evidence-Based Reasoning*), which is viable for course evaluation (cf. Groß Ophoff et al., 2017). Person measures for each sub-dimension were determined on the basis of a dichotomous response format of the items ("1 = correctly solved" and "0 = not correctly solved")

---

[3] After the study had begun, students were given the option to opt out of answering the question about their teaching subjects. Some students had been worried that they could be identified by their unusual combination of teaching subjects.

using the WLE estimator (Warm, 1989). The person measures from the manifest data were estimated using a maximum likelihood function (Hartig and Kühnbach, 2006; Strobl, 2010). Item difficulties were fixed (external anchor design; see Wright and Douglas, 1996; Mittelhaëuser et al., 2011) in order to compare results from this study to those from the standardization study (Groß Ophoff et al., 2014).

Difference Values and Effect Sizes

In order to compare differences (Δ) between the points of assessment, the following values were calculated. For self-assessment: (a) $\Delta_{self} t_2 - t_1$ to determine *indirect* differences in competencies; (b) $\Delta_{self} t_2 - t_1^{retro}$ to determine the *quasi-indirect* adjusted differences in competencies via retrospective assessment without response shift; (c) $\Delta_{self} t_1 - t_1^{retro}$ to identify response shift, if difference values are not equal to zero, and misjudgments, whereby positive difference scores indicate overestimation and negative difference scores indicate underestimation (Schwartz and Sprangers, 2010). For the competence test, (d) $\Delta_{test} t_2 - t_1$ was calculated to determine the *direct* differences in competencies.

The calculation of effect sizes for the differences was carried out according to Lakens' recommendations inspecting d_{AV} (2013). The effects were interpreted in line with Cohen's (1992) benchmarks : ≥ 0.20 "small," ≥ 0.50 "medium," and ≥ 0.80 "large."

Longitudinal Analyses

Multivariate and multifactorial variance analyses (MANOVA) with repeated measurement were conducted in SPSS 25. Because of the small sample size, the hierarchical structure of the data was not taken into account. MANOVA with repeated measurement was performed according to the general linear model (GLM) with the factor time. If the Mauchly test was not applied, the Huynh field corrected degrees of freedom were used (see **Table 3**; Field, 2009). After this, univariate analyses were performed to identify major effects. Additionally, individual comparisons between factor levels were determined using the Bonferroni correction to account for the problem of multiple comparisons. The overall significance level was set at $p = 0.05$, while effect sizes η^2 were interpreted according to Cohen's (1988) benchmarks, with ≥ 0.01 "small," ≥ 0.06 "medium," and ≥ 0.14 "large."

For the analyses of self-assessment, MANOVA included the time points of assessment t_1, t_1^{retro}, t_2 for the five R-Comp scales. Although t_1^{retro} is a theoretical time point of assessment, it was included in MANOVA to investigate the occurrence of response shift. For the analyses of the competence test, MANOVA included test values for the three ERL scales at t_1 and t_2.

Missing Values

Non-response occurs when participants either do not take part at one of the time points or do not answer individual items (Sax et al., 2003). Only 37% of the students who took part at the first time point remained in the sample at the second time point. There was no systematic dropout by students' personal characteristics or competencies at the first time point.[4]

---

[4] However, the teacher training students in our study were also the target group in two other surveys that took place at the same time.

Unanswered items in the self-assessment of competencies did not occur more than twice per variable at any (theoretical) time point of assessment and were not imputed or replaced. For the analyses of the competence test, omissions were not imputed or replaced but treated as missing values (Groß Ophoff et al., 2017). With regard to personal information, 9.7% missing values occurred altogether, most of them for the first (10.2%) and second (22.0%) teaching subject.[5]

RESULTS

Self-Assessed Competencies

For *self-assessed research competencies*, a significant and large multivariate effect was found in MANOVA (**Table 2**). Moreover, results of univariate analysis indicated that this was caused by all five skills and knowledge dimensions, which showed large effects (**Table 3**). To find out how the individual effects are distributed over the individual time points of measurement, we consider the differences in mean values below. All mean values, standard deviations, differences in mean values, and effect sizes can be found in **Table 4**.

Indirect differences in competencies ($\Delta_{self} t_2 - t_1$): Significant increases occurred in Skills in *Reviewing the State of Research, Methodological Skills,* and *Communication Skills. Skills in Reviewing the State of Research* showed a large effect, *Communication Skills* a medium effect, *Methodological Skills* and

---

[5] These missing values probably occurred because students were given the option of not answering these items (see above).

TABLE 2 | Multivariate effects for self-assessed and tested research competencies.

	Pillai's Trace	F	df	Error df	p	η^2
Self-assessed research competencies	0.48	4.27	10	134	0.00	0.24
Tested research competencies	0.29	4.48	3	33	0.01	0.29

F = F-statistic; df = degrees of freedom; p = level of significance; η^2 = effect size.

TABLE 3 | Univariate effects for self-assessed and tested research competencies.

	df	SS	MS	F	p	η^2
Self-assessed research competencies						
Skills in Reviewing the State of Research	2	7.44	3.72	14.72	0.00	0.30
Methodological Skills	2	4.40	2.20	12.83	0.00	0.27
Skills in Reflecting on Research Findings	1.74[a]	2.69	1.52	7.80	0.00	0.18
Communication Skills	2	4.43	2.22	12.07	0.00	0.26
Content Knowledge	1.61[a]	4.91	3.10	6.55	0.01	0.16
Tested research competencies						
Information Literacy	1	3.96	3.96	4.60	0.04	0.12
Statistical Literacy	1	6.22	6.22	8.00	0.01	0.19
Evidence-Based Reasoning	1	0.00	0.00	0.00	0.95	0.00

df = degrees of freedom; SS = sum of squares; MS = mean square; F = F-statistic; p = level of significance; η^2 = effect size.
[a] Huynh-Feldt corrected df.

TABLE 4 | Mean values, standard deviations, differences in mean values, effect sizes, and Cronbach's alpha.

	Research-learning projects (RLP, $n = 36$)												
	$_{self}t_1$		$_{self}t_1^{retro}$		$_{self}t_2$		$_{self}t_1$-t_1^{retro}		$_{self}t_2$-t_1		$_{self}t_2$-t_1^{retro}		$\alpha_{t_1/t_1^{retro}/t_2}$
	M	SD	M	SD	M	SD	Δ	d_{av}	Δ	d_{av}	Δ	d_{av}	
Skills in Reviewing the State of Research	3.02	0.62	2.76	0.89	3.40	0.65	0.26	0.34	0.38**	0.80	0.64***	0.83	0.64/0.83/0.72
Methodological Skills	2.93	0.67	2.68	0.82	3.17	0.66	0.25	0.34	0.24*	0.48	0.49***	0.67	0.87/0.92/0.88
Skills in Reflecting on Research Findings	3.16	0.75	2.89	0.98	3.27	0.91	0.27	0.31	0.11	0.26	0.38***	0.40	0.85/0.93/0.91
Communication Skills	2.80	0.83	2.63	0.91	3.12	0.87	0.17	0.20	0.32**	0.67	0.49***	0.55	0.80/0.87/0.83
Content Knowledge	2.61	0.73	2.20	0.83	2.68	0.83	0.41	0.52	0.07	0.17	0.48***	0.58	0.82/0.85/0.91
	$_{test}t_1$				$_{test}t_2$				$_{test}t_2$-t_1				
	M	SD			M	SD			Δ	d_{av}			
Information Literacy	0.35	1.19			0.82	1.01			0.47*	0.79			
Statistical Literacy	0.02	1.12			0.61	1.11			0.59**	1.17			
Evidence-Based Reasoning	0.40	0.82			0.39	0.77			−0.01	−0.02			

Notes. $_{self}t_1$ = values of self-assessment at first time of assessment; $_{self}t_1^{retro}$ = values of self-assessment in retrospective pre-assessment version; $_{self}t_2$ = values of self-assessment at second time of assessment; $_{self}t_1$–t_1^{retro} = differences between values of self-assessment at first time of assessment and retrospective pre-assessment version; $_{self}t_2$–t_1 = differences between values of self-assessment at second time and first time of assessment; $_{self}t_2$–t_1^{retro} = differences between values of self-assessment at second time and retrospective pre-assessment version; α = Cronbach's alpha; M = means; SD = standard deviations; Δ = difference values; d_{av} = effect size for differences by Lakens (2013); $_{test}t_1$ = values of tested competencies at first time of assessment; $_{test}t_2$ = values of tested competencies at second time of assessment; $_{test}t_2$–t_1 = differences between values of tested competencies at second and first time of assessment; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Skills in Reflecting on Research Findings a small, positive effect, and *Content Knowledge* no effect.

Quasi-indirect differences in competencies ($\Delta_{self}t_2$–t_1^{retro}): Significant increases occurred in all skills and knowledge scales. Skills in *Reviewing the State of Research* showed a large effect, *Methodological Skills, Communication Skills*, and *Content Knowledge* a medium effect, and *Skills in Reflecting on Research Findings* a small effect.

Differences indicating response shift and overestimation of competencies ($\Delta_{self}t_2$–t_1^{retro}): Positive but no significant differences occurred. *Content Knowledge* showed a medium effect, and all skills scales a small effect. All identified effects indicate initial over-estimation.

Tested Competencies

For *tested research competencies*, a significant and large multivariate effect was found in MANOVA (**Table 2**). Moreover, results of univariate analysis indicated that this effect was caused by *Information Literacy*, with a medium effect, and *Statistical Literacy*, with a large effect (**Table 3**). These results are consistent with differences in mean values (direct differences in competencies; $\Delta_{test}t_2$-t_1): Significant increases occurred for *Information Literacy* and *Statistical Literacy*, both with large, positive effects, whereas *Evidence-Based Reasoning* showed no effect.

DISCUSSION

The aim of this study was to develop, implement, and evaluate a RLP format to promote EBP in teacher training students, enabling them to acquire competencies for EBP in the context of their university studies. These competencies can be broken down into the categories of *using research*, which involves reflection on and use of evidence to solve problems in teaching practice, and *establishing research*, which involves investigating a research question independently by applying research methods. We conducted a longitudinal study to evaluate the increase in competencies based on a self-assessment of competencies (indirect measurement) focusing on *establishing research*, and a competence test (direct measurement) focusing on *using research*. We also added retrospective pre-assessment version (quasi-indirect measurement) to consider response shift and overestimation or underestimation in self-assessments.

Overall, the results show increases in the competencies examined, albeit to varying degrees. The indirect measurement showed that teacher training students in the RLP perceived an increase in S*kills in Reviewing the State of Research, Methodological Skills*, and *Communication Skills*. Moreover, the quasi-indirect measurement indicated that students perceived an increase in all skill dimensions as well as in *Content Knowledge*. Thus, a difference between indirect and quasi-indirect increases in competencies became apparent, indicating the occurrence of response shift. The direct measurement showed that students improved in *Information Literacy* and *Statistical Literacy*. The results of the competence test correspond to the results of the self-assessment of competencies. Students showed increases in similar dimensions: *Information Literacy* and *Skills in Reviewing the State of Research* as well as *Statistical Literacy* and *Methodological Skills*. In general, it can be noted that students seemed to benefit from the RLP format in every aspect of competence except for *Skills in Reflecting on Research Findings* and *Evidence-Based Reasoning*.

In summary, the RLP provided teacher training students the opportunity to go through the entire research process themselves.

The similar increases found between the self-assessment and the competence test can be explained well: While the self-assessment with R-Comp deals with *establishing research*, the competence test deals with *using research* and thus one facet of *establishing research*. This indicates that in the RLP, students learned not only to apply evidence but also to generate evidence themselves in order to use it in teaching practice. The design of the seminar around the structure of research process with close supervision by instructors seems to have contributed to the acquisition of EBP competencies by students. Only in the areas of *Interpreting Evidence* and *Drawing Conclusions for Practice* did the students appear not to have benefited from the RLP. On the one hand, the RLP did not explicitly promote reasoning through targeted exercises. The interpretation of results and reflection on their implications were only addressed in consultations about students' concrete findings. Since people rate their competencies in a more differentiated way when they have increased experience (Bach, 2013; Mertens and Gräsel, 2018), the students may still not have been able to assess their competencies in this area accurately. On the other hand, reasoning is an extremely complex process and the corresponding test items are therefore highly challenging (Groß Ophoff et al., 2014, 2017). In the future, we will adapt the RLP format to take this into account. There will be specific exercises for reflection, argumentation, how to contextualize one's own findings in relation to current research, and how to draw conclusions and implications for one's own practice. The work of Fischer et al. (2014) on *scientific reasoning* provides important insights in this respect.

The combination of self-assessments of competencies and a competence test allowed us to gain a comprehensive picture of the increase in competence. Each method has its own advantages and disadvantages, and the two therefore complement each other well (Lucas and Baird, 2006). Although self-assessments may be inaccurate due to misjudgments or bias (Lucas and Baird, 2006; Chevalier et al., 2009), they may still affect people's actions (Bach, 2013). Moreover, the quasi-indirect measurement of competence may have an impact on self-efficacy beliefs (Hill and Betz, 2005). As self-assessed research competencies are highly correlated with self-efficacy beliefs (Mertens and Gräsel, 2018; Böttcher-Oschmann et al., 2019), it would be desirable if the self-evaluation of competence as well as the belief in having completed the research process successfully had positive effects on future EBP. Future studies could include a follow-up survey in their design to examine whether the effects translated into practice.

Through the additional use of the retrospective pre-assessment version of R-Comp, differences were identified in the increase in competence between indirect and quasi-indirect measurement. These differences indicate that a response shift (Schwartz and Sprangers, 1999, 2010; Sprangers and Schwartz, 1999) and misjudgments (Kruger and Dunning, 1999; Lucas and Baird, 2006) occurred in the longitudinal measurement of self-assessed competencies. First, the fact that the difference between the first measurement point and the retrospective estimate is not equal to zero indicates a response shift. This assumption is supported by the fact that the quasi-indirect, adjusted difference values differ from the indirect difference values: In indirect measurement, two dimensions—*Skills in Reflecting on Research Findings,* and *Content Knowledge*—did not change over time, whereas in the quasi-indirect adjusted measurement, an increase was found in all areas. It seems that a change occurred in the individual's internal standards of measurement had occurred in these dimensions, a recalibration response shift (Schwartz and Sprangers, 1999, 2010; Sprangers and Schwartz, 1999). Second, students overestimated their competencies prior to the RLP. The reason for the initial overestimation may be the students' low level of experience (Mertens and Gräsel, 2018) at the beginning of the semester. Hence, students might have had a vague idea about research competencies and therefore misjudged themselves, although some of them achieved good test scores. At the end of the semester, students may have been better able to evaluate themselves in a differentiated way.

Some limitations to our study need to be considered. On the one hand, the explanatory power is limited by the small sample size. This can be seen, for example, in the fact that large effects sometimes did not reach the level of significance. Although a larger sample was originally planned, the high dropout rate meant that only a small number of students could be reached at both time points. Students may have been oversaturated by the other surveys that took place at university at the same time (Sax et al., 2003).

The interdisciplinary self-assessment and the domain-specific competence test are only partially compatible, as the two instruments cover different levels of EBP. A competence test that also covers *establishing research* would be desirable, although developing such a test would be challenging. Nevertheless, the combination of self-assessment and competence testing should be maintained in future studies.

The phenomenon of response shift could not be sufficiently investigated in this study. With the retrospective pre-assessment version of R-Comp we were only able to detect a recalibration response shift. A reconceptualization response shift is also conceivable, however, since RLPs are designed with the explicit aim of imparting the skills and knowledge needed to complete the entire research process, but this cannot be answered here based on the available data. Measurement invariance testing would be necessary as a statistical approach (Meredith, 1993; Piwowar and Thiel, 2014), but invariance testing for response shift was not possible here due to the small sample size. Such testing would be useful to compare results from our design-based approach to results from a statistical approach (Schwartz and Sprangers, 1999; Piwowar and Thiel, 2014). It should also be noted that in retrospective measurement, other forms of bias may occur, such as social desirability, recall bias, and the implicit theory of change (Hill and Betz, 2005).

Further analyses are necessary to empirically confirm the results reported here. An implementation check should be carried out to take account of composition effects, teaching subject combinations, school type, or previous experience with

the research process. Moreover, due to the small sample size, the hierarchical structure of the data has not been considered. To increase internal validity, and in particular to verify the results against a general increase in competencies as a result of attending other courses, it would be advisable to include a control group that attended a seminar on research-based, practical approaches to teaching and learning. As one cannot fully eliminate environmental influences, studies should control for the influence of courses attended in parallel.

CONCLUSION

Our study provides evidence that teacher training students can be prepared for EBP through RLP. One strength of this study is its design: The longitudinal approach used while students were in the field shows good external validity. Another strength is the learning environment, which offers students the opportunity to acquire the necessary competencies in EBP in the course of their teacher training based on international examples. The results show that through RLP, teacher training students learn to apply methods of self-assessment and external assessment. If it would succeed better to strengthen the competencies *Interpreting Evidence* and *Drawing Conclusions for Practice*, then teacher training students should be able to develop and monitor the quality of their teaching practice by, first, reflecting on their own experiences and competencies, second, by drawing conclusions from these reflections, and third, by applying what they have learned to their professional practice. We recommend that research-oriented teaching formats, such as the RLP format, should be integrated in other German master's degree programs to fulfill the directives of the *National Standards for Teacher Education* (Kultusministerkonferenz, 2004, 2014) that teacher trainings students should be prepared for an EBP.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because data privacy must be guaranteed. Requests to access the datasets should be directed to FB-O, franziska.boettcher-oschmann@fu-berlin.de.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethikkommission der Freien Universität Berlin Fachbereich Erziehungswissenschaft und Psychologie (the Ethics Committee of Freie Universität Berlin Department of Education and Psychology, own translation). Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

REFERENCES

Bach, A. (2013). *Kompetenzentwicklung im Schulpraktikum: Ausmaß und Zeitliche Stabilität von Lerneffekten Hochschulischer Praxisphasen [Competence Development in school Internships: Extent and Temporal Stability of Learning Effects of Practical Phases in Higher Education, 1st Edn.]*, Münster: Waxmann.

Borg, S. (2010). Language teacher research engagement. *Lang. Teach.* 43, 391–429. doi: 10.1017/S0261444810000170

Böttcher, F., and Thiel, F. (2016). "Der Fragebogen zur Erfassung studentischer Forschungskompetenzen - Ein Instrument auf der Grundlage des RMRKW-Modells zur Evaluation von Formaten forschungsorientierter Lehre [The questionnaire to assess university students' research competences – An instrument based on theRMRC-K model to evaluate research-oriented teaching and learning arrangements]," in *Neues Handbuch Hochschullehre*, eds B. Berendt, A. Fleischmann, N. Schaper, B. Szczyrba, and J. Wildt (Berlin: DUZ Medienhaus), 57–74.

Böttcher, F., and Thiel, F. (2018). Evaluating research-oriented teaching: a new instrument to assess university students' research competences. *Higher Educ.* 75, 91–110. doi: 10.1007/s10734-017-0128-y

Böttcher-Oschmann, F. (2019). *Erfassung und Förderung Studentischer Forschungskompetenzen. Validierung eines Instrumentes zur Erfassung studentischer Forschungskompetenzen sowie Gestaltung und Evaluation einer forschungsorientierten Lernumgebung. [Assessment and Promotion of Students' Research Competences]*. Dissertation thesis, Freie Universität Berlin, Berlin.

Böttcher-Oschmann, F., Groß Ophoff, J., and Thiel, F. (2019). Validierung eines Fragenbogens zur Erfassung studentischer Forschungskompetenzen über Selbsteinschätzungen – Ein Instrument zur Evaluation forschungsorientierter Lehr-Lernarrangements [Validation of a questionnaire to assess university students' research competences via self-evaluation – An instrument for evaluating research-oriented teaching and learning arrangements]. *Unterrichtswissenschaft* 47, 1–27. doi: 10.1007/s42010-019-0 0053-8

Chevalier, A., Gibbons, S., Thorpe, A., Snell, M., and Hoskins, S. (2009). Students' academic self-perception. *Econ. Educ. Rev.* 28, 716–727. doi: 10.1016/j.econedurev.2009.06.007

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Erlbaum.

Cohen, J. (1992). A power primer. *Psychol. Bull.* 112, 155–159. doi: 10.1037/0033-2909.112.1.155

Collins, A., Brown, J. S., and Newman, S. E. (1989). "Cognitive apprenticeship: teaching the craft of reading, writing, and mathematics," in *Knowing, Learning, and Instruction: Essays in Honor of Robert Glaser*, ed. L. B. Resnick (Hillsdale, NJ: Lawrence Erlbaum Associates), 453–494. doi: 10.4324/9781315044408-14

Cramer, C. (2010). "Kompetenzerwartungen Lehramtsstudierender. Grenzen und Perspektiven selbsteingeschätzter Kompetenzen in der Lehrerbildungsforschung [Expectations of competence of teacher training students. Limits and perspectives of self-assessed competencies in teacher education research]," in *Bildungsstandards und Kompetenzmodelle: Beiträge zu einer aktuellen Diskussion über Schule, Lehrerbildung und Unterricht*, eds A.

Gehrmann, U. Hericks, and M. Lüders (Bad Heilbrunn: Klinkhardt Verlag), 85–97.

Davies, P. (1999). What is evidence-based education? *Br. J. Educ. Stud.* 47, 108–121. doi: 10.1111/1467-8527.00106

Deci, E. L., and Ryan, R. M. (2000). The" what" and" why" of goal pursuits: human needs and the self-determination of behavior. *Psychol. Inq.* 11, 227–268. doi: 10.1207/S15327965PLI1104_01

Field, A. (2009). *Discovering Statistics Using SPSS: (and sex and drugs and rock 'n' roll)*. Los Angeles, CA: SAGE.

Fischer, F., Kollar, I., Ufer, S., Sodian, B., Hussmann, H., Pekrun, R., et al. (2014). Scientific reasoning and argumentation: advancing an interdisciplinary research agenda in education. *Higher Educ. Res. Dev.* 2, 28–45.

Ghali, W. A., Saitz, R., Eskew, A. H., Gupta, M., Quan, H., and Hershman, W. Y. (2000). Successful teaching in evidence-based medicine. *Med. Educ.* 34, 18–22. doi: 10.1046/j.1365-2923.2000.00402.x

Green, J. L., Schmitt-Wilson, S., Versland, T. M., Kelting-Gibson, L., and Nollmeyer, G. E. (2016). Teachers and data literacy: a blueprint for professional development to foster data driven decision making. *Jcepd* 3, 14–32.

Groß Ophoff, J., Schladitz, S., Lohrmann, K., and Wirtz, M. (2014). "Evidenzorientierung in Bildungswissenschaftlichen Studiengängen [Evidence Orientation in Educational Science Programmes]," in *Empirische Bildungsforschung und Evidenzbasierte Reformen im Bildungswesen*, eds K. Drossel, R. Strietholt, and W. Bos (Münster: Waxmann), 251–275.

Groß Ophoff, J., Wolf, R., Schladitz, S., and Wirtz, M. (2017). Assessment of educational research literacy in higher education: construct validation of the factorial structure of an assessment instrument comparing different treatments of omitted responses. *J. Educ. Res. Online* 9, 37–68.

Hammersley, M. (1997). Educational research and teaching: a response to David Hargreaves' TTA lecture. *Br. Educ. Res. J.* 23, 141–161.

Hammersley, M. (ed.) (2007). *Educational Research and Evidence-Based Practice*. Los Angeles, CA: SAGE.

Hargreaves, D. H. (1996). *Teaching as a Research-Based Profession: Possibilities and Prospects*. London: Teacher Training Agency Annual Lecture, 1996.

Hargreaves, D. H. (1997). In defence of research for evidence-based teaching: a rejoinder to Martyn Hammersley. *Br. Educ. Res. J.* 23, 405–419. doi: 10.1080/0141192970230402

Hartig, J., and Kühnbach, O. (2006). "Schätzung von Veränderung mit „plausible values "in mehrdimensionalen Rasch-Modellen [Estimating change using the plausible value technique within multidimensional Rasch-models]," in *Veränderungsmessung und Längsschnittstudien in der Empirischen Erziehungswissenschaft*, eds A. Ittel and H. Merkens (Wiesbaden: VS Verlag für Sozialwissenschaften/GWV Fachverlage GmbH Wiesbaden), 27–44. doi: 10.1007/978-3-531-90502-0_3

Hill, L. G., and Betz, D. L. (2005). Revisiting the retrospective pretest. *Am. J. Eval.* 26, 501–517. doi: 10.1177/1098214005281356

Holbrook, N. J., and Devonshire, E. (2005). Simulating scientific thinking online: an example of research-led teaching. *Higher Educ. Res. Dev.* 24, 201–213. doi: 10.1080/07294360500153844

Humpert, W., Hauser, B., and Nagl, W. (2006). Was (zukünftige) Lehrpersonen über wissenschaftliche Methoden und Statistik wissen sollen und wollen [What (future) teachers should and want to know about scientific methods and statistics]. *Beiträge zur Lehrerinnen-und Lehrerbildung* 24, 231–244.

Kippers, W. B., Poortman, C. L., Schildkamp, K., and Visscher, A. J. (2018). Data literacy: what do educators learn and struggle with during a data use intervention? *Stud. Educ. Eval.* 56, 21–31. doi: 10.1016/j.stueduc.2017.11.001

Kruger, J., and Dunning, D. (1999). Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *J. Personal. Soc. Psychol.* 77, 1121–1134. doi: 10.1037/0022-3514.77.6.1121

Kultusministerkonferenz (2004). *Standards für die Lehrerbildung: Bildungswissenschaften. Beschluss der Kultusministerkonferenz vom 16.12.2004 [Standards for Teacher Training in the Educational Sciences: Resolution of the Standing Conference of the Ministers of Education and Cultural Affairs of the Länder in the Federal Republic of Germany of 16/12/2004]*. Bonn: Kultusministerkonferenz.

Kultusministerkonferenz (2014). *Standards für die Lehrerbildung: Bildungswissenschaften. Beschluss der Kultusministerkonferenz vom 16.12.2004 i. d. F. vom 12.06.2014 [Standards for Teacher Training in the Educational*

Sciences: Resolution of the Standing Conference of the Ministers of Education and Cultural Affairs of the Länder in the Federal Republic of Germany of 562 12/06/2014]. Bonn: Kultusministerkonferenz.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Front. Psychol.* 4:863. doi: 10.3389/fpsyg.2013.00863

Lucas, R. E., and Baird, B. M. (2006). "Global self-assessment," in *Handbook of Multimethod Measurement in Psychology*, eds M. Eid and E. Diener (Washington: American Psychological Association). doi: 10.1037/11383-003

Mandinach, E. B., and Gummer, E. S. (2016). What does it mean for teachers to be data literate: laying out the skills, knowledge, and dispositions. *Teach. Teach. Educ.* 60, 366–376. doi: 10.1016/j.tate.2016.07.011

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika* 58, 525–543. doi: 10.1007/BF02294825

Mertens, S., and Gräsel, C. (2018). Entwicklungsbereiche bildungswissenschaftlicher Kompetenzen von Lehramtsstudierenden im Praxissemester [The development of educational competencies during long-term internships in teacher education]. *Z Erziehungswiss* 21, 1109–1133. doi: 10.1007/s11618-018-0825-z

Mittelhaëuser, M.-A., Béguin, A. A., and Sijtsma, K. (2011). *Comparing the Effectiveness of Different Linking Designs: The Internal Anchor Versus the External Anchor and Pre-Test Data*. Arnheim: Cito.

Moon, B., Butcher, J., and Bird, E. (eds) (2000). *Leading Professional Development in Education*. London: Routledge Falmer.

Nückles, M., and Wittwer, J. (2014). "Lernen und Wissenserwerb [Learning and knowledge acquisition]," in *Pädagogische Psychologie: Mit Online-Materialien*, eds T. Seidel and A. Krapp (Weinheim: Julius Beltz), 225–252.

Piwowar, V., and Thiel, F. (2014). Evaluating response shift in training evaluation: comparing the retrospective pretest with an adapted measurement invariance approach in a classroom management training program. *Eval. Rev.* 38, 420–448. doi: 10.1177/0193841X14546932

Reeves, T. D., and Chiang, J.-L. (2017). Building pre-service teacher capacity to use external assessment data: an intervention study. *Teach. Educ.* 52, 155–172. doi: 10.1080/08878730.2016.1273420

Reeves, T. D., and Honig, S. L. (2015). A classroom data literacy intervention for pre-service teachers. *Teach. Teach. Educ.* 50, 90–101. doi: 10.1016/j.tate.2015.05.007

Reinmann, G., and Mandl, H. (2006). "Unterrichten und Lernumgebungen gestalten [Teaching and designing learning environments]," in *Pädagogische Psychologie: Ein Lehrbuch*, eds A. Krapp and B. Weidenmann (Weinheim: Beltz PVU), 613–658.

Sackett, D. L., Rosenberg, W. M. C., Gray, J. A. M., Haynes, R. B., and Richardson, W. S. (1996). Evidence based medicine: what it is and what it isn't. *BMJ* 312, 71–72. doi: 10.1136/bmj.312.7023.71

Sax, L. J., Gilmartin, S. K., and Bryant, A. N. (2003). Assessing response rates and nonresponse bias in web and paper surveys. *Res. Higher Educ.* 44, 409–432. doi: 10.1023/A:1024232915870

Schwartz, C. E., and Sprangers, M. A. G. (1999). Methodological approaches for assessing response shift in longitudinal health-related quality-of-life research. *Soc. Sci. Med.* 48, 1531–1548. doi: 10.1016/S0277-9536(99)00047-7

Schwartz, C. E., and Sprangers, M. A. G. (2010). Guidelines for improving the stringency of response shift research using the thentest. *Qual. Life Res.* 19, 455–464. doi: 10.1007/s11136-010-9585-9

Sprangers, M. A. G., and Schwartz, C. E. (1999). Integrating response shift into health-related quality of life research: a theoretical model. *Soc. Sci. Med.* 48, 1507–1515. doi: 10.1016/S0277-9536(99)00045-3

Strobl, C. (2010). *Das Rasch-Modell: Eine Verständliche Einführung für Studium und Praxis [The Rasch-Model: A Comprehensible Introduction for Study and Practice*, 1st Edn. München: Hampp.

Thiel, F., and Böttcher, F. (2014). "Modellierung fächerübergreifender Forschungskompetenzen - Das RMKR-W-Modell als Grundlage der Planung und Evaluation von Formaten forschungsorientierter Lehre [Modeling cross-disciplinary research competences - The RMRC-K model as a basis for planning and evaluating formats of research-oriented teaching]," in *Neues Handbuch Hochschullehre*, eds B. Berendt, A. Fleischmann, N. Schaper, B. Szczyrba, and J. Wildt (Berlin: Raabe), 109–124.

Thoren, K., Wißmann, J., Harks, M., Wenger, M., Kinder, A., and Hannover, B. (2020). "Förderung von Datennutzungskompetenzen in der Lehramtsausbildung: Konzeption und Evaluation dreier Seminare [Promotion of data literacy skills in teacher training: Design and evaluation of three seminars]: Konzeption und Evaluation dreier Seminare," in *Evidenzbasierung in der Lehrkräftebildung*, eds I. Gogolin, B. Hannover, and A. Scheunpflug (Wiesbaden: Springer VS), 39–71. doi: 10.1007/978-3-658-2 2460-8_3

Valter, K., and Akerlind, G. (2010). Introducing students to ways of thinking and acting like a researcher: a case study of research-led education in the sciences. *Int. J. Teach. Learn. Higher Educ.* 22, 89–97.

van der Scheer, E. A., Glas, C. A. W., and Visscher, A. J. (2017). Changes in teachers' instructional skills during an intensive data-based decision making intervention. *Teach. Teach. Educ.* 65, 171–182. doi: 10.1016/j.tate.2017.02.018

Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika* 54, 427–450. doi: 10.1007/BF02294627

Wecker, C., and Fischer, F. (2014). "Lernen in gruppen [Learning in groups]," in *Pädagogische Psychologie: Mit Online-Materialien*, eds T. Seidel and A. Krapp (Weinheim: Julius Beltz), 277–296.

Wright, B. D., and Douglas, G. A. (1996). Estimating Rasch (person, ability, theta) measures with known dichotomous item difficulties: anchored maximum likelihood estimation (AMLE). *Rasch Meas. Trans.* 10:499.