

1 Introduction

Across the field, there is a growing awareness that knowledge about teaching and schools, however extensive, is not sufficient to prepare capable teachers. Nor is it enough to hope that the necessary routines and practical skills will be acquired in internships and practica, especially if the aim is to prepare teachers to enact more ambitious and equitable forms of instruction.

Urban Fraefel and Pam Grossman (2024, p. 139)

Teachers have a large impact on their learners'¹ achievement (Hattie, 2023, p. 216)². However, the potential positive impact that teachers could have on their learners' achievement is not fully realized, as a substantial share of learners fails to achieve the desired outcomes. This discrepancy is supported by findings of large-scale studies, such as PISA³ and the IQB⁴ *Trends in Student Achievement*, that show mixed results for learner achievement in Germany (OECD, 2023, p. 183; Stanat et al., 2023). This raises questions about the effectiveness of teacher education, which is supposed to prepare teachers to maximize their impact on learners' achievement. This concern is corroborated by findings from Hattie, who states that "[teachers']⁵ initial teacher training programs have little impact on how well teachers influence their students' achievement" (Hattie, 2023, p. 243), quantifying the impact with a small effect size of $d = 0.10$ (Hattie, 2023, p. 229). Thus, these findings point to potential areas for improvement in the current approach to teacher education in Germany.

A central challenge in teacher education is the gap between theory and practice. Often, a lack of action-orientation and a perceived disconnect between the acquired theoretical, academic knowledge at university and supposed actual practice in the classroom are named as reasons for this gap (Rothland, 2020, p. 133). A commonly proposed solution is to increase the share of practical phases, allowing future teachers to apply their knowledge in school internships

¹ In this thesis, the term *learner* is usually used when referring to K-12 education, i.e. primary and secondary education, while *students* refers to higher education, i.e., university. Exceptions are, of course, direct quotations.

² In this thesis, the APA guidelines are generally followed for citations. However, in-text citations will also include the page number for indirect citations, if the publication is not addressed as a whole. This is meant to increase transparency and traceability of the scientific process.

³ PISA is the abbreviation of *Programme for International Student Assessment*.

⁴ IQB is the abbreviation of Germany's *Institut zur Qualitätssicherung im Bildungswesen* (Institute for Educational Quality Improvement).

⁵ For reasons of legibility, parentheses that indicate grammatical adjustment in direct quotations always include the full word.

(Rothland, 2020, p. 133). However, there are indications that the actual application of knowledge during such authentic practical phases is quite challenging for students (Zeichner, 2010, p. 91). Possibly, the complexity of the classroom might be a reason for this finding.

One approach to address the gap between theory and practice is through an *approximation of practice* (Grossman, Compton, et al., 2009, p. 2076). In such an approach, pre-service teachers (PSTs), i.e., teacher students at university, are allowed to perform in an authentic, but reduced-complexity situation, which allows them to practice the application of knowledge without necessarily interacting with real learners (Grossman, Compton, et al., 2009, p. 2076). However, PSTs are rarely given the opportunity to approximate practice in terms of action-oriented sequences during their university-based studies (Grossman, Compton, et al., 2009, pp. 2094–2095). In medical education, such opportunities are provided, for instance, in the form of role-play-based simulations (RobS⁶), in which trained actors take on the role of standardized patients (Schnabel, 2018). In teacher education, RobS would allow future teachers to take on the role of the teacher and engage in an authentic situation with an actor, portraying, for example, a standardized learner (SL) or parent (Dotger et al., 2010).

RobS can also be used as a performance-oriented format of summative assessment. In medical education, simulations are used for the assessment of clinical skills before (future) doctors engage with real patients in high-stakes situations and potentially severe consequences (Sopka et al., 2018). Considering the relevance that assessment has for students and their learning behavior, the implementation of RobS as performance-oriented assessment formats might lead students being better prepared for actual practice in the classroom, as “assessment drives learning” (Sopka et al., 2018; see also Gulikers, Bastiaens, et al., 2008). In teacher education, however, RobS are rarely found as an assessment format in teacher education programs or addressed in research.

If RobS were used as an assessment format in teacher education, a focus should also be on subject-specific practices, i.e. feedback on writing in the context of English as a Foreign Language (EFL). While teachers need to acquire a diverse set of competences for their profession, they are still primarily responsible for teaching their subjects, such as EFL, and mastering the required practices. In the EFL classroom, for instance, providing feedback on writing can be viewed as such a relevant practice (Grossman, 2018, p. 168; Pang, 2019, p. 265; Summer & Plötz, 2024, p. 164). Feedback has proven to be one of the most impactful

⁶ RobS (Role-play-based Simulation) is used as a fixed term for the test that is at the focus of this thesis as well as for an abbreviation of any role-play-based simulation. For legibility, RobS is used as the singular as well as plural form.

variables influencing learning (Hattie, 2023, p. 320; Hattie & Timperley, 2007). However, there is also a great variance in the effect that different kinds of feedback have, indicating that it is essential for teachers to learn how to provide feedback appropriately (Hattie, 2023, p. 319; Wisniewski et al., 2020, p. 12). Feedback on writing in the EFL context has shown to be specifically relevant (Scherer et al., 2024). Conceptualizing writing as an iterative and nonlinear process, feedback is required in order for revisions to take place (Hayes & Flowers, 1980; Hyland, 2019, p. 12). As writing is also a skill that is statutorily part of Germany's EFL curricula, EFL teachers need to be able to provide appropriate feedback on writing products of their learners (Council of Europe, 2020, p. 68; Kultusministerkonferenz [KMK], 2023a, p. 16). Considering the importance of feedback in writing, this context offers a relevant focus for developing a performance-oriented assessment format.

To evaluate the quality of such a newly developed performance-oriented assessment format, the criterion of validity must also be tackled. Within an argument-based approach to validation, different claims of validity are made, investigated and discussed critically (M. T. Kane, 2013a). Following this approach, validity does not refer to the test itself but the extent to which theory and evidence support the intended interpretation of its results (AERA, APA, NCME, 2014, p. 11). From a theoretical perspective, the construct of the test must be described transparently. From an empirical point of view, data must be collected that allows to discuss the claims. All findings are then used to make an overall evaluation concerning the appropriateness of the intended score interpretation.

Based on these considerations, a gap in teacher education can be identified that is addressed in this thesis. Teacher education programs do not prepare PSTs adequately for the profession, as teachers do not fully capitalize on their potential positive impact. The gap between theory and practice is often cited as a challenge in teacher education programs. Using RobS as a way to approximate practice might be one approach to bridge this gap. As an assessment format in teacher education, RobS have the potential to influence PST's learning behavior, focusing more on action-oriented competences which are relevant for actual practice in the classroom. However, there is a lack of subject-specific, performance-oriented assessment formats. This gap is addressed specifically for pre-service EFL teachers (PSETs) in this thesis, as a RobS is developed for the assessment of PSETs' feedback competence on writing. To ensure the quality of the developed assessment format, the aspect of validity is addressed in the argument based-approach. Consequently, the following can be stated:

The aim of this thesis is to develop and validate a role-play-based simulation in the EFL context that can be used for summative assessment.

Structure of this Thesis

This thesis is structured in five main parts, each addressing a specific aspect of the research process. Each part is divided into chapters and concludes with a summary. The following section provides an overview of each part and its chapters.

In **Part I**, a more detailed look at EFL teacher education in Germany is presented and the goal of this thesis derived. In Chapter 2, the specific goals and structural elements that are relevant for this thesis are described. In Chapter 3, the term competence is focused on. Specific models of (teacher) competence are presented and the relevant statutory frameworks are summarized. Thereafter, selected empirical findings are summarized regarding professional knowledge of (pre-service) EFL teachers. Lastly, the extent to which the current approach of teacher education suffices to prepare PSETs for practice is critically discussed. In Chapter 4, a performance-oriented approach to university-based teacher education is outlined, focusing on learning objectives, teaching and learning activities, and assessment formats. Lastly, the current assessment practices in teacher education and a potential gap of performance-oriented assessment are critically discussed. As a consequence of the discussion, the aim of this thesis, i.e., developing a RobS, is described as one approach to address this gap.

In **Part II**, the design of the RobS is presented. In Chapter 6 the reasoning for the content focus of the RobS, i.e., assessing feedback competence on writing, is provided using theoretical models and empirical evidence. In Chapter 7, a framework is presented, which includes the relevant competence facets for describing and assessing feedback competence on writing holistically. In Chapter 8, the developed RobS is described as a potential assessment format.

In **Part III**, the methodological background of this thesis is outlined. In Chapter 10, different criteria for test evaluation are described. The relevance of validity is highlighted by providing a diachronic perspective on different conceptualizations of this criterion. It concludes with the argument-based approach to validation, which describes the postulation and evaluation of specific claims for an intended use of a test and is followed in this thesis (M. T. Kane, 2013a). In Chapter 11, the specific *interpretation/use argument*, which can be referred to as a list of claims for which evidence must be provided to evaluate the RobS, is presented.

In **Part IV**, the research process and the empirical findings for the test evaluation are presented. In Chapter 13, an overview of this thesis' two research cycles is given. In Chapter 14, the first cycle, i.e., the pre-pilot phase and the pilot survey, is focused on. The development and evaluation of the test material is detailed, as is the training of the actors for the RobS. In Chapter 15, all the studies that are executed in the main survey are subsumed. In Study 1, the perspective of the participants on the RobS is described. In Study 2, the qualitative expert

judgements of RobS performances and the scoring instrument is focused on. In Study 3, the process of designing the scoring instrument, based on different reliability indices on the item and test score level is detailed. In Study 4, findings of the testing of several hypotheses that are stated *a priori* to discuss the interpretation of test score are presented.

In **Part V**, the findings are discussed in terms of the test's quality. In Chapter 17, the previously made claims are evaluated using the findings generated in this thesis and a conclusion is drawn regarding the validation of the RobS. In Chapter 18, the overall thesis is summarized, limitations drawn and implication made. The appendix includes the relevant material that complements this thesis, i.e., the material for the RobS, the rating manual and transcription guideline for the RobS, the guidelines, transcription rules and coding manuals for the interviews and further statistical information.

**Part I -
(EFL) Teacher Education in Germany: From Developing Competence
to Assessing Performance**

*[The] quickest way to change student learning
is to change the assessment system.*

Lewis R.B. Elton and Diana Laurillard (1979, p. 100)

In the first part, the theoretical background for this thesis is given and the research aim derived. Firstly, the focus is put on the specific context of the study and the structure and goals of teacher education in Germany are described. Secondly, the notion of competence is focused on. The term is discussed theoretically first, before specific models of teacher competence are presented. The statutory guidelines for EFL teacher education are outlined and empirical findings presented that allow for an indication of competence development. The dissatisfaction that is associated with the status quo of teacher education is discussed as a result of these observations. Thirdly, a performance-oriented perspective is provided to address this issue. Based on the framework of constructive alignment by Biggs (1996) and current research, performance-oriented learning objectives, teaching and learning activities and assessment formats are proposed. Regarding the latter, an assessment gap in teacher education is identified, which leads to this thesis' research aim.

2 Structural Elements of Teacher Education in Germany

The aim of this first chapter is to describe the status quo of teacher education in Germany. Firstly, the aims of teacher education in Germany will be presented. Secondly, the history, structure, and organization of (EFL) teacher education in Germany will be outlined, so as to show how the development of achieving these aims is organized.

2.1 The Goal of Teacher Education

Teachers have a large impact on their learners' achievement, as stated in the introduction (Hattie, 2009, 2023). Thus, it is the job of the teacher to enable student learning in the most beneficial way. Consequently, teacher education programs should prepare PSTs effectively to do their job and impact student learning positively. The nature of the relationship can heuristically be depicted by a causal flow chart as seen in Figure 1 (Terhart, 2012b, p. 7).

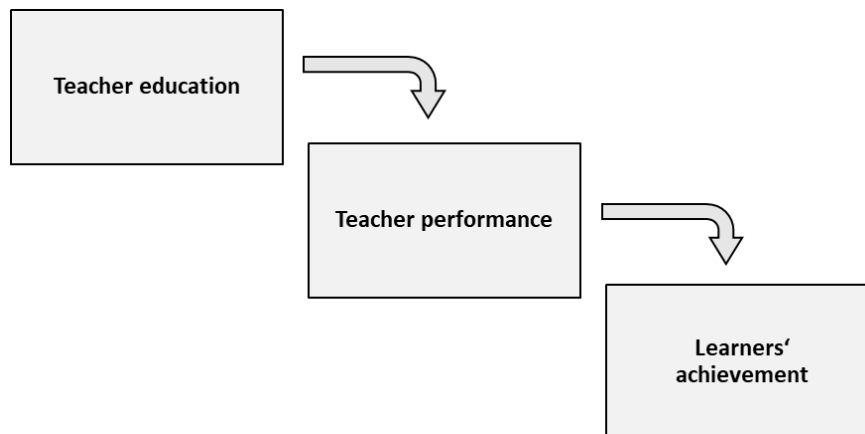


Figure 1: Causal flow chart depicting the influence of teacher education on teacher actions and the consequential influence on student learning (Terhart, 2012b, p. 7; translation TJ).

The assumption is that through teacher education, people acquire the necessary resources to fulfill the demands of the teaching profession (Vogelsang, 2014, p. 20). These resources can then be drawn upon to act adequately in concrete situations, which, in turn, have an impact on learning achievement. Consequently, learners' achievement is thus a direct result of teacher education. The exact nature of this relationship is, however, not that clear cut, as other potentially influencing variables are not included. Diez's model (2010; see Figure 2) takes mediating factors into account and adds an evaluative connotation as it includes questions to the proposed flow (Vogelsang, 2014, p. 21). In the model, the desired outcome of teacher education programs is described as "knowledge, skills, and dispositions" (Diez, 2010, p. 442), which are to be defined in standards that explicate the goals.

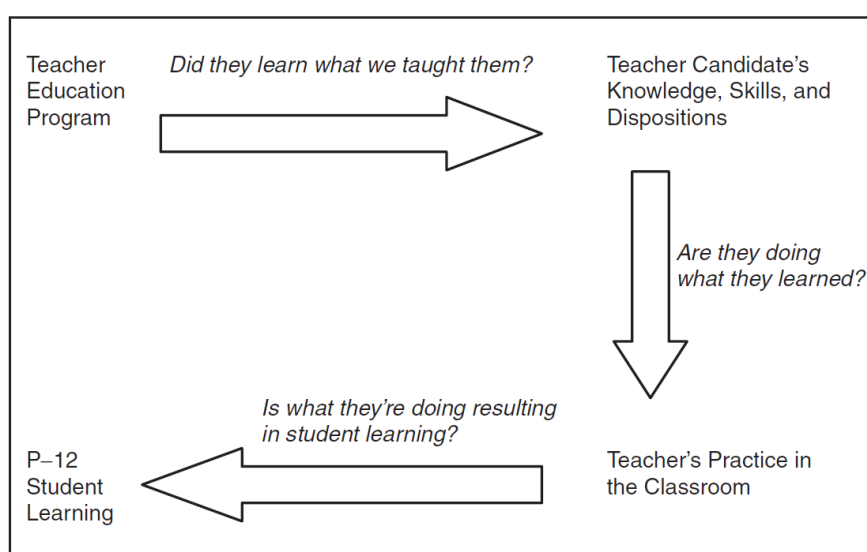


Figure 2: Flow diagram of impact on student learning (Diez, 2010, p. 442).

The first inquiry of evaluation then, should be to investigate the degree to which PSTs have achieved those standards. The second question is if teachers actually implement what they have learned in the classroom, with the subsequent final question being what the impact of their actions on P-12 student learning is, i.e., from pre-school to twelfth grade.

More recent models on the relation between teacher education and student learning have focused on teachers' professional competence. In the COACTIV⁷ study, teachers' professional competence is considered to be influential for the instructional quality, which in turn influences the students in their achievement⁸ (Krauss et al., 2020, p. 315). In her heuristic model, Voss links teacher education with a constructive understanding of teaching, creating a double offer-use-model⁹ that describes teacher education as one offer-use-situation, and the school's classroom as the other (Voss, 2019, p. 10). This model describes teachers' professional competence (see Chapter 3.2.1) as the result of teacher education, and a prerequisite for practice, similar to Diez. However, it also includes the learners and other variables like context, family or their individual learning (Helmke, 2022, p. 77; Voss, 2019, p. 11).

The cascade model by Krauss et al. (2020) is to date the most differentiating because it merges the model of competence as a continuum¹⁰ (Blömeke et al., 2015) with other relevant and salient paradigms in teacher education to describe the goal of teacher education (see Figure 3). The model includes five columns: Columns 1, 2 and 3 cover competence understood as a continuum as described by Blömeke et al. (2015), which, contrary to other models of competence, sees performance as part of competence, not a requirement for it (see Chapter 3). Columns 3 and 4, respectively, cover the offer-use-model as described, for example, by Helmke (2022, p. 77), while column 5 focuses on the outcome of the learning process on the learners' side. Evidently, teachers' competence, especially the performative aspect, i.e., the teacher's actual behavior, is central in all of these models. This part of the cascade intersects most clearly with learning – depending on how the teacher performs, learners might be more or less inclined to actually use the offers and achieve learning gains. The next chapter will describe how teacher education seeks to achieve its goal on a structural level.

⁷ COACTIV is the abbreviated form for *Cognitive Activation in the Mathematics Classroom and Professional Competence of Teachers* (Baumert & Kunter, 2013).

⁸ See Chapter 3.2.1 for a detailed overview of the COACTIV model of professional competence.

⁹ See also Helmke (2022, p. 77).

¹⁰ See Chapter 3.2.3 for a detailed description of the model.

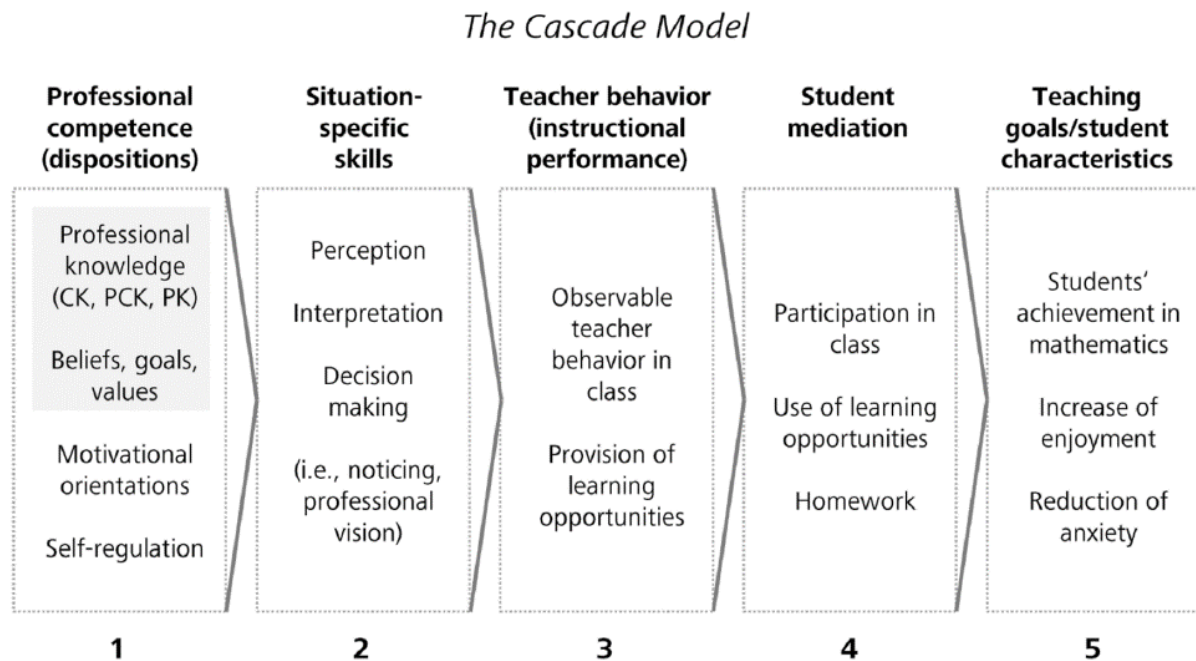


Figure 3: *The Cascade Model* (Krauss et al., 2020, p. 316).

2.2 The Structure of (EFL) Teacher Education in Germany

In Germany, the structure of teacher education, like education in general, is the responsibility of the individual federal state. In most federal states, teacher education is differentiated in three different phases:

The first phase: university-based education;

The second phase: teacher training or the induction phase, which takes place at training schools, as well as teacher education institutes (e.g., *Studienseminare, Zentren für schulpraktische Lehrkräfteausbildung*);

The third phase: further education as in-service teachers (Pasternack et al., 2017, p. 20; Terhart, 2007).

The main objective of the first phase is to lay the foundation for teaching the respective subjects. This includes a deeper knowledge about the subject matter, but also other pedagogical (content) dispositions (Pilypaitytė, 2013, p. 33). As a result of the Bologna process, the first phase is separated into a Bachelor's and a Master's program in most of Germany's federal states and, thus, covers at least five years of studies in total. The first phase also includes school internships of various lengths, allowing prospective teachers to gain experience in the role of the teacher. These practical phases are closely accompanied by the universities to ensure an adequate guidance and reflection (J. Meier et al., 2024, p. 312). In the state of North Rhine-Westphalia, the local context of the presented studies in Chapter 15, there are three practical phases (LABG NRW 2009, §12). In the