TEACHING QUALITY IN LITERATURE LESSONS OF STUDENT TEACHERS

Findings from a video study

FLORIAN HESSE

Department of German Literary Studies, Friedrich Schiller University Jena, Germany

Abstract

High-quality teaching is a pivotal element in achieving favourable student learning outcomes. However, previous research has hardly investigated the quality of literature teaching, especially not among student teachers. This study focuses on addressing this gap by analyzing literature lessons taught by student teachers during their five-month field experience in Germany. For the purpose of this study, high-quality teaching is defined as teaching that successfully integrates normative ideas about good teaching with effective practices that support student learning. Drawing on a newly developed, subject-specific model of teaching quality, this study assesses subject-specific and generic dimensions of teaching quality as well as variation in teaching quality among student teachers. To this end, the study designed new instruments for video-based observation and used them to evaluate *N* = 22 videotaped literature lessons on narrative texts. The results show that student teachers perform considerably better in generic quality dimensions than in subject-specific ones. Furthermore, standard deviations and cluster analyses indicate considerable student heterogeneity. These findings imply that teacher education programs should emphasize developing subject-specific teaching competencies in literature instruction even more and pay more attention to student teacher heterogeneity.

Keywords: teaching quality, literature teaching, student teachers, pre-service teachers, teaching practicum

Hesse, F. (2025). Teaching quality in literature lessons of student teachers: Findings from a video study. L1-Educational Studies in Language and Literature, 25, 1-29. https://doi.org/10.21248/l1esll.2025.25.1.856

Corresponding author: Florian Hesse, Department of German Literary Studies, Friedrich Schiller University Jena, Germany. Email: florian.hesse@tu-dresden.de

1. INTRODUCTION

According to theoretical assumptions and empirical findings of quantitative teaching research, the learning outcomes of pupils are primarily dependent on the subject-specific quality of the learning opportunities in the classroom (Decristan et al., 2020; Seidel & Shavelson, 2007). In the field of mathematics and science teaching, such learning opportunities have been subject of extensive research in recent years. However, only a few studies have been conducted to date on first language and especially literature teaching (Wieser, 2019). One of the reasons for this is that the development of a literature teaching-specific perspective on general educational approaches is still in its infancy, and there are only few instruments available to measure quality in literature lessons (Hesse & Winkler, 2022; Praetorius, Herrmann et al., 2020).

The research gap is particularly pronounced when one considers that the studies conducted thus far have concentrated almost exclusively on experienced teachers, while the teaching of student teachers has been largely unexplored. This is surprising given that the proportion of practical phases in teacher education has been increasing internationally in recent years, thereby rendering the teaching of student teachers more and more relevant in teacher training (Darling-Hammond, 2017; Gröschner & Hascher, 2022). In addition, the shortage of teachers in many countries—in Europe alone, 35 education systems report a shortage (UNESCO, 2024, p. 30)—has led to a situation in which, in some cases, over 50% of student teachers at a university are recruited to teach as temporary or substitute teachers without having completed their studies (Goldhaber et al., 2021; Hesse & Krause, 2024; Porsch & Reintjes, 2023). Consequently, it is important to systematically examine the learning opportunities designed by student teachers in order to ascertain their strengths and areas for improvement concerning instructional quality at an early stage of their training, with the aim of incorporating these insights into the curriculum of their teacher training.

The present article addresses the research gaps identified by linking the results of a German video study (Hesse, 2024) to the international discourse. The study, for the first time, used a subject-specific model of teaching quality alongside a newly developed rating instrument to analyze literature lessons taught by student teachers in Germany. The investigation placed a particular focus on the quality characteristics within both the subject-specific and generic dimensions, as well as on the extent to which the respective teachers differed in these dimensions and whether different realization variants could be identified. This insight could contribute to teacher training and further education. The study was conducted in three main steps: first, a comprehensive model of quality in literature teaching was developed; second, instruments for video-based assessment by external observers were constructed; and third, these instruments were used to assess a total of 22 lessons on narrative texts with regard to their teaching quality. This article will present the central results of the project after outlining the state of research and the methodological approach.

2. THEORETICAL BACKGROUND

2.1 Teaching and teaching quality

Despite decades of research, the terms *teaching* and *teaching quality* remain subject to a lack of uniform definition (Praetorius & Charalambous, 2023). The understanding of teaching on which this study is based is oriented towards the opportunity-and-use-models of teaching that are prevalent in German teaching research (Lipowsky et al., 2009; Vieluf et al., 2020). In accordance with constructivist teaching and learning theories, it is assumed that learning opportunities are provided in the classroom and are then utilized by learners in accordance with their cognitive and emotional prerequisites, thereby enabling learning. Consequently, pupils' learning is not predetermined by the teacher; rather, pupils must actively engage with the learning opportunities presented to them.

To support students in engaging with the subject matter in depth, teachers need to make sure that learning opportunities have a high quality. Berliner (2005) argues that one can speak of high-quality teaching when *good teaching* and *effective teaching* coincide. Good teaching follows normative ideas or didactic theories, whereas effective teaching helps pupils learn from an empirical point of view. For example, studies show that direct instruction is effective for student learning. However, the concept is often not considered good from a didactic perspective because it does not enable students to learn independently and through discovery (de Jong et al., 2023). Hence, good and effective teaching may not always coincide. When they do, however, one can speak of *high-quality teaching*.

Research on the quality of teaching has produced a wide range of considerations for modeling teaching quality. In many studies, a fundamental distinction is first made between surface and deep structures of teaching (Decristan et al., 2020). Surface structures of teaching refer to the teaching methods (e.g., direct instruction) and social arrangement of the lesson (e.g., group work, pair work), while deep structures refer directly to the facilitation of pupils' learning processes. Accordingly, the deep structures of teaching include characteristics that deal with the stimulation of deeper processes of understanding (e.g., cognitive activation, higher order thinking) or that ensure the necessary conditions for this (e.g., classroom management, classroom climate).

In recent years, numerous models of instructional quality have been proposed. Many of these models assess teaching from a generic (i.e., non-subject-specific), perspective. For example, the model of the *Three Basic Dimensions of Teaching Quality* assumes that the dimensions of cognitive activation, classroom management and constructive support can be used to assess teaching across all subjects, grades and possibly also cultures (Praetorius et al., 2018; Praetorius, Klieme, et al., 2020).

In the ongoing discourse on the assessment of teaching quality, there is a growing interest in the extent to which the dimensions identified in this and similar models (e.g., CLASS, PLATO) are adequate for a comprehensive description of teaching

quality. From both an educational science and a subject teaching perspective, there is a critique of the limited consideration given to subject-specific content and the approach taken to it (Praetorius, Herrmann, et al., 2020; Praetorius & Gräsel, 2021).

From an educational science perspective, the work of Praetorius and Charalambous (e.g., 2018) has helped to develop further the Three Basic Dimensions in recent years. The starting point was a review of four subject-specific, four generic and four hybrid frameworks for describing teaching quality, which the authors combined into a so-called *synthesis framework*. This showed that the Three Basic Dimensions have numerous blind spots, since, for example, aspects of the selection and presentation of subject matter are not considered at all in this framework. Accordingly, the synthesis framework estimates further dimensions beyond the Three Basic Dimensions, which, for example, concern the selection and presentation of subject matter, high-quality exercises or formative assessment (Praetorius, Rogh, et al., 2020).

From a subject teaching perspective, studies in the field of mathematics in particular have led to a rethinking of existing quality models. For example, Brunner (2017) was able to show that, depending on the weighting of subject-related quality criteria in the different frameworks, the same lesson could be rated as very good, average or below average. This led the author to develop a quality model in which the dimension *correctness* is a prerequisite for the other quality dimensions. Similarly, Lipowsky et al. (2018) were able to show that the subject-specific quality of the lessons in the Pythagoras study (cf. Drollinger-Vetter, 2011 for the operationalisation) forms an independent dimension that can be described separately from the other Three Basic Dimensions.

Beyond the question of which quality characteristics must be taken into account in research, the question of how quality models must be structured is the subject of intense debate in both educational science and subject-specific research. While quality dimensions have been modeled for many years as catalogues of quality features (e.g., Brophy, 2000; Helmke, 2014) or as loosely juxtaposed dimensions (as in the Three Basic Dimensions), the view that teaching quality should be thought of hierarchically is now gaining ground. This implies that some quality characteristics are more directly associated with student learning than others. For instance, it can be posited that quality characteristics such as cognitive activation are directly linked to the learning process, whereas quality characteristics such as classroom management or classroom climate exert an indirect influence on learning, serving as prerequisites.

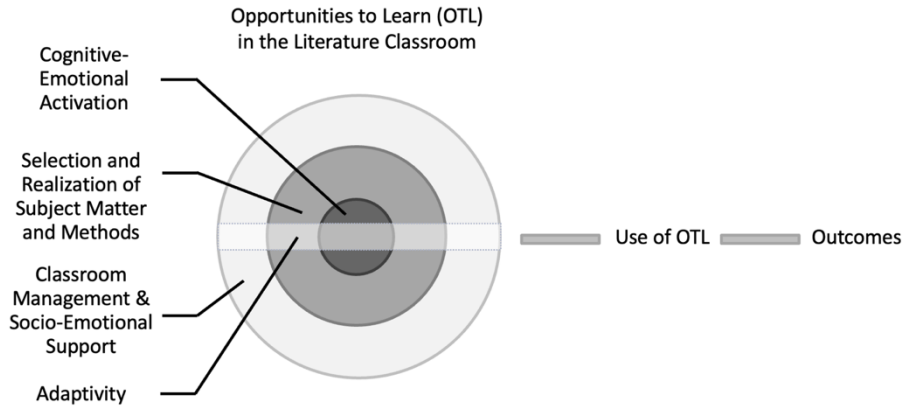
In conclusion, it can be stated that, in addition to generic characteristics, the modeling of teaching quality must also take subject-specific characteristics into account. In practical terms, this implies that the constructs which are predominantly influenced by educational psychology (e.g., cognitive activation, classroom management) must either be substantiated in subject-specific terms or complemented by subject-specific characteristics. Moreover, quality models must

exhibit a hierarchical structure, wherein the interrelationships between the diverse quality characteristics are elucidated.

2.2 Modeling teaching quality in the literature classroom

In view of the aforementioned considerations, a novel quality model of literature teaching was devised as part of a project entitled "Literature Teaching in the Teaching Practicum" (Hesse, 2024). The model includes subject-specific concretizations and is designed in a hierarchical structure (Fig. 1). As it serves as the foundation for operationalizing and measuring the quality of learning opportunities in student teachers' literature teaching in the present study, it will be presented in greater detail below.

Figure 1. Hierarchical model of teaching quality of OTL in the literature classroom



The model is based on the Three Basic Dimensions of teaching quality, which have not only shown their worth in numerous studies in German-speaking countries, but are also well aligned with international modeling of teaching quality (e.g., CLASS; Pianta et al., 2012; Pianta & Hamre, 2009). According to these models, teaching can be described as high-quality teaching if it is characterized by effective classroom management, constructive support and cognitive activation of the pupils. For the present quality model of literature teaching, the basic idea was adopted that quality literature teaching must (also) be characterized by good classroom management and socio-emotional support (classroom climate) in order for learning processes to succeed. In addition, the idea that good literature lessons are characterised by a high degree of cognitive activation was also incorporated into the model. This means that pupils must be given opportunities by the teacher to engage with the literary text in depth in order to be able to understand the text and learn something about themselves and others.

However, as Figure 1 already shows, some modifications must also be made to the Three Basic Dimensions. These initially concern subject-specific concretizations and additions. For example, Winkler (2020) has already emphasized in her concept of cognitive activation in literature lessons that these appear to be both good and effective if they succeed in activating learners not only cognitively but also emotionally. According to her, the results of existing intervention studies provide evidence for the "assumption that reader-oriented and text-oriented approaches to literature should be combined and complement each other in literature classes" (Winkler, 2020, p. 7). In line with more recent conceptualizations of teaching quality in artistic subjects (e.g., Frederking & Albrecht, 2016; Kranefeld, 2021; Rakoczy et al., 2021), the term *cognitive-emotional activation* is therefore used in this model to emphasize the importance of considering cognitive activation not merely as a stimulus for acquiring knowledge, but also as a way of becoming emotionally involved with the literary text.

A further modification to this model concerns the addition of the dimension *selection and realization of subject matter and methods*. This dimension captures subject-specific aspects of teaching quality that have often been neglected in previous modeling. For example, it records whether the potential of a literary text is fully utilized in the lesson, whether the learning opportunities follow a comprehensible logic or whether the teacher also pays attention to correctness and appropriateness. In relation to literature lessons, the latter means, for example, that the teacher ensures that terminology (e.g., in the field of narratology) is used correctly or that students' interpretations of literary texts are not arbitrary but are based on the text. There are two reasons in favor of adding this new dimension: On the one hand, it has already been noted in observations on literature lessons that it is necessary to consider ratings for cognitive or cognitive-emotional activation against the background of subject-specific correctness and appropriateness. For example, Winkler (2020, p. 25) remarks in the application of her rating instrument that the "tension between content appropriateness and transferability" is reflected but not yet made explicit enough in the rating instrument. On the other hand, there are also increasing calls from an educational science perspective to no longer think of modeling teaching quality exclusively in generic terms, but to take into account the subject-specific appropriateness of teaching. For example, the synthesis framework by Praetorius and Charalambous (2018) assumes a *selection and presentation of content* dimension, on which the present modeling is based.

Finally, a third modification concerns the hierarchical arrangement of the quality dimensions in a concentric model. In line with recent teaching quality modeling, it was initially assumed that the adaptivity of learning opportunities is of fundamental importance for all other dimensions. Teaching can only be successful if it is orientated towards the individual requirements of the learners. In a second layer of the model, it is assumed that literature lessons can only succeed if certain generic conditions for success, i.e. conditions that are important for all subjects, reach a certain threshold value: The teacher must ensure good classroom management to

maximize learning time and create a positive climate to increase learner motivation. If these basic conditions are not met, it cannot be assumed that the pupils will engage intensively and willingly with the literary texts. Engagement with the literary texts is then the subject of the other two dimensions: At the center of the model is the cognitive-emotional activation, which captures the extent to which learners are encouraged to engage cognitively and emotionally with the text. The dimension selection and realization of subject matter and methods captures whether this engagement is appropriate.

Prior to delineating the operationalization of the model for the observation of videotaped lessons in greater detail, it is first necessary to summarize the results of previous studies dedicated to the observation of everyday literature or language lessons.

3. STATE OF RESEARCH

As the literature lessons of student teachers have hardly been researched to date, other studies that are at least comparable with the phenomenon must be used as a point of reference. On the one hand, these are observational studies on the teaching of experienced teachers in first language (literature) lessons. On the other hand, there are observational studies on the teaching of student teachers in other subjects.

3.1 Observational studies on first language teaching with experienced teachers

In the German-speaking countries, there are only a few observational studies to date that examine first language teaching in general or literature teaching in particular and thereby connect to the discourse on teaching quality.

In her PhD thesis, Lotz (2016) analyzes the reading practice phases of $N = 37$ primary school classes on the picture book *Lucy rettet Mama Kroko* (engl. *Lucy rescues Mama Croc*) with regard to the potential for cognitive activation as well as generic dimensions such as classroom management and classroom climate. As a result, Lotz comes to the rather disillusioning conclusion that although teachers are generally successful in creating the underlying conditions for cognitively activating lessons (in terms of classroom management and classroom climate), much potential remains unutilized in terms of cognitive activation itself (Lotz, 2016, p. 336). This is expressed, for example, by the fact that teachers predominantly set simple tasks, ask learners knowledge questions rather than thinking questions and—if reflection takes place at all—it only lasts a few seconds, and the students are hardly involved (Lotz, 2016).

A comparable result was obtained by Schmidt (2016), who analyzed $N = 37$ primary school classes in art lessons in the same project context as Lotz. The subject of the lesson is the painting *The Tilled Field* by Miró. Schmidt assesses the lessons with a total of 20 subject-specific and non-subject-specific items. Similar to Lotz (2016), it was found that rather few measures were taken to cognitively activate the

learners during the reception of art in the sample analyzed (Schmidt, 2016, p. 138). Schmidt makes this clear from the finding that the theoretical scale mean of 2.5 was not exceeded in any of the 20 items assessed (Schmidt, 2016, p. 140).

Comparable findings are available in German-speaking countries from video studies on first language teaching in the areas of grammar (*N* = 7 classes; Stahns, 2013) and spelling (*N* = 42 classes; Riegler et al., 2022). These studies also indicate that the potential for cognitive activation remains suboptimal and that the lessons often show weaknesses, particularly in terms of correctness.

In the international discourse, observation studies have emerged in recent years primarily from the QUINT network (Quality in Northern Teaching). Here, language arts lessons in Norway, Sweden, Finland and Iceland have been analyzed from various perspectives. A study by Nissen et al. (2021) can show on a fundamental level that literature lessons play an important role in first language teaching across all the countries analyzed. Nevertheless, a more in-depth analysis by Gabrielsen and Blikstad-Balas (2020) based on *N* = 47 classes (178 videos) found that the repertoire of literature made available to students is limited.

In addition to missed opportunities in the text selection, the instructional quality also reveals weaknesses. Magnusson et al. (2019) can show for 47 classes (178 videos) from Norway that there is insufficient explicit teaching of reading strategies in the lessons observed (see also Tengberg et al., 2022, for similar results). Based on an analysis of 38 Language Arts classes (141 videos) using PLATO (Grossman et al., 2015), Tengberg et al. (2021) were able to show that, similar to German-speaking countries, the quality of teaching only achieved good results in the areas of behavior management and time management. All other dimensions do not reach the theoretical scale average of 2.5 on a four-point scale from 1 to 4 in almost all dimensions. There are particular weaknesses in the areas of Modeling, Connections to Prior Knowledge, and Strategy Use and Instruction. Nissen (2023) is able to substantiate this impression in an in-depth analysis of 49 language arts classes in Norway and Sweden, in which she examines the potential for cognitive activation of a total of 174 tasks. She found that these tasks most often implied cognitive activation coded as *recall and share* or *understand and explain*, comparatively low-level cognitive processes. In contrast, tasks requiring *analyze and create* were rarely observed.

Both the German-language and the international discourse thus suggest that a high level of teaching quality is not always to be expected in the first language lessons of experienced teachers. It is true that in most studies, teachers are able to create a learning-friendly environment (classroom management, classroom climate). However, aspects of subject-related teaching quality are weaker in comparison. This applies to aspects of cognitive or cognitive-emotional activation (critical thinking, use of strategies, etc.) as well as aspects of the selection and realization of subject matter and methods (e.g., text selection, subject-specific correctness). Finally, a high degree of heterogeneity within the data is noticeable in many studies. For example, Tengberg et al. (2021, p. 775) state “that the instruction in different classrooms,

sometimes neighboring classrooms in the same school, is of systematically different quality".

3.2 Observational studies on first language teaching among student teachers

The above-mentioned studies all consider experienced teachers. Although studies repeatedly show that mere professional experience is not a decisive criterion for teaching quality (e.g., Graham et al., 2020) and that experienced teachers do not represent a homogeneous group in terms of teaching quality (see Tengberg et al. above), the following section will nevertheless look at some studies that deal explicitly with student teaching. As there are no studies on literature teaching in this area to date, studies on the teaching of other school subjects will be referred to.

In mathematics and science teaching, for example, the "Φactio study" (Korneck et al., 2017; Szogs et al., 2016) assessed the quality of short teaching sequences for 125 physics student teachers using ratings. This study showed, among other things, that the students only achieved a mean score of 1.65 on a four-point scale from 0 to 3 in the grammar school teaching degree, and only 1.39 in the regular school teaching degree. Vogelsang (2014) came to similar conclusions when he assessed the teaching of eight trainee teachers and 14 student teachers as part of his study. He was able to show that most of the sub-dimensions were roughly in the middle of the four-point Likert scale (1–4). The students were particularly successful in providing challenging learning opportunities (M = 2.9), while there was greater potential for development, for example, in raising awareness of the learning status (M = 1.9) or in dealing with students' ideas (M = 2.0; cf. Vogelsang 2014, pp. 401–402). Similar to the studies on experienced teachers, however, student teachers in both studies succeed well in realizing framing conditions (e.g., classroom management).

Comparable results can also be found in a longitudinal study by Greve et al. (2021) on physical education among 11 student teachers. Their lessons were assessed using the CLASS instrument (Pianta & Hamre, 2009) and, analogous to the above studies, only showed positive developments in the area of negative climate, while items such as instructional support were rated low from the outset and did not improve over the course of the internship. An interdisciplinary longitudinal analysis by Baer et al. (2014) shows comparable results in an analysis of 98 videos, insofar as cognitive activation also shows significant potential for improvement here.

Overall, the findings available show a picture across various subjects that corresponds with the findings on experienced teachers. While generic quality dimensions tend to be realized well, there are considerable problems with the subject-specific dimensions, in particular with cognitive activation and with the selection and realization of subject matter and methods.

4. METHODS

4.1 Research questions and hypotheses

Against this background, this study examines the extent to which the above findings can also be replicated for literature lessons held by student teachers. The first step is to examine generic and subject-specific quality dimensions. In line with the other studies, it is assumed here that generic quality dimensions tend to be better realized than subject-specific ones.

> RQ 1.1: To what extent are the generic quality dimensions of teaching developed (classroom management, socio-emotional support)?
>
> RQ 1.2: To what extent are the subject-specific quality dimensions of teaching developed (cognitive-emotional activation, selection and realization of subject matter and methods)?

In addition to the question of the average value of the individual dimensions, the second step is to look at the variance in the data. Previous studies suggest great heterogeneity in this respect. Since it is not only of interest in the professionalization of student teachers that there is diversity, but also in which aspects the students differ, cluster-analytical methods will be used to explore whether different patterns of teaching quality can be identified in the various quality dimensions.

> RQ 2.1: To what extent is there heterogeneity in the equality assessments of the lessons examined?
>
> RQ 2.2: Can different realization variants be identified within the individual quality dimensions?

Given the absence of reliable findings on this question to date, a hypothesis-testing procedure is not employed, and exploratory methods are used instead.

4.2 Data collection and sample

To examine the research questions and hypotheses, a corpus of videotaped literature lessons is used, which were held by students from the University of Jena, Germany, during their practical semester as part of the OVID-PRAX project (Gröschner et al., 2019). The aim of this project was to investigate the effectiveness of online-based video feedback on the development of professional skills. To this end, a total of 38 students in a video group were asked to videotape a lesson of their own choice and upload an excerpt of the lesson to a Moodle platform for feedback purposes. While the OVID-PRAX project focused on the development of professional competence through feedback on the lesson excerpts, the present study concentrated on the analysis of the lesson videos themselves. Here, however, not only the video excerpts were considered, but the complete lessons (approx. 45–90 minutes each), which were available from a total of 22 students ($N = 22$ videos/classes).

In terms of content, the student teachers were largely free to decide how to design the lesson. The only requirement was that the students had to negotiate a short epic text and conduct at least one plenary discussion. However, the latter requirement should not be overestimated, as the student teachers were allowed to determine the purpose and length of the plenary phase themselves. There were also no specifications regarding the grade level to be taught, as student teachers are generally not able to choose their classes in the teaching practicum but are assigned them by the practicum schools or mentor teachers. Accordingly, the sample includes lessons in different class levels at lower and upper secondary level. The sample did not include any student teachers studying to become primary school teachers, as these are not trained at the university where the study was conducted.

Concerning the prerequisites of the student teachers, it should be noted that the majority of them were in their fifth or sixth semester at the time of the videography. In the previous semesters, all students had already attended basic lectures on topics such as literary history, literary analysis, literary theory and literature teaching. As part of their five-months practical semester, in which the students first observe lessons and then design and conduct lessons themselves, the student teachers were also supported by the university: In the accompanying lectures and seminars, selected example texts and tasks were used to discuss how literary texts can be analyzed with a view to student learning, how tasks can be set that promote learning and how one can deal productively with pupil responses in classroom discussions. Moreover, student teachers received an overview about generic quality dimensions of teaching (e.g., classroom management, classroom climate) and techniques for effective discussion (based on dialogic teaching research). Accordingly, when interpreting the results, it must be taken into account that the students were familiar with many aspects of (subject-specific) teaching quality, at least on a theoretical level.

4.3 Rating of the quality dimensions

The assessment of the quality features of the lessons was carried out by trained raters and was based on a specially developed and piloted manual including 23 items (Table 1). The conceptual basis was the model for assessing quality in literature lessons described above (Figure 1), which comprises a total of two subject-specific dimensions (cognitive-emotional activation; selection and realization of subject matter and methods) and two generic quality dimensions (classroom management; socio-emotional support).

Table 1. Overview of the subject-specific and generic quality dimensions and their items

Quality Dimension 1: Cognitive-emotional activation (CEA)	
CEA_1	Encouraging precise perception of the text
CEA_2	Encouraging subjective involvement
CEA_3	Encouraging the interplay of precise text perception and subjective involvement
CEA_4	Encouraging reflection on the text comprehension process
CEA_5	Encouraging the elaboration of student statements
CEA_6	Providing support in tasks
CEA_7	Providing support in the teacher-student-interaction
Quality Dimension 2: Selection and realization of subject matter and methods (SEL)	
SEL_1	Selection and use of the text
SEL_2	Structuring the lesson
SEL_3	Correctness and appropriateness
SEL_4	Suitability of methods
Quality Dimension 3: Classroom management (CM)	
CM_1	Withitness
CM_2	Group focus
CM_3	Clarity and enforcement of rules
CM_4	Dealing effectively with disturbances
CM_5	Absence of disturbances
CM_6	Time management
CM_7	Transition management
CM_8	Preparation of the classroom
Quality Dimension 4: Socio-emotional support (SES)	
SES_1	Positive charisma of the teacher
SES_2	Appreciation of the students by the teacher
SES_3	Dealing positively with comprehension difficulties and errors
SES_4	Positive learning atmosphere

For the two subject-related dimensions, 11 rating items were newly developed. With regard to the cognitive-emotional activation, the present study aligns with existing research (Lipowsky, 2020; Nissen, 2023; Winkler, 2020) by assessing cognitive-emotional activation through both tasks (CEA_1–CEA_4 & CEA_6) and the quality of the teacher-student interaction (CEA_5 & CEA_7). In line with Winkler, emphasis was also placed on the realization, integration and reflection of text- and reader-related perspectives (CEA_1–CEA_4). The items in the area of selection and realization of subject matter and methods are inspired by subdimensions suggested in the synthesis framework by Praetorius and Charalambous (2018) and at the same time modified with regard to teaching literature. For example, the item on correctness in Praetorius and Charalambous (2018) was modified since statements in literature lessons cannot always be judged as correct or incorrect, but rather as more or less appropriate. Accordingly, the item refers to correctness and appropriateness (SEL_3). For the two generic dimensions, a total of 12 rating items were taken from existing studies (e.g., Gabriel & Lipowsky, 2013a, 2013b; Pianta et al., 2012; Thiel et al., 2013).

All of the items shown here were designed in four levels (1–4) and relate to the entire lesson. Descriptions were given for each level, providing information on when

the level should be assigned. In addition, there were brief descriptions for each item as well as a list of indicators that suggest a particularly high level of the item. An example for the item *structuring the lesson* (SEL_2) can be found in Appendix 1. The complete rating manual for the study can be found in Hesse (2024).[1]

When adapting or designing the items, care was also taken to ensure that high values (levels 3 and 4) of an item stand for a frequent and adaptive realization of the relevant quality characteristic in the lesson, while low values (levels 1 and 2) stand for a low frequency and adaptivity. This is illustrated by the example of *encouragement to reflect on the text comprehension process* (CEA_4). This item is about the teacher encouraging pupils to reflect on the generation of interpretation hypotheses or feelings when reading (i.e., meta-cognition and meta-emotion). In this context, high-quality teaching does not consist of reflecting throughout the entire lesson, but rather of ensuring that the reflection phases are in an appropriate proportion to previous phases of the lesson and that those aspects are reflected that were actually central (and not peripheral) to the lesson. In terms of the cross-sectional dimension adaptivity of the quality model (Figure 1), it is therefore not the mere quantity of reflection opportunities that is decisive, but their quality and embedding in the lesson as a whole.

4.4 Inter-rater reliability and internal consistency of the scales

In order to test the reliability of the items used, a reliability study was conducted with three independent raters (author of the article + two trained student assistants). In this study, each lesson was assessed independently by all three raters, who had previously received intensive training with the manual and comparable video examples from another video corpus. To calculate the inter-rater reliability, two-factor intra-class correlations were calculated on the basis of unadjusted individual values (Wirtz & Caspar, 2002) after 10 and after 22 videos. With the exception of one item (encouraging the elaboration of students' statements[2], CEA_5), the inter-rater reliabilities for the items were consistently satisfactory (.722 > ICC > .907), measured against the high requirements of the rating systems. In cases where the raters rated items differently, the rating value of the author of the study (= master rater) was used in accordance with the procedure of other studies (see e.g., Lotz, 2016).

---

1 Link to the manual: https://static-content.springer.com/esm/chp%3A10.1007%2F978-3-662-68775-8_9/MediaObjects/618076_1_De_9_MOESM1_ESM.pdf

2 Due to a lack of project resources, it was not possible to revise and retest the reliability of the item as part of the project. Due to the importance of the item for the construct of cognitive-emotional activation, it was nevertheless decided to include the item in the further statistical analysis. However, this is accompanied by the strong limitation that the reliability of the assessment of the item could not be ensured. In current follow-up projects, the item has already been revised and will hopefully achieve more reliable values.

Cronbach's Alpha was calculated to check the internal consistency of the items within the respective scales (Döring & Bortz, 2016, pp. 467–468). Sufficient values were found in all four dimensions, so that the scale values could also be used for further calculations (Table 2).

Table 2. Internal consistency (Cronbach's Alpha)

Dimension	*Items*	*N*	*α*
Cognitive-emotional activation	7	22	.748
Selection and realization of subject matter and methods	4	22	.825
Classroom management	8	22	.930
Socio-emotional support	4	22	.746

4.5 Analysis of the data

To answer RQ 1.1 and RQ 1.2, the mean rating values for the items and scales were calculated for both the subject-specific and generic quality dimensions. With regard to the RQ 2, different statistical methods were applied. Standard deviations, minimum and maximum values were calculated at item and scale level in order to gain an initial impression of the dispersion of the quality measures (RQ 2.1). For insights into possible realization variants of the different quality dimensions (RQ 2.2), agglomerative, hierarchical cluster analyses were used, which are particularly suitable for small samples (Bacher et al., 2010; Backhaus et al., 2016). In this form of cluster analysis, each class/student teacher is initially considered as an individual case and merged into larger clusters until a jump in the fusion coefficient is reached, which indicates a poorer fit to the data. This jump can be determined both graphically using the elbow criterion and mathematically using, for example, the Mojena test (Backhaus et al., 2016, p. 495–497). After determining the final cluster solution with an appropriate number of clusters, t-values were calculated which provide information on whether a characteristic within a cluster is over- or underrepresented compared to the overall group. This makes it possible to better interpret the clusters.

5. RESULTS

5.1 General descriptive results

Looking first at RQ 1, it can be seen that the generic teaching quality dimensions, classroom management and socio-emotional support, perform far better than the subject-specific dimensions (Table 3). It is striking that the generic quality dimensions, and here in particular socio-emotional support, almost reach the maximum of the scale, while the subject-specific dimensions only just exceed or fall short of the theoretical scale mean (2.5).

Table 3. Descriptive statistics and internal consistency of the quality dimensions

Dimension	*Items*	*N*	*M*	*SD*	*Min*	*Max*
Cognitive-emotional activation	7	22	2.32	.54	1.29	3.29
Selection and realization of subject matter and methods	4	22	2.60	.65	1.50	3.75
Classroom management	8	22	3.35	.63	2.13	4.00
Socio-emotional support	4	22	3.48	.48	2.50	4.00

However, when looking at the standard deviations as well as the minimum and maximum values, it is also noticeable that there is a considerable variance in most dimensions (RQ 2.1). For example, there are 2.25 points between the worst and best rating score in the selection and realization of subject matter and methods dimension. This shows that lessons on comparable literary texts can be realized by teachers with a comparable level of experience to very different degrees of quality.

As it is not only interesting for teacher training whether there *is* variance in the data, but also whether certain teaching patterns can be identified with regard to the individual dimensions, agglomerative hierarchical cluster analyses were also carried out. For reasons of space, only the subject-specific dimensions (i.e., cognitive-emotional activation, selection and realization of subject matter and methods) are presented in detail here because they are regarded to be particularly relevant for this article. The results for the two generic dimensions, classroom management and classroom climate, are only summarized briefly below.

5.2 Cognitive-emotional activation

A closer look at the individual items of the cognitive-emotional activation dimension initially illustrates that the various facets of the construct were rated differently (Table 4).

Table 4. Descriptive statistics: Cognitive-emotional activation

Code	*Item*	*M*	*SD*	*Min*	*Max*
CEA_1	Encouraging precise perception of the text	2.91	.97	1	4
CEA_2	Encouraging subjective involvement	2.64	.95	1	4
CEA_3	Encouraging the interplay of precise text perception and subjective involvement	1.95	.79	1	4
CEA_4	Encouraging reflection on the text comprehension process	1.55	.67	1	3
CEA_5	Encouraging the elaboration of student statements	2.55	.96	1	4
CEA_6	Providing support in tasks	2.41	.80	1	4
CEA_7	Providing support in the teacher-student-interaction	2.77	.77	1	4

It is evident, for example, that on average the lessons focus more on the text (CEA_1) than on the learners (CEA_2) and that text-related and learner-related perspectives are at best only partially related to each other (CEA_3). It is also noticeable that the comprehension process in the lessons is hardly reflected upon to any considerable

extent (CEA_4) and that the two interaction-related items (CEA_5; CEA_7) are only slightly above the theoretical scale average, despite intensive processing in the preparatory and accompanying seminars from a didactic and educational science perspective. The support in the learning tasks themselves (CEA_6) is even slightly lower.

Beyond the mean value analyses, a considerable variance of the quality assessments is noticeable, both in Table 2 and, above all, in Table 4. With the exception of item CEA_4, which is low across almost all lessons and does not reach the theoretical maximum of 4, all values were awarded for the other items. The standard deviations also indicate a large variance in the data. As already explained in the previous section, this finding was taken as an opportunity to conduct a cluster analysis to examine whether certain distribution patterns can be found in the data. As no sufficient variation was observed between the cases for item CEA_4, the item was not included in the cluster analysis.

As a result of the cluster analysis, four clusters of cognitive-emotional activating teaching programs were identified using the Elbow criterion and Mojena's test, which are shown in Table 5. The mean values refer to the average values of the characteristics within a cluster; the corresponding t-values indicate whether the mean value is overrepresented (positive value) or underrepresented (negative value) compared to the total values in the sample. Values greater than |0.5| indicate greater deviations from the overall survey population and are highlighted in color (orange = underrepresented, green = overrepresented).

Table 5. Results of the cluster analysis: Cognitive-emotional activation

		Cluster 1 *N* = 3		Cluster 2 *N* = 10		Cluster 3 *N* = 4		Cluster 4 *N* = 5	
Code	Item	*M1*	*t1*	*M2*	*t2*	*M3*	*t3*	*M4*	*t4*
CEA_1	Encouraging precise perception of the text	1.33	-1.62	2.80	-0.11	3.00	0.09	4.00	1.12
CEA_2	Encouraging subjective involvement	2.33	-0.32	2.60	-0.04	3.75	1.16	2.00	-0.67
CEA_3	Encouraging the interplay of precise text perception and subjective involvement	1.00	-1.21	2.00	0.06	3.00	1.34	1.60	-0.45
CEA_4	Encouraging the elaboration of student statements	2.00	-0.57	1.90	-0.67	3.50	0.99	3.40	0.88
CEA_5	Providing support in tasks	1.00	-1.77	2.30	-0.14	3.25	1.06	2.80	0.49
CEA_6	Providing support in the teacher-student-interaction	2.00	-0.35	1.90	-0.48	2.50	0.30	3.00	0.95

The first cluster comprises lessons (*N* = 3) with *little cognitive-emotionally activating learning opportunities*, insofar as all assessments here are low or underrepresented both in terms of mean values and in relation to the survey total. These lessons are particularly characterized by the fact that the literary text is hardly present and therefore no interplay between text perception and subjective involvement is possible.

The second cluster describes *average cognitive-emotionally activating lessons with weaknesses in interaction* (*N* = 10). Students in this cluster are much more successful in creating opportunities that focus on the text, and the interplay between subjective involvement and perception of the text is also initiated at least to some extent. However, there is no difference compared to the previous cluster with regard to interaction-related items, which achieve the lowest values in this cluster. This becomes clear, for example, in teaching situations in which students continuously ask for further interpretative hypotheses from the students using a series of add-on questions (e.g., “Does anyone else want to say anything about this?”) without being clear which other hypotheses (beyond those already mentioned) the learners could put forward.

Compared to the two previous clusters, the third cluster (*N* = 4) can be described as *strongly cognitive-emotionally activating teaching*, in that both mean and t-values are consistently high. The comparatively high values for the encouragement of subjective involvement and the interplay between involvement and text perception

are particularly noteworthy. Only in the area of stimulating precise text perception and providing support in the interaction does the cluster score only average. An average score in the area of precise text perception is characterized by the fact that the majority of students focus on content- or presentation-related aspects of the text and do not or only rarely relate these to each other.

In the fourth cluster (*N* = 5), the items that are less well realized in the third cluster are particularly well developed. The encouragement to perceive texts accurately achieves maximum values here and the interaction-related items are also consistently high both in terms of mean and t-values. Against this background, the cluster can also be labelled as *text- and interaction-related activating teaching*. The only weaknesses in this cluster are that the lessons have the lowest value in the cluster comparison in the area of encouraging subjective involvement and also achieve the second-lowest value in the interplay between involvement and text perception. The tasks set by the teachers are therefore not, or only to a limited extent, aimed at the learners' prior knowledge, their text-related emotions and judgements.

5.3 Selection and realization of subject matter and methods

In this second subject-related quality dimension, it is measured to what extent the teachers utilize the potential of the text, structure the lesson logically in terms of content, prepare it appropriately and select suitable methods and social forms. As with cognitive-emotional activation, the examined participants only achieved values around the theoretical scale mean (Table 6).

Table 6. Descriptive statistics: Selection and realization of subject matter and methods

Code	*Item*	*M*	*SD*	*Min*	*Max*
CEA_1	Selection and use of the text	2.73	.77	1	4
CEA_2	Structuring the lesson	2.55	.86	1	4
CEA_3	Correctness and appropriateness	2.59	.80	1	4
CEA_4	Suitability of methods	2.55	.80	1	4

In addition, there is also a large variance in the data in this dimension, as can be seen from the high standard deviations and the large ranges (from 1 to 4). Accordingly, it also seemed appropriate to carry out a hierarchical cluster analysis with all items in this dimension.

Table 7. Results of the cluster analysis: Selection and realization of subject matter and methods

	Cluster 1 (*N* = 7)		Cluster 2 (*N* = 5)		Cluster 3 (*N* = 9)	
Item	*M1*	*t1*	*M2*	*t2*	*M3*	*t3*
Selection and use of the text	1.86	-1.09	3.60	1.13	2.89	0.22
Structuring the lesson	2.14	-0.60	3.60	1.22	2.44	-0.22
Correctness and appropriateness	1.86	-1.10	3.40	1.00	2.89	0.31
Suitability of methods	2.29	-0.35	3.60	1.27	2.22	-0.43

Looking at Table 7, it becomes clear that Cluster 1 and Cluster 2 are almost opposite. Cluster 1 (*N* = 7) contains cases that have low values across all items, but particularly for the items *text selection and use* as well as *correctness* and *appropriateness*, which are also underrepresented with regard to the sample as a whole (cf. the strongly negative t1 values highlighted in dark gray). Here, one could therefore also speak of a *predominantly inadequate selection and realization of subject matter and methods*. Cluster 2 (*N* = 5) on the other hand, bundles cases with high values that express a *predominantly adequate selection and realization of subject matter and methods*. Cluster 3 (*N* = 9) ultimately lies between the other two clusters, whereby the proximity to the first cluster predominates, particularly with regard to the negatively pronounced items *structuring the lesson* and *suitability of methods*. Consequently, these cases can be considered as a *partially inadequate selection and selection and realization of subject matter and methods*. The last cluster is particularly interesting because it suggests that structuring lessons and selecting suitable methods are additional requirements that go beyond the use of text potential and subject-specific appropriateness and that not all student teachers have mastered.

5.4 Classroom management and socio-emotional support

Finally, looking at the two generic dimensions of teaching quality, classroom management and socio-emotional support (Table 8), it is noticeable that these are rated significantly better overall than the subject-specific items. This is evident not only from the fact that all items exceed an average value of 3 on a four-point scale, but also from the fact that the range of the rating scale is not fully utilized. For example, there are only two items for which the rating scale is fully utilized. In addition, there are two items for which only the values 3 and 4 were assigned.

Table 8. Descriptive statistics: Classroom management and socio-emotional support

Dimension	*Items*	*M*	*SD*	*Min*	*Max*
Classroom management	Withitness	3.23	.75	2	4
	Group focus	3.27	.63	2	4
	Clarity and enforcement of rules	3.41	.80	1	4
	Dealing effectively with disturbances	3.36	.85	2	4
	Absence of disturbances	3.32	.95	1	4
	Time management	3.32	.65	2	4
	Transition management	3.27	.83	2	4
	Preparation of the classroom	3.59	.59	2	4
Socio-emotional support	Positive charisma of the teacher	3.45	.60	2	4
	Appreciation of the students by the teacher	3.55	.51	3	4
	Dealing positively with comprehension difficulties and errors	3.64	.49	3	4
	Positive learning atmosphere	3.36	.79	2	4

The results of a cluster analysis for both quality dimensions showed that there are three groups of student teachers in both dimensions. The first group represents student teachers who achieve *moderate* results across all items in the respective dimension. However, only four (classroom management) and five prospective teachers (socio-emotional support) were counted in this group. The remaining students were assigned in roughly equal proportions to the other two groups, which represent *good* or *very good* classroom management and socio-emotional support respectively.

6. DISCUSSION

This article presents the results of a study which, for the first time, used a subject-specific model of teaching quality and a newly developed rating instrument to analyze student teachers' literature lessons. A particular focus was placed on the quality characteristics in the subject-specific and generic quality dimensions (RQ 1). At the same time, it was also investigated to what extent the respective teachers differ in the quality dimensions and whether different realization variants can be identified within the dimensions, which can be referred to in the training and further education of teachers (RQ 2).

As a result, it was found that students perform very well on average in the generic quality dimensions (RQ 1.1), which is in line with the current state of research. On the one hand, this result is encouraging as it shows that students in the practical semester are already able to pave the way for central *prerequisites* for high-quality subject-specific learning opportunities. On the other hand, it is important to bear in mind that aspects of classroom management and socio-emotional support in particular can be subject to bias in videographic studies. This does not only refer to camera effects, which have tended to be relativized in methodological research in the past (Petko et al., 2003). Rather, the positively biased selection of the videotaped

classes is likely to play a role, as is the fact that the students may also benefit from classroom management and teaching climate aspects that have been already established by their mentor teachers (Junker et al., 2021). Accordingly, the results should not be interpreted as a free pass not to continue to address classroom management and classroom climate aspects in teacher education and training in the future.

Less positive assessments were found in the subject-specific dimensions, cognitive-emotional activation and selection and realization of subject matter and methods. Here, the students only achieved values that were below or only just above the theoretical scale mean. This result is also not specific to the present sample but is in line with existing research findings on student teaching (Section 1.3.2).

This result is problematic in that not every lesson that is well designed on the level of generic quality dimensions must also be designed in a subject-specifically appropriate and stimulating way. This confirms the assumption of the hierarchical quality model, according to which generic quality dimensions are an important prerequisite, but not sufficient conditions for lessons that are conducive to learning. However, since the subject-specific quality of teaching is decisive for pupils' learning outcomes, it is important to consider the extent to which student teachers can be trained in the future in such a way that they can design more appropriate and activating lessons. Simply familiarizing students with quality dimensions as part of university courses does not appear to be sufficient. After all, the students in this study knew subject-specific quality features (e.g., tasks that promote learning, productive handling of student responses) from the seminars accompanying the practical semester (Section 2.1). Moreover, the low scores cannot be explained by the students' lack of experience, as the teaching of experienced teachers also shows considerable weaknesses in some cases (Section 1.3.1) and because other studies have shown that professional experience has no influence on teaching quality (Graham et al., 2020). It could therefore be more helpful to focus on long-term coaching programs in which (prospective) teachers are continuously supported and in which reflect on their own teaching (e.g., Allen et al., 2011; Gore et al., 2017).

In this regard, the heterogeneity among the students found in the data must also be taken into account (RQ 2). It is not only the standard deviations that show that the students differ considerably in their realization of the various dimensions, similar to the study by Tengberg et al. (2021) (RQ 2.1). The cluster analyses also identified different profiles of high-quality literature teaching (RQ 2.2), similar to a study by Wiprächtiger et al. (2023).

By using cluster analyses, the article was able to show that student teachers already differ considerably in terms of the quality of their teaching in the fifth semester of their training: While some students perform poorly in almost all items of this dimension, other students primarily have difficulties in the area of interaction or in relation to the integration of text- and learner-related perspectives. Still other students achieve the highest possible scores in almost all items. Accordingly, the results underline a claim already made by Tengberg et al. (2021, p. 775), according

to which a “one-size-fits-all model for professional development” is not suitable for teaching development. Rather, adaptive teaching-learning settings are needed that address both students with fundamental professionalization needs and advanced students who may only need further training in specific areas.

One limitation of the present study is that, although it depicts different realization variants of the individual teaching quality dimensions, it does not include any context variables. However, for the conceptualization and empirical testing of learning opportunities in initial and continuing teacher training, it is necessary to know what influences the quality of teaching. According to competence theory in teacher research (e.g., Blömeke et al., 2015; Krauss et al., 2020), professional knowledge, beliefs and situation-specific skills are the main factors that come into question here. In the context of the present study, it is possible to consider relevant influencing factors in follow-up analyses, as knowledge tests (Jähne et al., 2022; Winkler & Seeber, 2020) and data on professional vision (Hesse & Seeber, 2025) are available in addition to the teaching quality ratings. Findings in this area can help to develop tailor-made learning opportunities for specific groups of students.

In addition to explanatory approaches from competence theory-based teacher research, other explanations are also conceivable. With regard to the poor performance in subject-specific quality dimensions, it can be asked whether a lack of interest in reading and a lack of reading experience can also explain the observed deficits. In this regard, several existing studies show a connection between the reading behavior of the teacher and the quality of teaching (Applegate et al., 2014; Applegate & Applegate, 2004; Nathanson et al., 2008). In addition, the extent to which curricular influences affect (the quality of) observable teaching practices could be examined more closely with regard to contextual conditions. In this regard, Witte & Sâmihăian (2013), for example, can show that different curricula within Europe, but also different curricula within Germany, set different priorities in terms of learner or content orientation. It is possible that these also influence the focus of the observed lessons. Accordingly, international comparative studies on the teaching quality of prospective teachers would be beneficial in order to investigate the respective influences.

Another limitation concerns the generalizability of the available findings, as the sample consists of only 22 videos. Concerning this issue, it would be easy to call for the results to be validated on larger samples in the future. However, at least in Germany, the increasing shortage of teachers is making it more and more difficult to recruit study participants (Porsch & Reintjes, 2023). In this regard, it should be discussed whether complex videographic surveys are still feasible in the medium term or whether alternative survey methods should be used instead. For example, Magirius et al. (2024) are currently working on a tool that will enable students in the practicum to observe surface features, quality characteristics and objectives of literature lessons on a larger database. However, the extent to which such a tool proves to be reliable has yet to be demonstrated.

Finally, it should be noted that the present study focusses exclusively on the leaning opportunities of teaching and does not examine any correlations with learner data. This is problematic because there are currently hardly any studies that analyze the theoretically assumed impact of observed teaching quality on learning outcomes (in first language teaching). In addition, those studies that do analyze such correlations are hardly able to explain any variance in student performance by teaching characteristics (Praetorius et al., 2018; Tengberg et al., 2024). Accordingly, further research is also needed here to show that the quality criteria assumed in the present study are empirically relevant for the learning outcomes of the pupils and thus also relevant for professionalization of future teachers in teacher education programs.

7. CONCLUSION

This study set out to explore the teaching quality of student teachers in literature instruction, with a particular focus on both generic and subject-specific dimensions. The findings confirm that student teachers tend to perform well in generic aspects such as classroom management and socio-emotional support (RQ 1.1), which suggests that they are able to create a learning-conducive environment early in their professional development. However, subject-specific dimensions – particularly cognitive-emotional activation and the selection and realization of subject matter and methods – are less well developed and show considerable room for improvement (RQ 1.2). In addition, a high degree of heterogeneity was observed across the sample (RQ 2.1), and cluster analyses revealed qualitatively distinct patterns of how student teachers design literature lessons (RQ 2.2). These results underline the need for differentiated support structures in teacher education that address individual strengths and weaknesses. Simply familiarizing students with theoretical models of teaching quality is not sufficient. Instead, more intensive, practice-oriented formats – such as video-based reflection, peer feedback, or mentoring – may help student teachers to transfer theoretical knowledge into effective classroom practice.

Future research should further explore the factors that contribute to variation in subject-specific teaching quality among prospective teachers. In particular, studies could examine the role of professional knowledge, beliefs about teaching and learning, and personal reading experience in shaping instructional practices in literature. Moreover, international comparative studies may shed light on how curriculum frameworks and institutional settings influence how student teachers approach literature instruction. Finally, linking observed teaching quality to actual student learning outcomes would provide valuable evidence on the effectiveness of different teaching strategies and support the development of targeted professionalization efforts in teacher education.

REFERENCES

Allen, J. P., Pianta, R. C., Gregory, A., Mikami, A. Y., & Lun, J. (2011). An interaction-based approach to enhancing secondary school instruction and student achievement. *Science*, *333*(6045), 1034–1037. https://doi.org/10.1126/science.1207998

Applegate, A. J., & Applegate, M. D. (2004). The Peter Effect: Reading habits and attitudes of preservice teachers. *The Reading Teacher*, *57*(6), 554–563.

Applegate, A. J., Applegate, M. D., Mercantini, M. A., McGeehan, C. M., Cobb, J. B., DeBoy, J. R., Modla, V. B., & Lewinski, K. E. (2014). The Peter Effect Revisited: Reading Habits and Attitudes of College Students. *Literacy Research and Instruction*, *53*(3), 188–204. https://doi.org/10.1080/19388071.2014.898719

Bacher, J., Pöge, A., & Wenzig, K. (2010). *Clusteranalyse: Anwendungsorientierte Einführung in Klassifikationsverfahren* (3rd edn). Oldenbourg Verlag.

Backhaus, K., Erichson, B., Plinke, W., & Weiber, R. (2016). *Multivariate Analysemethoden: Eine anwendungsorientierte Einführung* (14th edn). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-46076-4

Baer, M., Guldimann, T., Kocher, M., & Wyss, C. (2014). Kognitive Aktivierung als Ausbildungsziel der Lehrer/innenbildung—Was zeigt der Blick in den Unterricht? In K.-H. Arnold, A. Gröschner, & T. Hascher (Eds), *Schulpraktika in der Lehrerbildung / Pedagogical field experiences in teacher education: Theoretische Grundlagen, Konzeptionen, Prozesse und Effekte / Theoretical foundations, programmes, processes, and effects* (pp. 183–200). Waxmann.

Berliner, D. C. (2005). The Near Impossibility of Testing for Teacher Quality. *Journal of Teacher Education*, *56*(3), 205–213. https://doi.org/10.1177/0022487105275904

Blömeke, S., Gustafsson, J.-E., & Shavelson, R. J. (2015). Beyond Dichotomies: Competence Viewed as a Continuum. *Zeitschrift Für Psychologie*, *223*(1), 3–13. https://doi.org/10.1027/2151-2604/a000194

Brophy, J. (2000). *Teaching*. International Bureau of Education.

Brunner, E. (2017). Qualität von Mathematikunterricht: Eine Frage der Perspektive. *Journal Für Mathematik-Didaktik*, *39*(2), 257–284. https://doi.org/10.1007/s13138-017-0122-z

Darling-Hammond, L. (2017). Teacher education around the world: What can we learn from international practice? *European Journal of Teacher Education*, *40*(3), 291–309. https://doi.org/10.1080/02619768.2017.1315399

de Jong, T., Lazonder, A. W., Chinn, C. A., Fischer, F., Gobert, J., Hmelo-Silver, C. E., Koedinger, K. R., Krajcik, J. S., Kyza, E. A., Linn, M. C., Pedaste, M., Scheiter, K., & Zacharia, Z. C. (2023). Let's talk evidence – The case for combining inquiry-based and direct instruction. *Educational Research Review*, *39*(100536), 1–10. https://doi.org/10.1016/j.edurev.2023.100536

Decristan, J., Hess, M., Holzberger, D., & Praetorius, A.-K. (2020). Oberflächen- und Tiefenmerkmale: Eine Reflexion zweier prominenter Begriffe der Unterrichtsforschung. *Zeitschrift Für Pädagogik*, *66*(66. Beiheft), 102–116. https://doi.org/10.25656/01:25867

Döring, N., & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (5., vollst. überarb., aktual. und erw. Aufl.). Springer.

Drollinger-Vetter, B. (2011). *Verstehenselemente und strukturelle Klarheit: Fachdidaktische Qualität der Anleitung von mathematischen Verstehensprozessen im Unterricht*. Waxmann.

Frederking, V., & Albrecht, C. (2016). Ästhetische Kommunikation im Literaturunterricht. Theoretische Modellierung und empirische Erforschung unter besonderer Berücksichtigung 'emotionaler Aktivierung'. In M. Krelle & W. Senn (Eds), *Qualitäten von Deutschunterricht* (pp. 57–81). Fillibach bei Klett.

Gabriel, K., & Lipowsky, F. (2013a). Hoch inferentes Rating: Klassenführung in Deutsch, Kunst und Mathematik. In M. Lotz, F. Lipowsky, & G. Faust (Eds), *Dokumentation der Erhebungsinstrumente des Projekts 'Persönlichkeits- und Lernentwicklung von Grundschulkindern' (PERLE): 3. Technischer Bericht zu den PERLE-Videostudien* (Vol. 23/3, pp. 145–168). GFPF.

Gabriel, K., & Lipowsky, F. (2013b). Hoch inferentes Rating: Unterrichtsklima in Deutsch, Kunst und Mathematik. In M. Lotz, F. Lipowsky, & G. Faust (Eds), *Dokumentation der Erhebungsinstrumente des Projekts 'Persönlichkeits- und Lernentwicklung von Grundschulkindern' (PERLE): 3. Technischer Bericht zu den PERLE-Videostudien* (Vol. 23/3, pp. 169–190). GFPF.

Gabrielsen, I. L., & Blikstad-Balas, M. (2020). Hvilken litteratur møter elevene i norskfaget?: En analyse av hvilke skjønnlitterære verk som inngår i178 norsktimer på åttende trinn. *Edda*, *107*(2), 85–99. https://doi.org/10.18261/issn.1500-1989-2020-02-02

Goldhaber, D., Krieg, J., Naito, N., & Theobald, R. (2021). Student Teaching and the Geography of Teacher Shortages. *Educational Researcher*, *50*(3), 165–175. https://doi.org/10.3102/0013189X20962099

Gore, J., Lloyd, A., Smith, M., Bowe, J., Ellis, H., & Lubans, D. (2017). Effects of professional development on the quality of teaching: Results from a randomised controlled trial of Quality Teaching Rounds. *Teaching and Teacher Education*, *68*, 99–113. https://doi.org/10.1016/j.tate.2017.08.007

Graham, L. J., White, S. L. J., Cologon, K., & Pianta, R. C. (2020). Do teachers' years of experience make a difference in the quality of teaching? *Teaching and Teacher Education*, *96*(103190), 1–10. https://doi.org/10.1016/j.tate.2020.103190

Greve, S., Elena Weber, K., Brandes, B., & Maier, J. (2021). What Do They Reflect on?—A Mixed-Methods Analysis of Physical Education Preservice Teachers' Written Reflections After a Long-Term Internship. *Journal of Teaching in Physical Education*, 1–11. https://doi.org/10.1123/jtpe.2021-0103

Gröschner, A., & Hascher, T. (2022). Praxisphasen in der Lehrerinnen- und Lehrerbildung. In M. Harring, C. Rohlfs, & M. Gläser-Zikuda (Eds), *Handbuch Schulpädagogik* (2nd edn, pp. 709–720). Waxmann.

Gröschner, A., Klaß, S., & Winkler, I. (2019). Lernbegleitung von Langzeitpraktika—Konzeption und Designelemente einer hochschuldidaktischen Intervention mittels Unterrichtsvideos. In J. Košinár, A. Gröschner, & U. Weyland (Eds), *Langzeitpraktika als Lernräume: Historische Bezüge, Konzeptionen und Forschungsbefunde* (Vol. 4, pp. 85–101). Waxmann.

Grossman, P., Cohen, J., & Brown, L. (2015). Understanding Instructional Quality in English Language Arts: Variations in PLATO Scores by Content and Context. In T. J. Kane, K. A. Kerr, & R. C. Pianta (Eds), *Designing Teacher Evaluation Systems* (1st edn, pp. 303–331). Wiley. https://doi.org/10.1002/9781119210856.ch10

Heins, J. (2018). Was sind typische Problemsituationen im Literaturunterricht? Ein Rahmenmodell zur Systematisierung von Unterrichtssituationen für die Entwicklung von Vignetten. *Didaktik Deutsch*, *23*(44), 27–43. https://doi.org/10.25656/01:22561

Helmke, A. (2014). *Unterrichtsqualität und Lehrerprofessionalität: Diagnose, Evaluation und Verbesserung des Unterrichts* (5th edn). Klett | Kallmeyer.

Hesse, F. (2024). *Qualitäten von Literaturunterricht. Eine Videostudie im Praxissemester*. Springer / J. B. Metzler.

Hesse, F., & Krause, J. (2024). Studierende unterrichten neben dem Studium als Vertretungslehrkräfte—Ein Bericht zum Stand der empirischen Forschung. *Didaktik Deutsch*, *56*, 66–86.

Hesse, F., & Seeber, A. (2025). Professionelle Unterrichtswahrnehmung im Praxissemester: Analysen von videobasierten Peer-Feedbacks zu Literaturstunden im Fach Deutsch. *Didaktik Deutsch*, *58*, 97–123. https://doi.org/10.21248/dideu.781

Hesse, F., & Winkler, I. (2022). Fachliche Qualität im Literaturunterricht. *SLLD-Z*, *2*, 1–29. https://doi.org/10.46586/SLLD.Z.2022.9605

Jähne, M., Seeber, A., & Winkler, I. (2022). Das Praxissemester als Gelegenheit zum Erwerb fachdidaktischer Kompetenzen?: Differenzielle Ergebnisse einer Interventionsstudie mit Deutschsstudierenden zur Wirksamkeit onlinebasierten Videofeedbacks. In J. Heins, K. Kleinschmidt-Schinke, D. Wieser, & E. Wiesner (Eds), *Üben. Theoretische und empirische Perspektiven in der Deutschdidaktik*. SLLD-B.

Junker, R., Gold, B., & Holodynski, M. (2021). Classroom management of pre-service and beginning teachers: From dispositions to performance. *International Journal of Modern Education Studies*, *5*(2), 339–363. https://doi.org/10.51383/ijonmes.2021.137

Korneck, F., Krüger, M., & Szogs, M. (2017). Professionswissen, Lehrerüberzeugungen und Unterrichtsqualität angehender Physiklehrkräfte unterschiedlicher Schulformen. In H. Fischler & E. Sumfleth (Eds), *Professionelle Kompetenz von Lehrkräften der Chemie und Physik* (pp. 113–134). Logos.

Kranefeld, U. (2021). Der Diskurs um Unterrichtsqualität in der Musikdidaktik zwischen generischen und fachspezifischen Dimensionen. *Unterrichtswissenschaft*, *49*(2), 221–233. https://doi.org/10.1007/s42010-021-00113-y

Krauss, S., Bruckmaier, G., Lindl, A., Hilbert, S., Birkner, K., Steib, N., & Blum, W. (2020). Competence as a continuum in the COACTIV study: The "cascade model". *ZDM Mathematics Education*, *52*(2), 311–327. https://doi.org/10.1007/s11858-020-01151-z

Lipowsky, F. (2020). Unterricht. In E. Wild & J. Möller (Eds), *Pädagogische Psychologie* (3., vollständig übearbeitete und aktualisierte Aufl., pp. 70–118). Springer.

Lipowsky, F., Drollinger-Vetter, B., Klieme, E., Pauli, C., & Reusser, K. (2018). Generische oder fachdidaktische Dimensionen von Unterrichtsqualität—Zwei Seiten einer Medaille? In M. Martens, K. Rabenstein, K. Bräu, M. Fetzer, H. Gresch, I. Hardy, & C. Schelle (Eds), *Konstruktionen von Fachlichkeit: Ansätze, Erträge und Diskussionen der empirischen Unterrichtsforschung* (pp. 183–202). Klinkhardt.

Lipowsky, F., Rakoczy, K., Pauli, C., Drollinger-Vetter, B., Klieme, E., & Reusser, K. (2009). Quality of geometry instruction and its short-term impact on students' understanding of the Pythagorean Theorem. *Learning and Instruction*, *19*(6), 527–537. https://doi.org/10.1016/j.learninstruc.2008.11.001

Lotz, M. (2016). *Kognitive Aktivierung im Leseunterricht der Grundschule: Eine Videostudie zur Gestaltung und Qualität von Leseübungen im ersten Schuljahr*. Springer VS. https://doi.org/10.1007/978-3-658-10436-8

Magirius, M., Scherf, D., & Hesse, F. (2024). *Literaturunterricht in verschiedenen Schularten. Ein Beobachtungstool zur Erfassung von Unterrichtsaktivitäten*. Symposion Deutschdidaktik 2024, Mainz. https://doi.org/10.13140/RG.2.2.12566.77129

Magnusson, C. G., Roe, A., & Blikstad-Balas, M. (2019). To What Extent and How Are Reading Comprehension Strategies Part of Language Arts Instruction? A Study of Lower Secondary Classrooms. *Reading Research Quarterly*, *54*(2), 187–212. https://doi.org/10.1002/rrq.231

Meyer, H. (2016). *Was ist guter Unterricht?* (11th edn). Cornelsen.

Nathanson, S., Pruslow, J., & Levitt, R. (2008). The Reading Habits and Literacy Attitudes of Inservice and Prospective Teachers: Results of a Questionnaire Survey. *Journal of Teacher Education*, *59*(4), 313–321.

Nissen, A. (2023). Cognitive Activation as an Aspect of Literature Instruction. *L1-Educational Studies in Language and Literature*, *23*, 1–20. https://doi.org/10.21248/l1esll.2023.23.1.447

Nissen, A., Tengberg, M., Svanbjörnsdóttir, B. M., Gabrielsen, I. L., & Blikstad-Balas, M. (2021). Function and use of literary texts in Nordic schools. *L1 Educational Studies in Language and Literature*, *21*, 1–22. https://doi.org/10.17239/L1ESLL-2021.21.02.10

Petko, D., Waldis, M., Pauli, C., & Reusser, K. (2003). Methodologische Überlegungen zur videogestützten Forschung in der Mathematikdidaktik: Ansätze der TIMSS 1999 Video Studie und ihrer schweizerischen Erweiterung. *ZDM*, *35*(6), 265–280.

Pianta, R. C., & Hamre, B. K. (2009). Conceptualization, Measurement, and Improvement of Classroom Processes: Standardized Observation Can Leverage Capacity. *Educational Researcher*, *38*(2), 109–119. https://doi.org/10.3102/0013189X09332374

Pianta, R. C., Hamre, B. K., & Mintz, S. (2012). *Classroom Assessment Scoring System: Secondary Manual*.

Porsch, R., & Reintjes, C. (2023). Teacher Shortages in Germany. Alternative Routes into the Teaching Profession as a Challenge for Schools and Teacher Education. In P. Hohaus & J.-F. Heeren (Eds), *The Future of Teacher Education: Innovations across Pedagogies, Technologies and Societies* (pp. 393–363). Brill. https://doi.org/10.1163/9789004678545

Praetorius, A.-K., & Charalambous, C. Y. (2018). Classroom observation frameworks for studying instructional quality: Looking back and looking forward. *ZDM*, *50*(3), 535–553. https://doi.org/10.1007/s11858-018-0946-0

Praetorius, A.-K., & Charalambous, C. Y. (Eds). (2023). *Theorizing Teaching: Current Status and Open Issues*. Springer. https://doi.org/10.1007/978-3-031-25613-4

Praetorius, A.-K., & Gräsel, C. (2021). Noch immer auf der Suche nach dem heiligen Gral: Wie generisch oder fachspezifisch sind Dimensionen der Unterrichtsqualität? *Unterrichtswissenschaft*, *49*(2), 167–188. https://doi.org/10.1007/s42010-021-00119-6

Praetorius, A.-K., Herrmann, C., Gerlach, E., Zülsdorf-Kersting, M., Heinitz, B., & Nehring, A. (2020). Unterrichtsqualität in den Fachdidaktiken im deutschsprachigen Raum – zwischen Generik und Fachspezifik. *Unterrichtswissenschaft*, *48*(3), 409–446. https://doi.org/10.1007/s42010-020-00082-8

Praetorius, A.-K., Klieme, E., Herbert, B., & Pinger, P. (2018). Generic dimensions of teaching quality: The German framework of Three Basic Dimensions. *ZDM Mathematics Education*, *50*(3), 407–426. https://doi.org/10.1007/s11858-018-0918-4

Praetorius, A.-K., Klieme, E., Kleickmann, T., Brunner, E., Lindmeier, A., Taut, S., & Charalambous, C. (2020). Towards Developing a Theory of Generic Teaching Quality: Origin, Current Status, and Necessary Next Steps Regarding the Three Basic Dimensions Model. *Zeitschrift Für Pädagogik*, *66. Beiheft*, 15–36. https://doi.org/10.25656/01:25861

Praetorius, A.-K., Rogh, W., & Kleickmann, T. (2020). Blinde Flecken des Modells der drei Basisdimensionen von Unterrichtsqualität? Das Modell im Spiegel einer internationalen Synthese von Merkmalen der Unterrichtsqualität. *Unterrichtswissenschaft*, *48*(3), 303–318. https://doi.org/10.1007/s42010-020-00072-w

Rakoczy, K., Wagner, E., & Frick, U. (2021). Wie in Mathe so auch in Kunst? Zur Konzeption von Unterrichtsqualität im Kunstunterricht. *Unterrichtswissenschaft*, 235–241. https://doi.org/10.1007/s42010-021-00104-z

Riegler, S., Wiprächtiger-Geppert, M., Kusche, D., & Schurig, M. (2022). Wie professionelles Wissen und gegenstandsbezogene Sachstruktur die Qualität von Rechtschreibunterricht beeinflussen. *SLLD*, *2*, 1–25. https://doi.org/10.46586/SLLD.Z.2022.9040

Schmidt, R. (2016). *Mit Kunstwerken zum Denken anregen: Eine empirische Untersuchung zur kognitiven Aktivierung im Rahmen der Kunstrezeption in der Grundschule*. kopaed.

Seidel, T., & Shavelson, R. J. (2007). Teaching Effectiveness Research in the Past Decade: The Role of Theory and Research Design in Disentangling Meta-Analysis Results. *Review of Educational Research*, *77*(4), 454–499. https://doi.org/10.3102/0034654307310317

Stahns, R. (2013). *Kognitive Aktivierung im Grammatikunterricht: Videoanalysen zum Deutschunterricht*. Schneider Verlag Hohengehren.

Szogs, M., Korneck, F., Krüger, M., Oettinghaus, L., & Kunter, M. (2016). Kognitive Aktivierung in standardisierten Unterrichtsminiaturen. In C. Maurer (Ed.), *Authentizität und Lernen – das Fach in der Fachdidaktik* (Vol. 36, pp. 605–607).

Tengberg, M., Blikstad-Balas, M., & Roe, A. (2022). Missed opportunities of text-based instruction: What characterizes learning of interpretation if strategies are not taught and students not challenged? *Teaching and Teacher Education*, *115*(103698), 1–13. https://doi.org/10.1016/j.tate.2022.103698

Tengberg, M., Skar, G. B., & Huebner, A. (2024). The impact of observable and perceived features of instruction on student achievement. *Teaching and Teacher Education*, *139*, 104457. https://doi.org/10.1016/j.tate.2023.104457

Tengberg, M., van Bommel, J., Nilsberth, M., Walkert, M., & Nissen, A. (2021). The Quality of Instruction in Swedish Lower Secondary Language Arts and Mathematics. *Scandinavian Journal of Educational Research*, *66*(5), 760–777. https://doi.org/10.1080/00313831.2021.1910564

Thiel, F., Ophardt, D., & Piwowar, V. (2013, January 1). *Abschlussbericht des Projekts Kompentenzen des Klassenmanagements (KODEK): Entwicklung und Evaluation eines Fortbildungsprogramms für Lehrkräfte zum Klassenmanagement*. https://www.ewi-psy.fu-berlin.de/einrichtungen/arbeitsbereiche/schulentwicklungsforschung/downloads/Abschlussbericht-KODEK.pdf

United Nations Educational, Scientific and Cultural Organization. (2024). *Global Report on Teachers: Addressing Teacher Shortages and Transforming the Profession*. United Nations. https://doi.org/10.18356/9789231006555

Vieluf, S., Praetorius, A.-K., Rakoczy, K., Kleinknecht, M., & Pietsch, M. (2020). Angebots-Nutzungs-Modelle der Wirkweise des Unterrichts: Ein kritischer Vergleich verschiedener Modellvarianten. *Zeitschrift Für Pädagogik*, *66*(66. Beiheft), 63–79.

Vogelsang, C. (2014). *Validierung eines Instruments zur Erfassung der professionellen Handlungskompetenz von (angehenden) Physiklehrkräften: Zusammenhangsanalysen zwischen Lehrerkompetenz und Lehrerperformanz*. Logos.

Wieser, D. (2019). Gegenwärtiger Stand der empirischen Unterrichtsforschung zum Literaturunterricht. In M. Kämper-van den Boogaart & K. H. Spinner (Eds), *Lese- und Literaturunterricht. Teil 2: Kompetenzen und Unterrichtsziele, Methoden und Unterrichtsmaterialien: Gegenwärtiger Stand der empirischen Unterrichtsforschung* (3., stark überarbeitete Aufl., Vol. 11/2, pp. 353–384). Schneider Verlag Hohengehren.

Winkler, I. (2020). Cognitive Activation in L1 Literature Classes: A content-specific framework for the description of teaching quality. *L1 – Educational Studies in Language and Literature*, *20*, 1–32. https://doi.org/10.17239/L1ESLL-2020.20.01.03

Winkler, I., & Seeber, A. (2020). Facetten literaturdidaktischer Kompetenz bei Deutschstudierenden vor und nach dem Praxissemester. Eine Interventionsstudie zur Wirksamkeit videobasierter Lernbegleitung. *Didaktik Deutsch*, *25*(49), 23–49. https://doi.org/10.25656/01:22286

Wiprächtiger, M., Stahns, R., Schurig, M., Riegler, S., & Harcsa, R. (2023). Fachspezifische Qualitätsprofile von Rechtschreibunterricht in Deutschland und der Schweiz. *Swiss Journal of Educational Research*, *45*(3), 228–241. https://doi.org/10.24452/sjer.45.3.1

Wirtz, M., & Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität*. Hogrefe.

Witte, T., & Sâmihăian, F. (2013). Is Europe open to a student-oriented framework for literature? A comparative analysis of the formal literature curriculum in six European countries. *L1 Educational Studies in Language and Literature*, *13*(1), 1–22. https://doi.org/10.17239/L1ESLL-2013.01.02

APPENDIX 1

Example item from the rating manual

SEL_2: Structuring the lesson
Main idea In order for literary learning and understanding to take place in classroom contexts, it is necessary that the learning tasks realized in class follow a clearly recognizable common thread and build on each other. In the best case scenario, the learning steps flow smoothly into one another so that there are no discontinuities in coherence. For the learners, the teacher also increases the coherence of the content by, for example, summarizing key discussion results in between/at transition points or explaining references between individual tasks.
Sources Praetorius et al. (2020, p. 311), Helmke (2014, pp. 190–200), Meyer (2016, pp. 25–38), Heins (2018, p. 34)
Indicators • The lesson follows clearly recognizable teaching/learning objectives. • The teacher sets tasks that build on each other. • The teacher explains to the learners how different tasks are related. to each other. • The teacher explains to the learners how the lesson fits into the lesson series. • The teacher creates plausible transitions between different learning steps in terms of content. • The teacher makes references to aspects already discussed in the lesson. • The teacher summarizes (interim) results of discussions or has them summarized by students.
Negative-Indicators • There is no recognisable common thread in the lesson. The sequence of tasks seems erratic and/or arbitrary. • At the transition from one task to the next, thematic breaks occur in the content. • Relevant aspects that were initially postponed are no longer addressed.

Rating levels
Level 4: A common thread is recognisable throughout the lesson. The tasks build on one another and flow smoothly into one another so that there are no breaks in coherence at any time. The teacher also creates coherence in the course of the lesson by referring back, making references or summarising.
Level 3: The lesson predominantly follows a common thread. The tasks generally build on each other well, with only minor breaks in coherence occurring in individual cases. The teacher uses numerous structuring impulses in the discussion. However, impulses are missing in some places where, for example, transitions are not understandable on their own or many discussion threads run side by side and need to be bundled.
Level 2: In the lessons observed, a common thread can only be reconstructed in phases. Although a few tasks build on each other, there is no recognisable connection with the other tasks. Consequently, the lessons are characterised by numerous breaks in content. This is also due to the fact that the teachers rarely make use of structuring options (summaries, transitions, references).
Level 1: No common thread can be reconstructed in the lessons observed. The tasks set have no recognisable connection; their arrangement seems arbitrary overall. This impression is also created by the fact that the teacher rarely or never uses impulses that make the content of the lesson transparent (e.g., through transitions, references, etc.).