

Florian Hesse

## **Praxisbezüge in der Unterrichtsqualitätsforschung**

### **Am Beispiel einer Videostudie zur Qualität von Literaturunterricht im Praxissemester**

#### **Abstract**

Die Unterrichtsqualitätsforschung erhebt den Anspruch, durch die Modellierung und Messung von Unterrichtsqualität einen Beitrag zur Verbesserung unterrichtlicher Praxis zu leisten. Zugespißt lautet die Grundannahme: Wenn empirisch nachgewiesen werden kann, dass bestimmte – auch aus theoretisch-normativer Sicht vielversprechende – Unterrichtsmerkmale die Lernerträge der Schüler\*innen positiv beeinflussen, dann können aus diesem Befund Empfehlungen hinsichtlich der Gestaltung von Unterricht abgeleitet werden. Der vorliegende Beitrag will diese These auf den Prüfstand stellen und dabei einerseits zeigen, dass ein Blick auf Qualitätsmerkmale von Unterricht durchaus für die Lehrer\*innenprofessionalisierung und Unterrichtsentwicklung ertragreich ist, insofern er Stärken und Schwächen aktueller Unterrichtspraxis differenziert und kriteriengeleitet sichtbar werden lässt. Andererseits soll deutlich gemacht werden, dass die Übersetzung entsprechender Ergebnisse in konkrete Handlungsempfehlungen in mehrererlei Hinsicht voraussetzungsreich und herausfordernd ist. Exemplarisch diskutiert werden in diesem Zusammenhang die Spannungsfelder (a) von Individualisierung und Standardisierung, (b) von fachspezifischen und generischen Zugängen zu Unterrichtsqualität sowie (c) von interner und externer Validität. Als Ausgangs- und Bezugspunkt der Überlegungen wird eine Untersuchung zum Literaturunterricht herangezogen, in der je eine Literaturstunde von  $N = 22$  Deutschstudierenden im Praxissemester videografiert und anschließend mittels kriteriengeleiteter Ratings hinsichtlich ihrer angebotsseitigen Qualität eingeschätzt wurde.

#### **Keywords**

Unterrichtsqualität, Videostudie, Literaturunterricht, Fachspezifik, Validität

#### **Abstract**

Teaching quality research claims to make an important contribution to improving teaching practice by modelling and measuring teaching quality. In a nutshell, the basic assumption is: If it can be empirically proven that certain teaching characteristics – which are also promising from a theoretical-normative point of view – positively influence the learning outcomes of the students, then recommendations regarding the design of teaching can be derived from this evidence. Examining this argument, this article aims to show that a look at the quality features of teaching is certainly useful for the professionalisation of teachers and the development of teaching, in that it makes the strengths and weaknesses of current teaching practice visible in a differentiated and criterion-guided way. On the other hand, it should be made clear that the translation of corresponding results into concrete recommendations for action is demanding and challenging in several respects. In this context, the tensions between (a) individualisation and standardisation, (b) subject-specific and generic approaches to teaching quality, and (c) internal and external validity will be discussed.

#### **Keywords**

Teaching Quality, Video Study, Literature Teaching, Subject Specificity, Validity

## 1 | Einleitung

Begrich et al. (2023) leiten einen jüngst erschienen Artikel zum aktuellen Stand und zur Zukunft der Unterrichtsqualitätsforschung folgendermaßen ein:

Eine hohe Qualität schulischen Unterrichts hat eine zentrale Funktion für die Entwicklung von Kindern und Jugendlichen und ist daher von hoher gesellschaftlicher und volkswirtschaftlicher Relevanz. Dementsprechend gibt es ein großes Interesse daran, die Qualität von Unterricht wissenschaftlich zu fassen und zu verbessern (Begrich et al., 2023, S. 64).

Für den vorliegenden Band ist diese Feststellung deshalb bedenkenswert, weil sie verdeutlicht, dass die Unterrichtsqualitätsforschung nicht nur das Ziel verfolgt, die Qualität von Unterricht zu rekonstruieren, sondern auch den Anspruch erhebt, durch die gewonnenen Erkenntnisse einen Beitrag zur Optimierung bestehender Unterrichtspraxis zu leisten. Ziel des vorliegenden Beitrags ist es, diesen Anspruch anhand einer Videostudie zur Qualität von Lernangeboten im Literaturunterricht angehender Deutschlehrpersonen einer Reflexion zu unterziehen. Dazu wird zunächst der Forschungsstand zur deutschdidaktischen Unterrichts(qualitäts)forschung kurz umrissen (Kap. 2), ehe die Fragestellung der Videostudie in diesem verortet wird (Kap. 3.1). Anschließend soll anhand einer exemplarischen Betrachtung von Befunden zur Qualitätsdimension der ‚kognitiv-emotionalen Aktivierung‘ einerseits gezeigt werden, dass die kriteriengeleitete Analyse unterrichtlicher Angebote in der Tat Einblicke ermöglicht, die für die Lehrpersonenprofessionalisierung und Unterrichtsentwicklung von Relevanz sind (Kap. 3.2). Andererseits wird anhand dreier Problempunkte diskutiert, warum konkrete Implikationen für die Ausbildungs-, Weiterbildungs- und Unterrichtspraxis – wenn überhaupt – nur mit Vorsicht abgeleitet werden können. Konkret geht es dabei um die Spannungsfelder (a) von Individualisierung und Standardisierung, (b) von fachspezifischen und generischen Zugängen zu Unterrichtsqualität sowie (c) von interner und externer Validität (Kap. 4). Aus den Überlegungen sollen abschließend Konsequenzen für zukünftige Studien im Bereich der deutschdidaktischen Unterrichtsqualitätsforschung gezogen werden (Kap. 5).

## 2 | Forschungsstand der deutschdidaktischen Unterrichts(qualitäts)forschung

Bemessen an der quantitativen wie qualitativen Bedeutung des Faches Deutsch im schulischen Fächerkanon ist die Anzahl von Untersuchungen zur Unterrichtspraxis des Deutschunterrichts insgesamt nach wie vor gering. Klammert man zudem Studien aus, die sich der qualitativen Rekonstruktion von Verstehensprozessen in Einzelstunden widmen oder im Rahmen von stark kontrollierten Interventionsdesigns Wirkungen bestimmter Lernsettings nachzuweisen versuchen, wird unmittelbar ersichtlich, dass systematische Einblicke in die Gestaltung und Qualität von ‚authentischem‘ (d. h. von Forschenden wenig beeinflusstem, ungescriptetem) Deutschunterricht auf breiter Datenbasis bislang äußerst rar sind (vgl. ausführlich Wieser, 2019). Dies muss eigentlich verwundern, insofern die Deutschdidaktik als Disziplin immer wieder einen (mehr oder weniger starken) Anspruch darauf erhebt, Deutschunterricht verbessern zu wollen (vgl. Meier & Emmersberger, i. d. B.), gleichzeitig aber über wenig systematisch gewonnenes Wissen verfügt, wie dieser Deutschunterricht eigentlich aussieht und welche Qualität Lernangebote im Deutschunterricht haben.

Angeregt durch Arbeiten der bildungswissenschaftlich-psychologischen Unterrichts(qualitäts)forschung haben sich in den letzten Jahren einige Forscher\*innen auf den Weg gemacht,

dieses Desiderat zu bearbeiten. In der Sprachdidaktik ist hier bspw. an die Arbeiten von Stahns (2013) oder Böhnert (2020) zum Grammtikunterricht oder von Riegler et al. (2020, 2022) zum Rechtschreibunterricht zu denken. Und auch in der Lese- und Literaturdidaktik liegen mittlerweile einige Arbeiten vor, die (vorwiegend konzeptionell) darum bemüht sind, Qualitätsaspekte des Unterrichts zu modellieren und z. T. auch empirisch zu erfassen (z. B. Depner et al., 2020; Hesse & Winkler, 2022; Lotz, 2016; Magirius et al., 2022; Schwahl, 2019; Winkler, 2017, 2020).

All diese Arbeiten orientieren sich mit Blick auf das Verständnis von Unterricht und Unterrichtsqualität an vier Grundannahmen, die sich auch im fächerübergreifenden Diskurs um Unterrichtsqualität (Begrich et al., 2023, S. 66–68) als konsensfähig erwiesen haben:

1. Unterricht ist dann ‚qualitätsvoll‘, wenn er theoretisch-normativ als ‚gut‘ befundene Unterrichtsmerkmale aufweist, die zugleich als ‚effektiv‘ für das Erreichen bestimmter Lernziele gelten (Berliner, 2005).
2. Unterricht lässt sich in Anlehnung an sog. Angebots-Nutzungs-Modelle als ein Zusammenspiel von Lernangeboten und deren individueller Nutzung durch die Schüler\*innen fassen, das zu unterschiedlichen Lernerträgen führt (Vieluf et al., 2020).
3. Unterricht kann auf einer Oberflächen- und einer Tiefenebene beschrieben werden. Entscheidend für die Zuordnung von Merkmalen zur einen oder anderen Ebene ist dabei die Frage, ob das betrachtete Merkmal ausgehend von einem konstruktivistischen (vs. behavioristischen) Lehr-Lern-Verständnis in unmittelbarem Zusammenhang mit dem Lernen der Schüler\*innen steht (z. B. kognitive Aktivierung) oder ob es sich um ein Gestaltungsmerkmal von Unterricht handelt (z. B. Sozialformen, Methoden), das das Lernen nicht direkt beeinflusst, sondern vielmehr einen Rahmen für die Initiierung bestimmter Lernprozesse schafft (Decristan et al., 2020; Hess & Lipowsky, 2020).<sup>1</sup> Aufgrund ihrer (mutmaßlich) höheren Lernwirksamkeit, stehen die Tiefenmerkmale des Unterrichts im Fokus der Forschung.
4. Für die Beschreibung der Tiefenebene des Unterrichts eignet sich grundsätzlich das in der deutschsprachigen Unterrichtsqualitätsforschung breit rezipierte Modell der drei Basisdimensionen von Unterrichtsqualität (Praetorius et al., 2018; Praetorius, Klieme, et al., 2020), welches die ‚kognitive Aktivierung‘, die ‚Klassenführung‘ und Aspekte des ‚Klassenklimas‘ für das Lernen zentral setzt. Konsens besteht allerdings auch dahingehend, dass das Modell insbesondere mit Blick auf die gegenstandsspezifische Dimension der kognitiven Aktivierung konkretisiert sowie um weitere Aspekte fachlicher Unterrichtsqualität (z. B. fachliche Korrektheit des Unterrichts, Gegenstandsauswahl) ergänzt werden muss.

<sup>1</sup> Ein Beispiel: Ob Schüler\*innen durch das Erstellen eines Standbildes einen literarischen Text (besser) verstehen, ist zunächst einmal weniger von der Methode als solcher abhängig als von der Art und Weise, wie die Methode eingesetzt wird. Zu fragen wäre beispielsweise, ob die Lehrperson eine für das Textverstehen zentrale (vs. randständige) Passage zugrundelegt oder ob die Schüler\*innen bei der Erstellung des Standbildes zu einer genauen Textlektüre angeregt werden (und nicht bloß aus einer vagen Erinnerung an die Lektüre heraus das Standbild erarbeiten).

### 3 | Literaturunterricht im Praktikum – exemplarische Einblicke in eine aktuelle Studie zur angebotsseitigen Qualität von Literaturunterricht

#### 3.1 | Fragestellung

Basierend auf den zuletzt genannten Annahmen und dem bislang übersichtlichen Studienbestand hat sich die nachfolgend näher betrachtete Studie *Literaturunterricht im Praktikum* (Hesse, 2024) mit der Frage beschäftigt, wie angehende Lehrpersonen des Faches Deutsch Lernangebote im Literaturunterricht auf Oberflächen- und Tiefenebene des Unterrichts gestalten. Dabei resultiert der Fokus auf angehende (vs. erfahrene) Lehrpersonen aus dem Umstand, dass in den letzten Jahren einerseits Praxisanteile in Lehramtsstudiengängen kontinuierlich erhöht wurden (Gröschner & Hascher, 2022) und andererseits in Zeiten des Lehrer\*innenmangels zunehmend die Möglichkeit besteht, bereits während des Studiums studienunabhängigen Unterrichtstätigkeiten (bspw. als Vertretungslehrkraft) nachzugehen (Bäuerlein et al., 2018; Scheidig & Holmeier, 2022). Wie angehende (Deutsch-)Lehrpersonen allerdings in solchen Settings Lernangebote im Unterricht gestalten und welche Qualität die Lernangebote aufweisen, ist bislang weitestgehend unerforscht. Zwar liegen in der Praxissemesterforschung zahlreiche Studien vor, die mittels Fragebogen die selbsteingeschätzte Kompetenz im Unterricht erheben. Konsens herrscht allerdings dahingehend, dass die so generierten Befunde eher als Selbstwirksamkeitserwartungen zu interpretieren sind und allenfalls in Ansätzen Rückschlüsse über das tatsächliche Verhalten im Unterricht zulassen (Cramer, 2010). Entsprechend wird zunehmend die Forderung nach Studien laut, die jenseits von Selbsteinschätzungen das Unterrichtshandeln mittels Videografie oder Tests in den Blick nehmen (vgl. Ulrich et al., 2020).

#### 3.2 | Qualitätsmodell zur Einschätzung von Lernangeboten im Literaturunterricht

Die Studie kam der gerade benannten Forderung nach, indem sie Literaturstunden im Jenaer Praxissemester u. a. hinsichtlich verschiedener Tiefenmerkmale eingeschätzt hat. Hierfür wurde ein Qualitätsmodell herangezogen (Abbildung 1), das sich an den bereits erwähnten drei Basisdimensionen von Unterrichtsqualität orientiert, diese jedoch für den Literaturunterricht konkretisiert und erweitert.

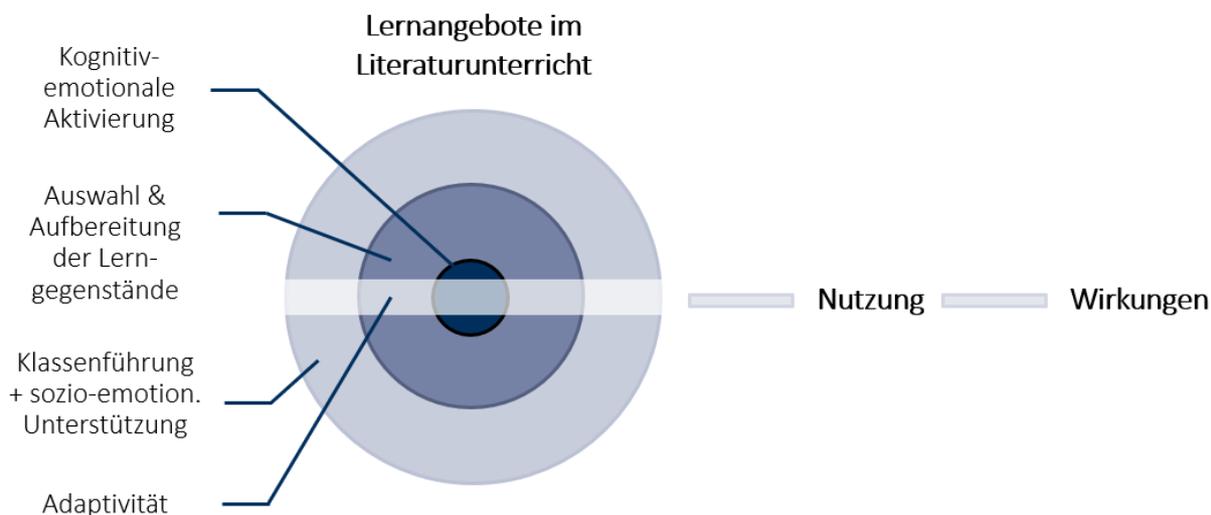


Abb. 1: Angebotsseitige Qualität von Literaturunterricht (nach Hesse & Winkler, 2022, S. 10)

Als Grundlage für die Ermöglichung fachlichen Lernens wurden aus den Basisdimensionen zunächst die beiden generisch angelegten Dimensionen ‚Klassenführung‘ und ‚sozio-emotionale Unterstützung‘ übernommen. Dieser Entscheidung lag die Annahme zugrunde, dass auch Literaturunterricht nur dort gelingen kann, wo Lehrpersonen für ein hohes Maß an effektiver Lernzeit sorgen und eine von Respekt, Interesse und Anerkennung geprägte Atmosphäre herstellen.

Übernommen wurde auch der Grundgedanke der Dimension der ‚kognitiven Aktivierung‘, da es auch im Literaturunterricht bedeutsam ist, dass die Schüler\*innen „zum vertieften Nachdenken und zu einer elaborierten Auseinandersetzung mit dem Unterrichtsgegenstand anregt“ werden (Lipowsky, 2020, S. 92). Gleichwohl ist die Beschäftigung mit Literatur nicht mit dem in der Bildungsforschung vielfach adressierten Wissenserwerb im Sinne eines vornehmlich kognitiv gedachten Informationsverarbeitungsprozesses gleichzusetzen. Zwar spielen fraglos auch beim Verstehen literarischer Texte kognitive Prozesse und das Vorwissen der Schüler\*innen eine zentrale Rolle (Christmann & Schreier, 2003). Der Umgang mit literarischen Texten, die in besonderer Weise Emotionen darstellen und intendieren, erschöpft sich aber nicht darin. Dies belegen unter anderem zahlreiche Rezeptionsstudien, die die Rolle von Emotionen oder auch damit verwandten Irritationen beim Lesen von Literatur hervorheben (siehe dazu z. B. die Beiträge in Bertschi-Kaufmann & Scherf, 2018; Lessing-Sattari & Freudenberg, 2020). Vor diesem Hintergrund wurde das Konstrukt der kognitiven Aktivierung um eine emotionale Facette hin zur ‚kognitiv-emotionalen Aktivierung‘ erweitert, wobei die Bindestrichschreibung bewusst gewählt ist, um zu verdeutlichen, dass Kognition und Emotion einander wechselseitig bedingen.

Gegenüber den drei Basisdimensionen neu hinzugefügt wurde schließlich die Dimension der ‚Auswahl und Aufbereitung der Lerngegenstände‘, die über die eher generischen Dimensionen der ‚Klassenführung‘ und der ‚sozio-emotionalen Unterstützung‘ hinausgehend erforderlich ist, damit die mit der ‚kognitiv-emotionalen Aktivierung‘ adressierte Auseinandersetzung von Schüler\*innen und Text auch fachlich grundiert wird. Ganz konkret werden in der Dimension beispielsweise Fragen von Textauswahl und -einsatz, der fachlichen Korrektheit bzw. Angemessenheit, der Strukturiertheit oder der Passung von Methoden und Sozialformen zu Lernziel, Gegenstand und Lernendengruppe in den Blick genommen. Das Qualitätsmodell knüpft damit sowohl an neuere konzeptionelle Entwicklungen der allgemeinen und auch mathematikdidaktischen Unterrichtsqualitätsforschung an, die die Bedeutung entsprechender Qualitätsmerkmale betonen (z. B. Brunner, 2017; Charalambous & Praetorius, 2020; Praetorius & Charalambous, 2018), als auch an Studien, die nachweisen können, dass sich eine Dimension „fachlicher Unterrichtsqualität“ von den anderen drei Unterrichtsqualitätsdimensionen empirisch trennen lässt (z. B. Lipowsky et al., 2018). Wichtig war bei der Konstruktion der neu hinzugefügten Dimension zudem, dass „typische Problemsituationen“ (Heins, 2018) von Literaturunterricht abgedeckt werden, die weder auf der Ebene von ‚Klassenführung‘ und ‚sozio-emotionaler Unterstützung‘ noch auf der Ebene der ‚kognitiv-emotionalen Aktivierung‘ liegen.

Aus den bisherigen Ausführungen sollte bereits deutlich geworden sein, dass das Qualitätsmodell einer hierarchischen Ordnung folgt, die über das abgebildete Kreismodell visualisiert wird. Angenommen wird also in Übereinstimmung mit vergleichbaren Vorschlägen aus Bildungswissenschaften und anderen Fachdidaktiken (Brunner, 2017; Charalambous & Praetorius, 2020; Praetorius, Herrmann et al., 2020), dass bestimmte generische und fachliche Qualitätsmerkmale zu einem gewissen Grad ausgeprägt sein müssen, damit sich die auf die Interaktion von

Text und Schüler\*innen gerichtete ‚kognitiv-emotionale Aktivierung‘ überhaupt entfalten und Textverstehen befördern kann. Zudem zeigt die Abbildung auch, dass alle Qualitätsratings so konzipiert wurden, dass die jeweils höchste Ausprägung dann erreicht wird, wenn die jeweiligen Merkmale *adaptiv* bezüglich Lernziel, Gegenstand und Lerngruppe realisiert wurden und nicht etwa dann, wenn ein bestimmtes Verhalten maximal häufig auftrat. Mit Blick auf die empirische Studie war es vor diesem Hintergrund interessant zu untersuchen, wie die Lernangebote der Studierenden hinsichtlich der unterschiedlichen Dimensionen einzuschätzen sind und inwiefern es innerhalb der untersuchten Stichprobe Unterschiede in der Realisierung der Qualitätsdimensionen und -merkmale gibt.

### 3.3 | Beispiel: Kognitiv-emotionale Aktivierung

Insofern der vorliegende Beitrag nicht den Raum bietet, die Operationalisierung und Befunde zu allen Qualitätsdimensionen vorzustellen, wird exemplarisch die Dimension der ‚kognitiv-emotionalen Aktivierung‘ thematisiert. Wie im vorherigen Abschnitt erläutert wurde, handelt es sich dabei um diejenige Qualitätsdimension, die den Kern von angebotsseitiger Unterrichtsqualität im Literaturunterricht ausmacht: Nur wenn die Schüler\*innen angeregt und dabei unterstützt werden, sich intensiv mit dem literarischen Text auseinanderzusetzen, wird Textverstehen überhaupt erst möglich.

Operationalisiert wurde die Dimension über insgesamt sieben Qualitätsmerkmale, die in Tabelle 1 in einem Vierfelderschema dargestellt werden.

	Anforderung bzw. Demand („D“)		Unterstützung bzw. Support („S“)		
<b>Aufgabenstellung bzw. Aufgabenset</b>	D1	Anregung genauer Textwahrnehmung	S1	Bereitstellung von aufgaben(set)immanentem Verstehenssupport	<b>Aufgabenstellung bzw. Aufgabenset</b>
	D2	Anregung subjektiver Involviertheit			
	D3	Anregung des Wechselspiels subjektiver Involviertheit und genauer Textwahrnehmung			
	D4	Anregung zur Reflexion des Verstehensprozesses			
<b>Aufgabenbearbeitung</b>	D5	Anregung zur Elaboration von Schüler:innen-Äußerungen	S2	Bereitstellung von Verstehenssupport im Bearbeitungsprozess	<b>Aufgabenbearbeitung</b>

Tab. 1: Item-Systematik ‚Kognitiv-emotionale Aktivierung‘

Um die Logik der Operationalisierung nachvollziehen zu können, ist der Blick zunächst auf den linken oberen Quadranten zu richten. Angenommen wird hier, dass für die Etablierung eines elaborierten Textverständnisses Aufgabenstellungen bzw. Aufgabensets vorliegen müssen, die die Schüler\*innen kognitiv-emotional zu genauer Textwahrnehmung, subjektiver Involviertheit, einem Wechselspiel der zuvor genannten Aspekte sowie einer Reflexion des Verstehensprozesses anregen. Der Fokus auf *diese* Merkmale lässt sich dabei sowohl konzeptionell als auch empirisch begründen: So weisen einschlägige Modellierungen literarischen Lernens und

Verstehens wiederkehrend darauf hin, dass die Verbindung text- und leser\*innenseitiger Perspektiven, von Immersion und Reflexion, von subjektiver Involviertheit und genauer Textwahrnehmung zentral für den Literaturunterricht seien (z. B. Spinner, 2006; Winkler, 2015, Zabka, 2015). Untersuchungen aus dem deutschsprachigen und internationalen Raum können zudem Belege dafür liefern, dass die Berücksichtigung und das Zusammenspiel ebendieser Merkmale tatsächlich die erhofften Wirkungen erzielt (vgl. Winkler, 2017, 2020, 2021).

Die Merkmale in den anderen drei Quadranten fragen danach, wie mit den Anforderungen umgegangen wird, die durch die Aufgaben an die Lernenden gestellt werden. Hier wird einerseits darauf geachtet, ob die Lehrperson auch im Zuge der Aufgabenbearbeitung und der damit verbundenen Interaktion zwischen Lehrperson und Schüler\*innen Impulse einsetzt, die mit Blick auf eines der zuvor genannten Items zu einer vertieften Auseinandersetzung anregen (D5). Ganz konkret ist hier beispielsweise an Nachfragen der Lehrperson zu denken, die Lernende dazu auffordern, eine Deutungshypothese zu explizieren, am Text zu falsifizieren oder mit Vorwissensbeständen und Erfahrungen in Beziehung zu setzen. Andererseits wird jenseits der aufgabenseitigen Anforderungen (*demand*) auch darauf geachtet, ob und inwiefern die Schüler\*innen aufgabenseitig (S1) oder in der Interaktion (S2) Unterstützung (*support*) erhalten. Hiermit knüpft die Studie an Befunde der literaturdidaktischen Aufgaben- und Gesprächsforschung an, die nahelegen, dass sich text- und lernendenangemessener Verstehenssupport in Abhängigkeit unterschiedlicher Voraussetzungen positiv auf das Textverstehen auswirkt (z. B. Harwart & Scherf, 2018; Heins, 2017; Magirius et al., 2022; Steinmetz, 2020).

Insgesamt betrachtet liegt der Studie also ein weites Verständnis kognitiv-emotionaler Aktivierung zugrunde, das die Initiierung und Unterstützung von Verstehensprozessen durch Aufgaben und deren Realisierung in der Interaktion umfasst. Dies wird – wie auch in anderen Studien (z. B. Hanisch, 2018; Lotz, 2016; Winkler 2017) – damit begründet, dass die einzelnen Merkmale konzeptionell sehr eng aufeinander bezogen sind. Inwiefern sich die unterschiedlichen Items tatsächlich einer Dimension ‚kognitiv-emotionaler Aktivierung‘ zuordnen lassen, ist zukünftig auf Basis eines größeren Datensatzes und unter Berücksichtigung der anderen Qualitätsdimensionen und -merkmale beispielsweise faktorenanalytisch zu prüfen. Ebenso beantwortet die Studie nicht die Frage, inwiefern die hier veranschlagten Merkmale eines kognitiv-emotional aktivierenden Unterrichts in Übereinstimmung mit den oben angedeuteten Befunden tatsächlich Textverstehensleistungen von Schüler\*innen vorhersagen können. Stattdessen liegt der Fokus zunächst auf einer differenzierten Einschätzung normativ erwünschter und potenziell verstehensrelevanter Qualitätsmerkmale des videografierten Lernangebots.

### 3.4 | Stichprobe und Rahmenbedingungen

Insgesamt wurden 22 Literaturstunden (je 45 bis 90 Min.) untersucht, die jeweils von einer\* einem Studierenden des Lehramts Deutsch für Gymnasium oder Regelschule ( $N = 22$ ) im Rahmen des Jenaer Praxissemesters gehalten wurden.<sup>2</sup> Letzteres wird von den Studierenden i. d. R. im fünften oder sechsten Semester absolviert. Die Studierenden verfügten folglich zum Erhebungszeitpunkt über fachwissenschaftliche und fachdidaktische Grundkenntnisse, die sie im Rahmen entsprechender Einführungsveranstaltungen erworben haben. Zudem nahmen alle Studierenden an deutschdidaktischen und bildungswissenschaftlichen Vorbereitungs- und Begleitveranstaltungen zum Praxissemester teil, die den lernförderlichen Umgang mit Schüler\*innenantworten und damit zusammenhängende Fragen von Unterrichtsqualität thematisierten. In der deutschdidaktischen Vorbereitungsveranstaltung wurde darüber hinaus exemplarisch besprochen, wie Unterrichtsstunden zu epischen Kurztexten geplant werden können, wobei Fragen der didaktischen Analyse literarischer Texte (Zabka, 2012) ebenso besprochen wurden wie das Stellen verstehensförderlicher Lernaufgaben (Heins, 2017; Winkler, 2011). Zwar dürfen diese Veranstaltungsinhalte nicht überbewertet werden – die Vorbereitungsveranstaltung dauerte lediglich einen Tag, die Begleitveranstaltungen fanden nur 14-tägig mit zwei Semesterwochenstunden statt –, allerdings kann zumindest festgestellt werden, dass die Studierenden mit einigen der Qualitätsdimensionen oder -merkmalen in Ansätzen vertraut waren.

Die Unterrichtsstunden als solche wurden im Wintersemester 2018/19 und im Sommersemester 2019 aufgezeichnet. Nach einer Einweisung in das Kameraequipment und dem Einholen notwendiger Einverständniserklärungen von Schüler\*innen, Eltern, Schulleitung und Ministerium videografierten die Studierenden die Unterrichtsstunden eigenverantwortlich und durften auch selbst entscheiden, welche Stunde in welcher Klasse videografiert werden sollte. Zwei relativ lose Vorgaben vonseiten der Forschenden bestanden lediglich darin, dass die Studierenden einen epischen Kurztext behandeln und darüber hinaus mindestens eine Phase des Klassengesprächs einplanen sollten. Erstere Vorgabe lässt sich damit begründen, dass Vergleichbarkeit zwischen den Stunden in Bezug auf einen Gegenstand hergestellt werden sollte, der (a) ohnehin in der Schule eingesetzt wird (vgl. zur Bedeutung von Kurzprosa im Literaturunterricht z. B. Siebenhüner et al., 2019a, 2019b), (b) für alle Klassenstufen und Schulformen gleichermaßen relevant ist und sich (c) im Rahmen einer Einzelstunde realisieren ließ. Die zweite Vorgabe resultierte aus dem Schwerpunkt der Begleitveranstaltungen, die – wie oben angesprochen – den lernförderlichen Umgang mit Schüler\*innenantworten im Klassengespräch fokussierten. Als lose sind diese Richtlinien deshalb zu bezeichnen, weil die Studierenden den konkreten Text frei wählen und die Plenumsphase beliebig lang ausdehnen sowie nach eigenen Vorstellungen im Unterricht positionieren konnten.

<sup>2</sup> Die Selbstvideografie war verpflichtender Bestandteil der Begleitveranstaltung im Praxissemester, sodass anders als bei sonstigen Videografieprojekten nicht von vornherein von einer Positivselektion der teilnehmenden Lehrkräfte auszugehen ist. Gleichwohl muss einschränkend eingeräumt werden, dass von ursprünglich 38 teilnehmenden Studierenden bei der Videoerhebung nur 22 Studierende mit ihren Stunden einbezogen werden konnten. Dies hatte vielfältige Ursachen (z. B. mangelnde Bild- und/oder Tonqualität, Datenverluste, Wahl eines anderen Gegenstandes), von denen eine auch darin bestand, dass die Studierenden einer Auswertung der Videos zu Forschungszwecken nicht zustimmten.

### 3.5 | Methodisches Vorgehen

Um die 22 Unterrichtsstunden hinsichtlich ihrer angebotsseitigen Qualität einschätzen zu können, waren drei sequenziell aufeinanderfolgende methodische Schritte notwendig:

1. *Datenaufbereitung*: Alle Videografien wurden zunächst durch den Autor in ein einheitliches Dateiformat konvertiert, nach den Konventionen von GAT-2 (Selting et al., 2009) in Form von Basistranskripten verschriftlicht und um Sitzpläne angereichert. Zudem wurden mit Blick auf die späteren Ratings kriteriengeleitete Basiskodierungen vorgenommen, die etwa das Festlegen von Unterrichtsbeginn und -ende, die Extraktion der im Unterricht gestellten Aufgaben sowie die Erstellung von Stundenübersichten bei jedem Video betrafen. Das Anliegen dieser Phase bestand also darin, einen zielgerichteten Zugriff auf die Daten zu ermöglichen. Dabei ist jedoch zu betonen, dass die Transkriptionen, Verlaufspläne usw. die Arbeit am Video nicht ersetzen, sondern *ergänzen*.
2. *Primäranalyse (Rating)*: Im Rahmen der Primäranalyse der Videodaten wurden die Lernangebote der Unterrichtsstunde entlang der im vorherigen Kapitel vorgestellten Qualitätsdimensionen hoch-inferent geratet (vgl. zu Ratings in der Unterrichtsforschung grundsätzlich Kleinbub, 2018; Pauli, 2014). Hierzu war es im Vorfeld nötig, die einzelnen Qualitätsdimensionen zu operationalisieren, indem für jede Dimension Qualitätsmerkmale und zugehörige Verhaltensindikatoren theoriegeleitet und in Anknüpfung an bestehende Untersuchungen bestimmt wurden (s. dazu beispielhaft die Operationalisierung zur ‚kognitiv-emotionalen Aktivierung‘ unten). Die Merkmalsbeschreibungen und Indikatoren wurden sodann in einem Ratingmanual zusammengefasst, das nach einer Pilotierung anhand von Unterrichtsvideos aus dem Projekt KoAla (Winkler & Steinmetz, 2016) sowohl als Grundlage für die Einschätzung der Angebote als auch für die Schulung von zwei weiteren Rater\*innen fungierte (fortgeschrittene Lehramtsstudierende des Faches Deutsch). Im Anschluss an die Schulung wurden die Stunden vom Autor und den beiden anderen Rater\*innen schrittweise unabhängig voneinander auf vierstufigen Skalen eingeschätzt (1 = geringe Ausprägung/Adaptivität; 4 = hohe Ausprägung/Adaptivität) und die Urteile mittels Intraklassenkorrelationskoeffizient hinsichtlich ihrer Reliabilität überprüft (Wirtz & Caspar, 2002).
3. *Sekundäranalyse (statistische Auswertung der Ratings)*: Die ermittelten Ratingscores wurden im Anschluss einer deskriptiv-statistischen Auswertung (Mittelwerte, Standardabweichung, Spannweite) unterzogen und mittels explorativer Methoden untersucht. Konkret wurden hierarchische agglomerative Clusteranalysen durchgeführt (Backhaus et al., 2016; Schendera, 2010), um herauszufinden, ob es innerhalb der Qualitätsdimensionen Studierendengruppen gibt, die sich hinsichtlich der Realisierung bestimmter Qualitätsmerkmale ähneln.

### 3.6 | Ergebnisse

Im Folgenden werden die Ergebnisse zur ‚kognitiv-emotionalen Aktivierung‘ zunächst auf dimensionaler Ebene betrachtet, ehe sich eine Detailanalyse der Items und der ermittelten Cluster anschließt.

Betrachtet man also die Ratings zur ‚kognitiv-emotionalen Aktivierung‘ im Vergleich mit den anderen Qualitätsdimensionen, fällt auf, dass diese Dimension durchschnittlich am schlechtesten eingeschätzt wurde (Tabelle 2).

Dimension	Anzahl Items	<i>N</i>	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>
Kognitiv-emotionale Aktivierung	7	22	2.32	.54	1.29	3.29
Auswahl und Aufbereitung ...	4	22	2.60	.65	1.50	3.75
Klassenführung	8	22	3.35	.63	2.13	4.00
sozio-emotionale Unterstützung	4	22	3.48	.48	2.50	4.00

**Tab. 2:** Skalenwerte der untersuchten Qualitätsdimensionen im Vergleich

Deutschstudierenden im Praxissemester gelingt es also offensichtlich sehr gut, angebotsseitig eher generische Dimensionen von Unterrichtsqualität zu realisieren, während v. a. im Bereich der ‚kognitiv-emotionalen Aktivierung‘, aber auch in der fach- bzw. gegenstandsbezogenen Dimension ‚Auswahl und Aufbereitung von Lerngegenständen‘ größere Qualitätsdefizite deutlich werden.

Ein genauerer Blick auf die einzelnen Merkmale der ‚kognitiv-emotionalen Aktivierung‘ veranschaulicht darüber hinaus, dass die verschiedenen Facetten des Konstrukts in ihrer Ausprägung z. T. sehr unterschiedlich bewertet wurden (Tabelle 3).

Items	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>
D1 Anregung genauer Textwahrnehmung	2.91	.97	1	4
D2 Anregung subjektiver Involviertheit	2.64	.95	1	4
D3 Anregung des Wechselspiels	1.95	.79	1	4
D4 Anregung zur Reflexion des Verstehensprozesses	1.55	.67	1	3
D5 Anregung zur Elaboration von SuS-Antworten	2.55	.96	1	4
S1 Support in Lernaufgaben	2.41	.80	1	4
S2 Support in der Interaktion	2.77	.77	1	4

**Tab. 3:** Itemwerte ‚kognitiv-emotionale Aktivierung‘

So wird beispielsweise deutlich, dass die Unterrichtsstunden angebotsseitig im Mittel eher auf den Text (D1) denn auf die Lernenden zielen (D2) und text- und lernendenbezogene Perspektiven allenfalls in Ansätzen aufeinander bezogen werden (D3). Auffällig ist darüber hinaus, dass der Verstehensprozess in den Unterrichtsstunden kaum nennenswert reflektiert wird (D4) und dass die beiden interaktionsbezogenen Items (D5, S2) trotz intensiver Bearbeitung in den Vorbereitungs- und Begleitveranstaltungen aus fachdidaktischer und bildungswissenschaftlicher Perspektive lediglich leicht über dem theoretischen Skalenmittelwert liegen. Der Support in den Lernaufgaben selbst (S1) liegt sogar leicht darunter.

Jenseits der Mittelwertsbetrachtungen sticht sowohl in Tabelle 2 als auch und v. a. in Tabelle 3 eine zum Teil erhebliche Streuung der Qualitätseinschätzungen ins Auge. Abgesehen vom Item D4, das über beinahe alle Stunden hinweg niedrig ausgeprägt ist und das theoretische Maximum von 4 nicht erreicht, wurden mit Blick auf die übrigen Items alle Werte vergeben. Zudem deuten auch die Standardabweichungen auf große Varianz in den Daten hin. Wie bereits im vorherigen Abschnitt erläutert, wurde dieser Befund zum Anlass genommen, um clusteranalytisch zu prüfen, ob sich in den Daten bestimmte Verteilungsmuster bezogen auf die Realisierung von Qualitätsmerkmalen in den einzelnen Dimensionen finden lassen.

Im Ergebnis ließen sich mittels Elbow-Kriterium und Test von Mojena (vgl. Backhaus et al., 2016; S. 495–497) vier Cluster kognitiv-emotional aktivierender Unterrichtsangebote bestimmen, die in Tabelle 4 dargestellt sind. Die Mittelwerte beziehen sich jeweils auf die mittleren Ausprägungen der Merkmale innerhalb eines Clusters; die zugehörigen t-Werte zeigen an, ob der Mittelwert im Vergleich zur Erhebungsgesamtheit überrepräsentiert (positive Ausprägung) oder unterrepräsentiert (negative Ausprägung) ist. Werte größer  $|0.5|$  zeigen dabei stärkere Abweichungen von der Erhebungsgesamtheit an und wurden farbig hervorgehoben.

Variable	Cluster 1 n = 3		Cluster 2 n = 10		Cluster 3 n = 4		Cluster 4 n = 5	
	M1	t1	M2	t2	M3	t3	M4	t4
D1 Textwahrnehmung	1.33	-1.62	2.80	-0.11	3.00	0.09	4.00	1.12
D2 subjektive Involviertheit	2.33	-0.32	2.60	-0.04	3.75	1.16	2.00	-0.67
D3 Wechselspiel	1.00	-1.21	2.00	0.06	3.00	1.34	1.60	-0.45
D5 Elaboration SuS-Antworten	2.00	-0.57	1.90	-0.67	3.50	0.99	3.40	0.88
S1 Support in Lernaufgaben	1.00	-1.77	2.30	-0.14	3.25	1.06	2.80	0.49
S2 Support in der Interaktion	2.00	-0.35	1.90	-0.48	2.50	0.30	3.00	0.95

Tab. 4: Clusteranalyse ‚kognitiv-emotionale Aktivierung‘

Nachfolgend werden die Cluster knapp beschrieben<sup>3</sup>, ehe die Befunde im nächsten Kapitel zusammenfassend diskutiert und eingeordnet werden:

<sup>3</sup> Eine ausführliche Beschreibung der Cluster inklusive einer Veranschaulichung anhand von Beispielstunden und Transkriptionsauszügen findet sich in Hesse (2024).

- Das erste Cluster umfasst drei Unterrichtsstunden mit *wenig kognitiv-emotional aktivierenden Lernangeboten*, insofern hier alle Einschätzungen sowohl kriterial (Mittelwerte) als auch in Bezug zur Erhebungsgesamtheit (t-Werte) niedrig ausgeprägt bzw. unterrepräsentiert sind. Besonders kennzeichnend für diese Stunden ist der Umstand, dass der literarische Text kaum präsent ist und somit auch keinerlei Wechselspiel zwischen Textwahrnehmung und subjektiver Involviertheit möglich ist.
- Das zweite Cluster beschreibt einen *durchschnittlich kognitiv-emotional aktivierenden Unterricht mit Schwächen in der Interaktion*. Studierenden in diesem Cluster gelingt es deutlich besser, Angebote zu schaffen, die den Text in den Blick nehmen, und auch das Wechselspiel zwischen subjektiver Involviertheit und Textwahrnehmung wird zumindest ansatzweise initiiert. Keine Verbesserung im Vergleich zum vorhergehenden Cluster besteht allerdings hinsichtlich interaktionsbezogenen Items, die in diesem Cluster die niedrigsten Werte erreichen. Deutlich wird dies beispielsweise in Unterrichtssituationen, in denen Studierende mittels aneinandergereihter Adding-on-Fragen (z. B. Will noch jemand etwas dazu sagen?) fortlaufend nach weiteren Deutungshypothesen der Schüler\*innen fragen, ohne dass klar wäre, welche weiteren Hypothesen (jenseits der bereits genannten) die Lernenden eigentlich anführen könnten.
- Das dritte Cluster kann im Vergleich zu den beiden vorherigen Clustern als *stark kognitiv-emotional aktivierender Unterricht* bezeichnet werden, insofern sowohl Mittel- als auch t-Werte durchweg hoch ausschlagen. Hervorzuheben sind dabei insbesondere die vergleichsweise hohen Werte bei der Anregung subjektiver Involviertheit und dem Wechselspiel von Involviertheit und Textwahrnehmung. Lediglich im Bereich der Anregung genauer Textwahrnehmung und der Bereitstellung von Support in der Interaktion schneidet das Cluster nur durchschnittlich ab. Dabei ist eine mittlere Ausprägung im Bereich der genauen Textwahrnehmung dadurch gekennzeichnet, dass die Studierenden mehrheitlich inhalts- oder darstellungsbezogene Aspekte des Textes fokussieren und diese nicht oder nur selten aufeinander beziehen.
- Im vierten Cluster kommen die Items, die im dritten Cluster weniger gut realisiert werden, mit besonders hoher Ausprägung vor. Die Anregung zu genauer Textwahrnehmung erreicht hier Maximalwerte (inhalts- und darstellungsbezogene Aspekte werden hier also gleichermaßen adressiert und aufeinander bezogen) und auch die interaktionsbezogenen Items sind sowohl kriterial als auch im Vergleich zur Erhebungsgesamtheit durchweg hoch ausgeprägt. Dabei zeichnen sich Unterrichtsstunden, die im interaktionsbezogenen Bereich hohe Werte erzielen, u. a. dadurch aus, dass es den Studierenden gelingt, die Beiträge der Lernenden produktiv aufeinander zu beziehen oder an Schlüsselstellen des Unterrichtsgesprächs durch Nachfragen oder pointierte Zusammenfassungen zu einer vertieften Auseinandersetzung mit dem Text anzuregen. Vor diesem Hintergrund kann das Cluster auch mit dem Label *gegenstands- und interaktionsbezogen aktivierender Unterricht* versehen werden. Schwachstellen ergeben sich in diesem Cluster lediglich daraus, dass die Unterrichtsstunden im Clustervergleich den niedrigsten Wert im Bereich der Anregung subjektiver Involviertheit aufweisen und auch beim Wechselspiel zwischen Involviertheit und Textwahrnehmung den zweitnied-

rigsten Wert erreichen. Auf das Vorwissen der Lernenden, ihre textbezogenen Emotionen und Wertungen richten sich die von den Lehrpersonen gestellten Aufgaben also nicht oder nur in geringem Maße.

### 3.7 | Einordnung in den Forschungsstand

Die Befunde der Untersuchung sind insgesamt gut an bereits durchgeführte Studien im Bereich der allgemein- und literaturdidaktischen Unterrichts- und Professionalisierungsforschung anchlussfähig.

- So ist etwa das Ergebnis, dass fachliche Qualitätsdimensionen in Ratings geringe Werte erzielen als generische, in bisherigen Studien wiederkehrend zu beobachten (siehe auch Lotz, 2016; Schmidt, 2016). Ursächlich für die beinahe exzellenten Werte in den generischen Dimensionen könnte die Anlage der Ratingsysteme (White & Klette, 2023), aber auch eine Positivselektion der videografierten Lehrkräfte und/oder Klassen sein. Speziell im Fall angehender Lehrpersonen ist zudem davon auszugehen, dass diese von bereits mentor\*innenseitig etablierten Klassenregeln und Beziehungen profitieren (Junker et al., 2021). Mitunter lassen sich auch verzerrende Beobachtungseffekte durch die Kamera nicht ausschließen, auch wenn die bisherige Forschung zu diesem Thema mahnt, Kameraeffekte nicht überzubewerten (Maak & Ricart Brede, 2014; Petko et al., 2003).
- Dass die Studierenden speziell bei der ‚kognitiv-emotionalen Aktivierung‘ Probleme haben, plausibilisieren auch bisherige Befunde aus der Forschung zum Wissen und zu situationsspezifischen Fähigkeiten von Lehrpersonen – also zwei Konstrukten, denen insbesondere im Expertise- bzw. kompetenzorientierten Ansatz der Lehrer\*innenforschung eine große Bedeutung für die Unterrichtsgestaltung zugeschrieben wird (Krauss, 2020). So können etwa die Studien von Anselm (2011), Pissarek & Schilcher (2017) sowie Winkler & Seeber (2020) zeigen, dass Studierende in Testsituationen z. T. große Probleme haben, Potenziale literarischer Texte zu erkennen oder für Schüler\*innen erstellte Lernaufgaben literarischer Texte selbst zu lösen. Zudem weisen beispielsweise die Studien von Heins (z. B. 2019) zur professionellen Unterrichtswahrnehmung in Übereinstimmung mit früheren Stufenmodellen der Lehrer\*innenprofessionalisierung (z. B. Berliner, 2004; Fuller & Bown, 1975) darauf hin, dass Studierende sich damit schwertun, die Perspektiven und Verstehenswege der Schüler\*innen zu erkennen und einzubeziehen. Entsprechend verwundert nicht, dass in der vorliegenden Studie sowohl gegenstands- als auch interaktionsbezogene Unterrichtsmerkmale vielfach nicht in hoher Qualität realisiert wurden.

Dass die generierten Befunde bestehende Forschungslinien fortschreiben, ist insofern erfreulich, als die empirische Basis für Handlungsimpulse oder gar -empfehlungen in Richtung der hochschulischen und schulischen Praxis damit gestärkt wird. Gerade im Bereich der Praktikumsforschung stellt die Studie diesbezüglich ein wichtiges Mosaikstück dar, insofern bisherige Studien zu unterrichtsbezogenen Kompetenzen oder zur Unterrichtsqualität im Praktikum überwiegend auf Selbsteinschätzungen der Studierenden beruhen, die sehr anfällig für selbstwertdienliche Verzerrungen sind (Gröschner & Hascher, 2022; Ulrich et al., 2020).

## 4 | Welche Praxisbezüge ergeben sich aus den Ergebnissen?

Obleich sich die vorliegenden Befunde gut an bestehende Erkenntnisse andocken lassen, stellt sich generell und speziell unter Berücksichtigung der thematischen Ausrichtung des vorliegenden Sammelbandes die Frage, inwiefern sich der einleitend thematisierte Anspruch auf Praxisrelevanz, den die Unterrichtsqualitätsforschung für sich reklamiert, durch die hier vorgestellten Ergebnisse einlösen lässt.

Zu konkretisieren ist dabei zunächst, von welcher Praxis eigentlich die Rede ist. Ertragreich erscheint in diesem Zusammenhang eine grundsätzliche Unterscheidung von Carl & Scherf (2022), die mit Blick auf die Ableitungen aus empirischen Studien dafür plädieren, das Handlungsfeld Schule und das Handlungsfeld Wissenschaft auseinanderzuhalten. Bezieht man sich auf ersteres, rückt die Frage ins Zentrum, inwiefern sich aus der Studie Impulse oder Empfehlungen für die Aus- und Fortbildung von (angehenden) Lehrpersonen ableiten lassen. Bezieht man sich auf zweiteres, geht es etwa um Konsequenzen einer Untersuchung für die Theoriebildung, die Konstruktmodellierung oder -operationalisierung. Aus Platzgründen sollen nachfolgend vor allem Ableitungen für das Handlungsfeld Schule im Mittelpunkt stehen, insofern diese im vorliegenden Sammelband über das „Kaskadenmodell“ eine besondere Gewichtung erfahren (s. Meier & Emmersberger i. d. B.). Drei Spannungsfeldern soll dabei ausgehend von den Befunden der oben dargestellten Studie besonderes Gewicht zukommen:

- a) dem Spannungsfeld von Individualisierung und Standardisierung,
- b) dem Spannungsfeld von fachspezifischen und generischen Zugängen,
- c) dem Spannungsfeld von interner und externer Validität.

### 4.1 | Spannungsfeld 1: Individualisierung vs. Standardisierung

Sowohl die Standardabweichungen als auch die Clusteranalysen verdeutlichen eindrücklich, dass Studierende im Praktikum Lernangebote nicht durchweg auf demselben Qualitätsniveau gestalten. Vielmehr zeigt sich eine große Vielfalt an Gestaltungsvarianten, die von eher qualitätsarmem bis hin zu unterschiedlichen Formen qualitätsvollen Literaturunterrichts reicht. Tengberg et al. (2022) kommen in einer großangelegten Videostudie in Schweden mit 141 videografierten und analysierten Language Arts-Stunden erfahrener Lehrpersonen zu einem vergleichbaren Ergebnis und halten fest:

The most important finding of the study is perhaps that teachers in different classrooms vary substantially and systematically on high-inference variables known from previous research to be vital for supporting student achievement gains over time. [...] This offers a key implication for future planning of instructional improvement and professional development initiatives. It suggests that a one-size-fits-all model for professional development—for instance, that all teachers ought to increase the amount of explicit strategy instruction, develop their capacity for intellectually challenging dialogue, or more carefully mind the time-on task ratio of their lessons—may be a less fruitful design because it does not respond to the individual development needs of teachers. (Tengberg et al. 2022, S. 765)

Tengberg et al. (2022) plädieren also für eine stärker individualisierte Aus- und Weiterbildung, wengleich sie weiterführend betonen, dass nicht für jede einzelne Lehrperson neue Programme aufgesetzt werden müssen. Vielmehr gehe es darum, die Fortbildungen an einschlägigen Entwicklungsverläufen auszurichten. Um solche Entwicklungsverläufe zu identifizieren,

seien wiederum Studien notwendig, die auf breiter Datenbasis Einblicke in die Unterrichtspraxis ermöglichen.

Die clusteranalytisch gewonnenen Befunde der vorliegenden Studie bieten zwar keine Einblicke in die längsschnittliche Entwicklung angebotsseitiger Unterrichtsqualität. Allerdings zeigen sie querschnittlich Stärken und Schwächen unterschiedlicher Gestaltungsweisen auf, die bei der Konzeption von Aus- und Weiterbildungssettings im Sinne von „Anstößen zum Überdenken profess. Lehrkonstruktionen“ (Carl & Scherf, 2022, S. 13) Berücksichtigung finden sollten. So gibt es einige (wenige) Studierende, deren Unterricht auf allen Ebenen professionalisierungsbedürftig ist (Cluster 1), während andere Studierende vorrangig im Bereich der Interaktion (Cluster 2) oder hinsichtlich der Berücksichtigung von Lernendenperspektiven (Cluster 4) Entwicklungspotenziale aufweisen. Zudem gibt es eine Gruppe von Studierenden, die bereits in nahezu allen Bereichen souverän agiert (Cluster 3), sodass sich hier vordergründig die Frage stellt, wie man das Qualitätsniveau halten und im oberen Bereich punktuell ausbauen kann. Über die Clusteranalyse gelingt es folglich, eine mittlere Ebene einzunehmen, die angehende Lehrpersonen weder als homogene Gruppe betrachtet noch in der Betrachtung des Einzelfalls verharret.

Zukünftig zu klären ist, wie der hier herausgearbeitete Befund, demzufolge unterschiedliche Gruppen von (angehenden) Lehrpersonen unterschiedliche Professionalisierungsbedarfe haben, in die Aus- und Weiterbildungspraxis übertragen werden kann. So liegt auf der Hand, dass die bei Tengberg et al. (2022) bereits angesprochenen „one-size-fits-all“-Ansätze kaum erfolgversprechend sein dürften, insofern sie an den Bedürfnissen zumindest einiger (Gruppen von) Lehrpersonen vorbeigehen und darüber hinaus Merkmale effektiver Aus- und Weiterbildung, wie bspw. zielgerichtetes Feedback, kaum berücksichtigen können (Lipowsky & Rzejak, 2021). Umgekehrt ist fraglich, ob sich wirksame Formate, die einzelne Lehrpersonen über lange Zeiträume hinweg individuell mittels Videografie und Coaching begleiten (z. B. Allen et al., 2011), angesichts knapper Ressourcen in der Breite implementieren lassen. Die Unterrichtsqualitätsforschung hat dieses Transferproblem erkannt (Begrich et al., 2023, S. 71), bislang allerdings noch keine nachhaltig wirkende und zugleich pragmatisch machbare Lösung identifizieren können. Bedenkenswert scheint der Hinweis von Begrich et al. (2023, S. 71), auf der Suche nach Lösungen für dieses Dilemma Aus- und Fortbildungskonzepte grundsätzlich neu zu denken, indem nicht nur klassische Top-down-Ansätze von der Wissenschaft in die Praxis, sondern stärker als bislang auch ko-konstruktive Ansätze verfolgt werden. So birgt die Arbeit in (auch digitalen) Lerngemeinschaften zahlreiche Potenziale, die bislang noch wenig genutzt und erforscht werden (Lipowsky & Rzejak, 2021, S. 46). Aber auch Studien im Design-Based-Research-Ansatz wären in dieser Hinsicht sicher vielversprechend (vgl. z. B. Dube & Prediger, 2017).

#### **4.2 | Spannungsfeld 2: Fachspezifische und generische Zugänge zu Unterrichtsqualität**

Eine weitere Frage, die im Diskurs zur Unterrichtsqualitätsforschung der letzten Jahre stark an Bedeutung gewonnen hat und auch für die Frage nach Praxisbezügen relevant ist, betrifft das Verhältnis von fachspezifischen und generischen Zugängen zu Unterrichtsqualität. Wurde insbesondere im Kontext der drei Basisdimensionen von Unterricht lange der Anspruch erhoben, die Qualität von Unterricht ohne Berücksichtigung fachlicher Spezifika beschreiben zu können (Praetorius et al., 2018, S. 408), zeichnet sich mittlerweile eine „*fachdidaktische Wende*“

(Reusser & Pauli, 2021, S. 192, Hervorh. im Orig.) in der Unterrichtsqualitätsforschung ab. Wie anhand des oben skizzierten Qualitätsmodells für den Literaturunterricht deutlich geworden ist, ist damit längst nicht mehr nur eine Konkretisierung einzelner Dimensionen gemeint, wie sie die Bildungswissenschaften bereits länger fordert (z. B. Klieme & Rakoczy, 2008). Vielmehr geht es auch um die Ergänzung fachlicher Dimensionen sowie Fragen der Hierarchisierung und gegenseitiger Kompensierbarkeit fachlicher und generischer Merkmale.

Je nachdem, wie man diese Fragen beantwortet, ergeben sich unterschiedliche Konsequenzen für die Aus- und Weiterbildung von Lehrpersonen. Begreift man fachliche Konkretisierungen als *Additum* zu generischen Qualitätsmerkmalen, wären modular gestaffelte Professionalisierungskonzepte denkbar, bei denen fachübergreifende Vorstellungen qualitätsvollen Unterrichts durch fachliche ergänzt werden. Denkt man das Verhältnis hingegen stärker *integrativ*, indem man etwa wie im obigen Modell annimmt, dass bestimmte (auch fachliche) Qualitätsdimensionen andere bedingen, wäre zu prüfen, inwiefern eine voneinander unabhängige Vermittlung tatsächlich sinnvoll ist.

Lohnend ist in diesem Zusammenhang auch ein genauerer Blick darauf, wie ‚Fachspezifik‘ im aktuellen Unterrichtsqualitätsdiskurs eigentlich modelliert wird. So argumentieren Wiprächtiger-Geppert et al. (2021) dafür, zwischen *fachspezifischen* und *fachbezogenen* Qualitätsmerkmalen zu unterscheiden. Erstere sind solche, die unmittelbar aus Spezifika des Lerngegenstandes, Lernbereichs oder Faches abgeleitet sind und daher nicht ohne Weiteres auf andere Kontexte übertragen werden können. Mit fachbezogenen Merkmalen sind indes solche gemeint, die prinzipiell für alle Fächer Gültigkeit haben und lediglich im konkreten Untersuchungssetting auf bestimmte Fächer, Lernbereiche und Gegenstände übertragen werden (Typisch sind etwa Itemformulierungen wie: ‚*Im Deutschunterricht* [vs. im Mathematikunterricht, Sportunterricht usw.] unterstützt uns unsere Lehrperson, wenn wir Fehler machen.‘).

Für die Ableitung von Praxisbezügen aus Forschungsbefunden ist diese Unterscheidung in doppelter Hinsicht relevant: Einerseits markiert sie, an welchen Punkten fachlich-fachdidaktische Überlegungen unentbehrlich werden, andererseits lässt sie Transferpotenziale bestimmter Qualitätsmerkmale auf andere Gegenstände, Lernbereiche und Fächer erkennen. Während im obigen Qualitätsmodell (Abbildung 1) beispielsweise die Dimension der ‚kognitiv-emotionalen Aktivierung‘ spezifisch auf die für den Literaturunterricht zentrale Relation von Text und Leser\*innen ausgerichtet ist, handelt es sich bei den Qualitätsmerkmalen in der Dimension der ‚Auswahl und Aufbereitung von Lerngegenständen‘ um solche, die auch in anderen Lernbereichen des Deutschunterrichts sowie in anderen Fächern prinzipiell relevant sind und hier nur auf den Umgang mit Literatur bezogen werden. Hinsichtlich der Konzeption von Professionalisierungsmaßnahmen ist die Betrachtung bedenkenswert, weil daran anschließend entschieden werden muss, an welcher Stelle man Lehrpersonen im Allgemeinen, Deutschlehrpersonen im Besonderen oder Literaturlehrpersonen im Speziellen adressiert.

Prägend für die aktuelle deutschdidaktische Forschung ist, dass das Thema Unterrichtsqualität noch sehr stark mit Blick auf einzelne Lernbereiche (Grammatikunterricht, Rechtschreibunterricht, Literaturunterricht usw.) und hier wiederum mit Bezug auf konkrete Gegenstände (z. B. Doppelkonsonantenschreibung, epische Texte) betrachtet wird. Für eine erste Phase der Relationierung bildungswissenschaftlicher und fachdidaktischer Perspektiven mag diese Engführung gewinnbringend gewesen sein. Zukünftig scheint es aber gerade mit Blick auf die Ableitung

von Anstößen und Empfehlungen für das Handlungsfeld Schule ratsam, die in Teilen sehr spezifischen Forschungen stärker aufeinander zu beziehen, und zwar sowohl innerhalb der Deutschdidaktik als auch im Vergleich zu anderen Fachdidaktiken und den Bildungswissenschaften.<sup>4</sup>

### 4.3 | Spannungsfeld 3: Interne und externe Validität

Schließlich besteht wie auch in anderen Forschungsfeldern eine Herausforderung hinsichtlich des Ableitens von Praxisbezügen aus Studien zur Unterrichtsqualität in der Sicherung geltender Gütekriterien und -standards. Fraglos wichtig sind dabei Aspekte wie Objektivität und Reliabilität, die in der Unterrichtsqualitätsforschung beispielsweise über die Durchführung von Inter-Rater-Reliabilitätsstudien gesichert werden. Gleichwohl reichen gute Ergebnisse im Bereich der Reliabilitätsprüfung allein nicht aus, um Anstöße oder Empfehlungen für das Handlungsfeld Schule auszusprechen. Entscheidend ist hier nämlich insbesondere die Validität der Befunde, bezüglich derer u. a. zwischen interner und externer Validität unterschieden werden kann.

*Interne Validität* ist laut Bortz und Döring (2016, S. 94) gegeben, „wenn die untersuchten Variablenzusammenhänge mit hoher Sicherheit als kausale Ursache-Wirkungs-Relationen zu interpretieren sind.“ Für das in der Unterrichtsqualitätsforschung vertretene Konzept von Unterrichtsqualität, das Annahmen ‚guten‘ Unterrichts mit Befunden zu ‚effektivem‘ Unterricht verzaubern möchte (Berliner, 2005), ist der Nachweis solcher Zusammenhänge im Rahmen der für die Unterrichtsforschung möglichen Bedingungen<sup>5</sup> zentral. Ansonsten ließe sich mit Blick auf Professionalisierungskonzepte berechtigterweise bezweifeln, warum bestimmte Merkmale der Tiefenstruktur für wichtiger befunden werden als andere.

In deutschdidaktischen Arbeiten zum Thema Unterrichtsqualität wird gemeinhin angenommen, dass die Wirksamkeit bestimmter Qualitätsmerkmale des Unterrichts, die vorrangig im mathematisch-naturwissenschaftlichen Unterricht nachgewiesen werden konnte, sich in ähnlicher Form auch im Deutschunterricht zeigt. Studien, die diese Annahme belegen, stehen tatsächlich aber noch weitestgehend aus (Lindl et al., 2023; Wieser, 2019). Ebenso zeigt ein Blick auf den mathematisch-naturwissenschaftlichen Diskurs, dass selbst dort nicht für alle Dimensionen gleichermaßen Effekte auf das Lernen der Schüler\*innen gezeigt werden konnten. So ist insbesondere in der fachdidaktisch relevanten Dimension der ‚kognitiven Aktivierung‘ noch immer eine gemischte Befundlage zu beobachten, die nur in einigen und längst nicht allen Fällen die erhofften Befunde belegen kann (Praetorius et al., 2018). Hier sind also zukünftig weitere Studien notwendig, die etwa im Rahmen von Interventionsdesigns die Wirksamkeit von Tiefenmerkmalen des Unterrichts prüfen.

<sup>4</sup> Vielversprechend erscheinen hier insbesondere die Resultate aus dem Projekt FALKO-PV (vgl. Lindl et al., 2023) sowie den Arbeiten des Leibniz-Netzwerkes Unterrichtsforschung (vgl. Begrich et al., 2023).

<sup>5</sup> Natürlich ist klar, dass die Unterrichtsforschung keine klassischen Experimentalstudien liefern kann, sondern allenfalls gut kontrollierte Interventionen (Berliner, 2002). Zudem sei daran erinnert, dass heutzutage Wirkungsforschung den Umstand berücksichtigt, dass bestimmte Merkmale des Lernangebots nicht direkt auf den Lernerfolg der Schüler\*innen wirken, sondern nur vermittelt über die je individuelle Nutzung der Lernenden (Vieluf et al., 2020).

Gleichwohl darf das Defizit an Interventionsstudien nicht dahingehend missverstanden werden, zukünftig nur noch solche Studien zu priorisieren, deren Design es erlaubt, (näherungsweise) kausale Schlüsse auf Wirkungen zu ziehen. Berücksichtigt werden muss nämlich neben der internen auch die *externe Validität*, die bei Studien immer in dem Maß vorliegt, „wie sich die Ergebnisse auf andere Orte, Zeiten, Operationalisierungen der abhängigen und unabhängigen Variablen oder auf andere Personen als die konkret untersuchten verallgemeinern lassen“ (Döring & Bortz, 2016, S. 95). Insofern Interventionsstudien in der Regel sehr starke Vorgaben hinsichtlich des Gegenstandes, der unterrichtlichem Vorgehen oder der Klassenstufe machen, ist diese Verallgemeinerbarkeit oft nur bedingt gegeben. So stellen etwa Carl & Scherf mit Blick auf eine der wenigen literaturdidaktischen Interventionsstudien im deutschen Raum, ÄSKIL, fest, dass die dort untersuchten, diametral entgegengesetzten Gesprächstypen KOKIL und ÄSKIL „nur bedingt an authentische Unterrichtskommunikation zu literarischen Texten [anknüpfen]“ und „Konsequenzen für eine schulische Praxis [...] nicht unmittelbar abgeleitet werden können (und sollen)“ (Carl & Scherf, 2022, S. 6). Die oben vorgestellte Studie zum Unterricht Praxissemesterstudierender berücksichtigt authentische Unterrichtsverläufe demgegenüber sicher stärker, sagt aber gleichwohl nichts über die Wirkungen der hochgradig heterogenen Settings aus. Mit Blick auf das Kaskadenmodell (vgl. Meier & Emmersberger i. d. B.) steht also weniger die schüler\*innenseitige Nutzung (Säule 4) oder der Lernertrag der Schüler\*innen (Säule 5), sondern die Gestaltung der Angebote (Säule 3) unter Alltagsbedingungen im Mittelpunkt.

Mit den vorherigen Überlegungen soll nicht der Eindruck erweckt werden, dass sich interne und externe Validität grundsätzlich gegenseitig ausschließen. Vielmehr soll herausgestellt werden, dass bei der Frage nach Praxisbezügen immer beide Seiten mitbedacht werden müssen: Welche Unterrichtsmerkmale erweisen sich als lernwirksam und wie lassen sie sich an die (z. T. doch sehr heterogene) Alltagspraxis des Deutschunterrichts rückbinden?

## 5 | Schluss

Die abschließenden Betrachtungen haben deutlich gemacht, dass der Weg von der Messung von Unterrichtsqualität – so aufschlussreich diese mitunter auch sein mag – hin zur Verbesserung von Unterrichtsqualität weniger geradlinig verläuft, als er auf den ersten Blick scheint. Gerade im Bereich des Deutschunterrichts wissen wir aktuell schlicht noch zu wenig darüber, welche Qualitätsmerkmale tatsächlich wirken, inwiefern diese spezifisch für das Fach Deutsch bzw. seine Lernbereiche und Gegenstände sind und wie die Qualität des Unterrichts unter Berücksichtigung von Erkenntnissen der Professionalisierungsforschung und angesichts der Heterogenität der Lehrpersonen verbessert werden kann. Diesbezüglich konzeptionell wie empirisch weiterzuarbeiten, bleibt also eine zentrale Aufgabe deutschdidaktischer Unterrichtsqualitätsforschung.

## 6 | Literatur

- 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>
- Anselm, S. (2011). *Kompetenzentwicklung in der Deutschlehrerbildung: Modellierung und Diskussion eines fachdidaktischen Analyseverfahrens zur empiriegestützten Wirkungsforschung*. Peter Lang.
- Backhaus, K., Erichson, B., Plinke, W., & Weiber, R. (2016). *Multivariate Analysemethoden: Eine anwendungsorientierte Einführung* (14. Aufl.). Springer. <https://doi.org/10.1007/978-3-662-46076-4>
- Bäuerlein, K., Reintjes, C., Fraefel, U., & Jünger, S. (2018). Selbstprofessionalisierung in der Schule? – Eine Bestandsaufnahme hinsichtlich der studienunabhängigen Lehrtätigkeit von Lehramtsstudierenden im Schulfeld. *Forschungsperspektiven*, 10, 27–46.
- Begrich, L., Praetorius, A.-K., Decristan, J., Fauth, B., Göllner, R., Herrmann, C., Kleinknecht, M., Taut, S., & Kunter, M. (2023). Was tun? Perspektiven für eine Unterrichtsqualitätsforschung der Zukunft. *Unterrichtswissenschaft*, 63–97. <https://doi.org/10.1007/s42010-023-00163-4>
- Berliner, D. C. (2002). Educational Research: The Hardest Science of All. *Educational Researcher*, 31(8), 18–20. <https://doi.org/10.3102/0013189X031008018>
- Berliner, D. C. (2004). Describing the Behavior and Documenting the Accomplishments of Expert Teachers. *Bulletin of Science, Technology & Society*, 24(3), 200–212. <https://doi.org/10.1177/0270467604265535>
- 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>
- Bertschi-Kaufmann, A., & Scherf, D. (Hrsg.). (2018). *Ästhetische Rezeptionsprozesse in didaktischer Perspektive*. Beltz Juventa.
- Böhnert, K. (2020). „Du musst dich immer fragen: Worin steckt die Info?“ – kognitive Aktivierung in inklusiven Lerngruppen. *k:ON - Kölner Online Journal für Lehrer\*innenbildung*, 26–49. <https://doi.org/10.18716/OJS/KON/2020.2.02>
- 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>
- Carl, M.-O., & Scherf, D. (2022). Über den angemessenen Umgang mit empirischen Erkenntnissen in der Literaturdidaktik. *Zeitschrift für Sprachlich-Literarisches Lernen und Deutschdidaktik*, 2, 1–25. <https://doi.org/10.46586/SLLD.Z.2022.8894>
- Charalambous, C. Y., & Praetorius, A.-K. (2020). Creating a forum for researching teaching and its quality more synergistically. *Studies in Educational Evaluation*, 67(100894), 1–8. <https://doi.org/10.1016/j.stueduc.2020.100894>
- Christmann, U., & Schreier, M. (2003). Kognitionspsychologie der Textverarbeitung und Konsequenzen für die Bedeutungskonstitution literarischer Texte. In F. Jannidis, G. Lauer, M. Martínez, & S. Winko (Hrsg.), *Regeln der Bedeutung: Zur Theorie der Bedeutung literarischer Texte* (S. 246–285). De Gruyter. <https://doi.org/10.1515/9783110907018.246>
- Cramer, C. (2010). Kompetenzerwartungen Lehramtsstudierender. Grenzen und Perspektiven selbst-eingeschätzter Kompetenzen in der Lehrerbildungsforschung. In A. Gehrman, U. Hericks, & M. Lüders (Hrsg.), *Bildungsstandards und Kompetenzmodelle: Beiträge zu einer aktuellen Diskussion über Schule Lehrerbildung und Unterricht* (S. 85–97). Klinkhardt.

- 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>
- Depner, S., Kernen, N., & Pieper, I. (2020). Gegenstandskonstitution und literarisches Lernen im Unterrichtsgespräch: Die Videostudie im Projekt TAMoLi – Texte, Aktivitäten und Motivationen im Literaturunterricht der Sekundarstufe I. In C. Dawidowski, A. R. Hoffmann, A. R. Stolle, & J. Witte (Hrsg.), *Schulische Literaturvermittlungsprozesse im Fokus empirischer Forschung* (S. 141–160). Peter Lang.
- Döring, N., & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (5., vollst. überarb., aktual. und erw. Aufl.). Springer.
- Dube, J., & Prediger, S. (2017). Design-Research – Neue Forschungszugriffe für unterrichtsnahe Lernprozessforschung in der Deutschdidaktik. *leseforum.ch*, 1, 1–14.
- Fuller, F. F., & Bown, O. H. (1975). Becoming a teacher. In K. Ryan (Hrsg.), *Teacher Education.: 74th Yearbook of the National Society for the Study of Education. Part II.* (S. 25–52). The University of Chicago Press.
- Gröschner, A., & Hascher, T. (2022). Praxisphasen in der Lehrerinnen- und Lehrerbildung. In M. Harring, C. Rohlf, & M. Gläser-Zikuda (Hrsg.), *Handbuch Schulpädagogik* (2. Aufl., S. 709–720). Waxmann.
- Hanisch, A.-K. (2018). *Kognitive Aktivierung im Rechtschreibunterricht: Eine Interventionsstudie in der Grundschule*. Waxmann.
- Harwart, M., & Scherf, D. (2018). „Vielleicht muss man aber auch so damit leben können und es aushalten.“: Zur Bedeutung des Lehrerhandelns in schulischen ästhetischen Rezeptionsprozessen. In A. Bertschi-Kaufmann & D. Scherf (Hrsg.), *Ästhetische Rezeptionsprozesse in didaktischer Perspektive* (S. 149–163). Beltz Juventa.
- Heins, J. (2017). *Lenkungsgrade im Literaturunterricht: Zum Einfluss stark und gering lenkender Aufgabensets auf das Textverstehen*. Springer Fachmedien. <https://doi.org/10.1007/978-3-658-17890-1>
- 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>
- Heins, J. (2019). Professionelle Unterrichtswahrnehmung: Wie Mentoren, Master- und Bachelorstudierende Literaturunterricht wahrnehmen und interpretieren. In C. Führer & F.-M. Führer (Hrsg.), *Dissonanzen in der Deutschlehrerbildung: Theoretische, empirische und hochschuldidaktische Perspektiven* (S. 123–144). Waxmann.
- Hess, M., & Lipowsky, F. (2020). Zur (Un-)Abhängigkeit von Oberflächen- und Tiefenmerkmalen im Grundschulunterricht: Fragen von Lehrpersonen im öffentlichen Unterricht und in Schülerarbeitsphasen im Vergleich. *Zeitschrift für Pädagogik*, 66(66. Beiheft), 117–131.
- Hesse, F. (2024). *Qualitäten von Literaturunterricht. Eine Videostudie im Praxissemester*. J. B. Metzler. <https://doi.org/10.1007/978-3-662-68775-8>
- Hesse, F., & Winkler, I. (2022). Fachliche Qualität im Literaturunterricht. *SLLD-Z*, 2, 1–29. <https://doi.org/10.46586/SLLD.Z.2022.9605>
- Junker, R., Gold, B., & Holodyski, 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>

- Kleinbub, I. (2018). Kriteriengeleitetes Rating: Ein Datenerhebungsverfahren zur Einschätzung von Unterrichtsmerkmalen. In J. M. Boelmann (Hrsg.), *Empirische Forschung in der Deutschdidaktik: Band 2: Erhebungs- und Auswertungsverfahren* (S. 341–358). Schneider Verlag Hohengehren.
- Klieme, E., & Rakoczy, K. (2008). Empirische Unterrichtsforschung und Fachdidaktik: Outcome-orientierte Messung und Prozessqualitäten des Unterrichts. *Zeitschrift für Pädagogik*, 54(2), 222–237.
- Krauss, S. (2020). Expertise-Paradigma in der Lehrerinnen und Lehrerbildung. In C. Cramer, J. König, M. Rothland, & S. Blömeke (Hrsg.), *Handbuch Lehrerinnen- und Lehrerbildung* (S. 154–162). Klinkhardt.
- Lessing-Sattari, M., & Freudenberg, R. (Hrsg.). (2020). *Zum Verhältnis von Irritation und Staunen im Rahmen literarästhetischer Erfahrung*. Peter Lang.
- Lindl, A., Böhringer, S., Ehrich, P., Gürtner, M., Gutmiedl, M., Krämer, A., Rader, M., & Simböck, L. (2023, Januar 1). *Modellierung fachspezifischer Merkmale von Unterrichtsqualität in den sechs Schulfächern Deutsch, Englisch, Evangelische Religion, Latein, Mathematik und Musik*.
- Lipowsky, F. (2020). Unterricht. In E. Wild & J. Möller (Hrsg.), *Pädagogische Psychologie* (3., vollständig überarbeitete und aktualisierte Aufl., S. 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 (Hrsg.), *Konstruktionen von Fachlichkeit: Ansätze, Erträge und Diskussionen der empirischen Unterrichtsforschung* (S. 183–202). Klinkhardt.
- Lipowsky, F., & Rzejak, D. (2021, Januar 1). *Fortbildungen für Lehrpersonen wirksam gestalten*. <https://doi.org/10.11586/2020080>
- 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>
- Maak, D., & Ricart Brede, J. (2014). Empirische Erfassung von Invasivität in videografierten Lehr-Lernsituationen: Entwicklung und Erprobung eines Beobachtungssystems. In A. Neumann & I. Mahler (Hrsg.), *Empirische Methoden der Deutschdidaktik: Audio- und videografierte Unterrichtsforschung* (S. 151–173). Schneider Verlag Hohengehren.
- Magirius, M., Scherf, D., & Steinmetz, M. (2022). Lernunterstützung im Literaturgespräch: Nachweis und Wirkung eines potenziellen Qualitätsaspekts gesprächsförmigen Literaturunterrichts. *SLLD*, 1–29. <https://doi.org/10.46586/SLLD.Z.2022.9552>
- Pauli, C. (2014). Ratingverfahren. *Journal für LehrerInnenbildung*, 14(1), 56–59.
- 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.
- Pissarek, M., & Schilcher, A. (2017). FALKO-D: Die Untersuchung des Professionswissens von Deutschlehrenden: Entwicklung eines Messinstruments zur fachspezifischen Lehrerkompetenz und Ergebnisse zu dessen Validierung. In S. Krauss, A. Lindl, A. Schilcher, M. Fricke, A. Göhring, B. Hofmann, P. Kirchoff, & R. H. Mulder (Hrsg.), *FALKO: Fachspezifische Lehrerkompetenzen: Konzeption von Professionswissenstests in den Fächern Deutsch, Englisch, Latein, Physik, Musik, Evangelische Religion und Pädagogik* (S. 68–109). Waxmann.

- 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., 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>
- Reusser, K., & Pauli, C. (2021). Unterrichtsqualität ist immer generisch und fachspezifisch. Ein Kommentar aus kognitions- und lehr-lerntheoretischer Sicht. *Unterrichtswissenschaft*, 49(2), 189–202. <https://doi.org/10.1007/s42010-021-00117-8>
- Riegler, S., Wiprächtiger-Geppert, M., Kusche, D., & Schurig, M. (2020). Wie Primarlehrpersonen Rechtschreiben unterrichten. Zur Praxis des Rechtschreibunterrichts in Deutschland und der Schweiz. *Didaktik Deutsch*, 25(49), 50–69.
- 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>
- Scheidig, F., & Holmeier, M. (2022). Unterrichten neben dem Studium – Implikationen für das Studium und Einfluss auf das Verlangen nach hochschulischen Praxisbezügen. *Zeitschrift für Bildungsforschung*, 12(3), 479–496. <https://doi.org/10.1007/s35834-022-00349-3>
- Schendera, C. F. G. (2010). *Clusteranalyse mit SPSS: Mit Faktorenanalyse*. Oldenbourg Verlag.
- Schmidt, R. (2016). *Mit Kunstwerken zum Denken anregen: Eine empirische Untersuchung zur kognitiven Aktivierung im Rahmen der Kunstrezeption in der Grundschule*. kopaed.
- Schwahl, M. (2019). „... Aber jetzt zerbreche ich mir den Kopf darüber!“—Kognitive Aktivierung im Literaturunterricht. In A. Gawatz & K. Stürmer (Hrsg.), *Kognitive Aktivierung im Unterricht: Befunde der Bildungsforschung und fachspezifische Zugänge* (S. 86–103). Westermann Schulbuchverlag.
- Selting, M., Auer, P., Barth-Weingarten, D., Bergmann, J., Bergmann, P., Birkner, K., Couper-Kuhlen, E., Deppermann, A., Gilles, P., Günthner, S., Hartung, M., Kern, F., Mertzluft, C., Meyer, C., Morek, M., Oberzaucher, F., Peters, J., Quasthoff, U., Schütte, W., ... Uhmann, S. (2009). Gesprächsanalytisches Transkriptionssystem 2 (GAT 2). *Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion*, 10, 353–402.
- Siebenhüner, S., Depner, S., Fässler, D., Kernen, N., Bertschi-Kaufmann, A., Böhme, K., & Pieper, I. (2019a). Unterrichtstextauswahl und schülerseitige Leseinteressen in der Sekundarstufe I: Ergebnisse aus der binationalen Studie TAMoLi. *Didaktik Deutsch*, 24(47), 44–64. <https://doi.org/10.25656/01:21015>
- Siebenhüner, S., Depner, S., Fässler, D., Kernen, N., Bertschi-Kaufmann, A., Böhme, K., & Pieper, I. (2019b, Januar 1). *Anhangsdokument zum Artikel „Unterrichtstextauswahl und schülerseitige Leseinteressen in der Sekundarstufe I: Ergebnisse aus der binationalen Studie TAMoLi“ in Didaktik Deutsch*

47. S. 44-64. [http://www.literaturunterricht.ch/downloads/anhang-zu-siebenhuener-et-al-2019-unterrichtstextauswahl-und-leseinteressen-tamoli/at\\_download/file](http://www.literaturunterricht.ch/downloads/anhang-zu-siebenhuener-et-al-2019-unterrichtstextauswahl-und-leseinteressen-tamoli/at_download/file)
- Spinner, K. H. (2006). Literarisches Lernen. *Praxis Deutsch*, 200, 6–13.
- Stahns, R. (2013). *Kognitive Aktivierung im Grammatikunterricht: Videoanalysen zum Deutschunterricht*. Schneider Verlag Hohengehren.
- Steinmetz, M. (2020). *Verstehenssupport im Literaturunterricht: Theoretische und empirische Fundierung einer literaturdidaktischen Aufgabenorientierung*. Springer VS. <https://doi.org/10.1007/978-3-658-28378-0>
- Tengberg, M., van Bommel, J., Nilsberth, M., Walkert, M., & Nissen, A. (2022). The Quality of Instruction in Swedish Lower Secondary Language Arts and Mathematics. *Scandinavian Journal of Educational Research*, 66(5), 760–777. <https://doi.org/jäge>
- Ulrich, I., Klingebiel, F., Bartels, A., Staab, R., Scherer, S., & Gröschner, A. (2020). Wie wirkt das Praxissemester im Lehramtsstudium auf Studierende? Ein Systematischer Review. In I. Ulrich & A. Gröschner (Hrsg.), *Praxissemester im Lehramtsstudium in Deutschland: Wirkung auf Studierende* (S. 1–66). Springer VS. [https://doi.org/10.1007/978-3-658-24209-1\\_1](https://doi.org/10.1007/978-3-658-24209-1_1)
- 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.
- White, M., & Klette, K. (2023). What's in a score? Problematizing interpretations of observation scores. *Studies in Educational Evaluation*, 77(101238), 1–8. <https://doi.org/10.1016/j.stueduc.2023.101238>
- Wieser, D. (2019). Gegenwärtiger Stand der empirischen Unterrichtsforschung zum Literaturunterricht. In M. Kämper-van den Boogaart & K. H. Spinner (Hrsg.), *Lese- und Literaturunterricht. Teil 2: Kompetenzen und Unterrichtsziele, Methoden und Unterrichtsmaterialien: Gegenwärtiger Stand der empirischen Unterrichtsforschung* (3., stark überarbeitete Aufl., Bd. 11/2, S. 353–384). Schneider Verlag Hohengehren.
- Winkler, I. (2011). *Aufgabenpräferenzen für den Literaturunterricht: Eine Erhebung unter Deutschlehrkräften*. VS Verlag für Sozialwissenschaften. <https://doi.org/10.1007/978-3-531-92698-8>
- Winkler, I. (2015). „Subjektive Involviertheit und genaue Wahrnehmung miteinander ins Spiel bringen“. Überlegungen zur Spezifikation eines zentralen Konzepts für den Literaturunterricht. *Leseräume*, 2, 155–168.
- Winkler, I. (2017). Potenzial zu kognitiver Aktivierung im Literaturunterricht: Fachspezifische Profilierung eines prominenten Konstrukts der Unterrichtsforschung. *Didaktik Deutsch*, 22(43), 78–97.
- 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. (2021). Literarisches Lernen im Spiegel internationaler Forschung. *SLLD - Zeitschrift für Sprachlich-Literarisches Lernen und Deutschdidaktik*, 1, 1–16. <https://doi.org/10.46586/SLLD.Z.2021.8771>
- 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>

- Winkler, I., & Steinmetz, M. (2016). Zum Spannungsverhältnis von deutschdidaktischen Fragestellungen und empirischen Erkenntnismöglichkeiten am Beispiel des Projekts KoALa. In M. Krelle & W. Senn (Hrsg.), *Qualitäten von Deutschunterricht* (S. 37–56). Fillibach bei Klett.
- Wiprächtiger-Geppert, M., Stahns, R., & Riegler, S. (2021). Fachspezifität von Unterrichtsqualität in der Deutschdidaktik. *Unterrichtswissenschaft*, 49(2), 203–209. <https://doi.org/10.1007/s42010-021-00109-8>
- Wirtz, M., & Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität*. Hogrefe.
- Zabka, T. (2012). Didaktische Analyse literarischer Texte. Theoretische Überlegungen zu einer Lehrerkompetenz. In D. A. Frickel, C. Kammler, & G. Rupp (Hrsg.), *Literaturdidaktik im Zeichen von Kompetenzorientierung und Empirie: Perspektiven und Probleme* (S. 139–162). Fillibach Verlag.
- Zabka, T. (2015). Was ist Hochschulreife im Umgang mit Literatur? *Didaktik Deutsch*, 20(38), 136–150.