

FLORIAN HESSE

Wie hängen die Qualitätsdimensionen von Literaturunterricht untereinander und mit Oberflächenmerkmalen des Unterrichts zusammen? Befunde einer Videostudie zur Unterrichtsqualität im Praxissemester

Abstract

Die literaturdidaktische Unterrichtsqualitätsforschung hat in den letzten Jahren Qualitätsmerkmale alltäglich stattfindenden Literaturunterrichts meist nur in isolierter Form, nicht aber in Zusammenhang mit anderen Merkmalen des Unterrichts untersucht. Deshalb wird jüngst die Forderung erhoben, Zusammenhänge zwischen den Qualitätsmerkmalen des Unterrichts sowie die Abhängigkeit derselben von der Oberflächengestaltung stärker in den Blick zu nehmen. Der Beitrag greift dieses Desiderat ausgehend von einer Studie zum Literaturunterricht im Praxissemester auf, in der 22 Unterrichtsstunden von 22 Studierenden zu epischen Kurztexten videografiert wurden. Die Stunden wurden hinsichtlich verschiedener Qualitätsdimensionen geratet (kognitive Aktivierung, Auswahl und Aufbereitung der Lerngegenstände und Fachmethoden, Klassenführung, sozio-emotionale Unterstützung) und bezüglich ausgewählter Oberflächenmerkmale kodiert (inhaltsbezogene Aktivitäten, Sozialformen). Mithilfe von Spearman-Korrelationen wurden Zusammenhänge zwischen den Qualitätsdimensionen bestimmt. Kruskal-Wallis-Tests wurden angewandt, um die Abhängigkeit der Qualitätsdimensionen von den Oberflächenmerkmalen zu prüfen. Hierbei zeigte sich einerseits, dass signifikante Korrelationen zwischen den generischen Qualitätsdimensionen, Klassenführung und sozio-emotionale Unterstützung, sowie zwischen den fachlichen Qualitätsdimensionen, Auswahl und Aufbereitung von Lerngegenständen und Fachmethoden sowie kognitiv-emotionale Aktivierung (hier insbesondere bezogen auf Aspekte der Unterstützung), bestehen. Demgegenüber ließen sich zwischen den generischen und fachlichen Qualitätsdimensionen keine Zusammenhänge ermitteln. Zudem konnte keine Abhängigkeit der Qualitätsmerkmale von den Oberflächenmerkmalen nachgewiesen werden. Damit legen die Befunde einerseits nahe, dass die fachlichen Qualitätsdimensionen eng miteinander verwoben sind, weshalb deren Merkmale bei der Vermittlung von Fähigkeiten der didaktischen Analyse aufeinander bezogen werden sollten. Andererseits zeigen die Ergebnisse, dass die angebotsseitige Unterrichtsqualität weniger an der Auswahl von Sozialformen oder inhaltsbezogenen Aktivitäten festzumachen ist, sondern an der davon unabhängigen Tiefenstruktur.

1 | Einleitung

Die Literaturdidaktik hat sich in den vergangenen Jahren zunehmend mit der Frage beschäftigt, was qualitätsvollen Literaturunterricht auszeichnet. Dabei wurde in Anlehnung an das Modell der drei Basisdimensionen von Unterrichtsqualität (Klieme et al. 2001; Praetorius et al. 2018) vor allem die Dimension der kognitiven Aktivierung in den Blick gerückt, die Auskunft darüber gibt, inwiefern die unterrichtlichen Lernangebote die Lernenden zu einer vertieften Auseinandersetzung mit dem fachlichen Unterrichtsgegenstand anregt (Depner et al. 2020; Schwahl 2019; Winkler 2017, 2020). Darüber hinaus lassen sich einige Arbeiten der Aufgaben- und Gesprächsforschung in der Literaturdidaktik mit der Qualitätsdimension der konstruktiven Unterstützung in Verbindung bringen, die erfasst, inwiefern Lernende bei der Bewältigung von Anforderungen unterstützt werden (Harwart/Scherf 2018; Heins 2017; Magirius et al. 2022; Steinmetz 2020).

Kennzeichnend für die benannten Studien ist insgesamt der Umstand, dass sie Literaturunterricht – wenn überhaupt – nur auf Basis klein(st)er Stichproben (i. d. R. Einzelstunden) betrachten und dabei stets nur einzelne Qualitätsdimensionen in den Blick nehmen. Eine theoretische Modellierung und empirische Untersuchung sowohl der o. g. Qualitätsdimensionen als auch der Oberflächengestaltung von Literaturunterricht (z. B. mit Blick auf Methoden, Sozialformen und inhaltsbezogene Aktivitäten) auf breiterer Datenbasis stellt folglich eine Forschungslücke dar (Wieser 2019).

Hesse (2024) hat dieses Desiderat jüngst in mehrfacher Hinsicht aufgegriffen. Ausgehend von aktuellen Entwicklungen der pädagogisch-psychologischen Unterrichtsqualitätsforschung wurde ein mehrdimensionales Modell zur Analyse der Gestaltung und Qualität von Literaturunterricht entworfen. Dieses erweitert nicht nur die Dimension der kognitiven Aktivierung um eine für die Auseinandersetzung mit Literatur wichtige emotionale Perspektive (‚kognitiv-emotionale Aktivierung‘), sondern berücksichtigt mit den Dimensionen ‚Auswahl und Aufbereitung von Lerngegenständen und Fachmethoden‘, ‚Klassenführung‘ und ‚sozio-emotionale Unterstützung‘ auch weitere (über-)fachliche Voraussetzungen für eine gelingende Anregung von gegenstandsbezogenen Lern- und Verstehensprozessen (siehe zum Modell auch Hesse/Winkler 2022). Auf Basis des Modells wurden ferner Instrumente zur beobachtungs-basierten Beschreibung von Literaturunterricht und zur Erfassung seiner Qualität konstruiert, welche bei einer Stichprobe von insgesamt 22 videografierten Literaturstunden Studierender im Praxissemester zur Anwendung kamen. Hierbei zeigte sich, dass die Studierenden die verschiedenen Qualitätsdimensionen in stark divergierender Qualität realisieren und sich bezogen auf die einzelnen Dimensionen unterschiedlichen Realisierungsvarianten zuordnen lassen.

Offen blieben im Rahmen dieser ersten Exploration allerdings zwei Fragen: Zum einen steht eine Untersuchung dazu aus, inwiefern die unterschiedlichen Ausprägungen der Qualitätsdimensionen zusammenhängen – ob also beispielsweise eine fachlich angemessene Aufbereitung der Lerngegenstände und Methoden auch mit einem kognitiv aktivierenden Unterricht einhergeht. Zum anderen gilt es zu klären, ob die heterogenen Qualitätsausprägungen der einzelnen Dimensionen von der Oberflächengestaltung (hier: den Sozialformen oder inhaltsbezogenen Aktivitäten) abhängen.

Der vorliegende Beitrag geht diesen beiden noch offenen Fragen nach und knüpft damit an die zunehmend laut werdende Forderung der deutschdidaktischen Unterrichtsqualitätsforschung an, Qualitätsdimensionen nicht nur isoliert, sondern in Zusammenhängen mit anderen Qualitätsdimensionen sowie Oberflächenmerkmalen des Unterrichts zu betrachten (Stahns 2021, 74–75; Wiprächtiger et al. 2023). Dazu werden im Beitrag zunächst aktuelle Modellierungen von Unterrichtsqualität vorgestellt und mit Blick auf bestehende empirische Befunde grundiert. Daran anschließend erkundet der Beitrag die adressierten Zusammenhänge mittels Korrelations- sowie Varianzanalysen und diskutiert diese im Lichte des aktuellen deutschdidaktischen Forschungsstands. Der Beitrag schließt mit einem Fazit und Ausblick.

2 | Grundannahmen und Forschungsstand

2.1 | Begriffsklärung: Tiefen- und Oberflächenmerkmale des Unterrichts

Eine Grundannahme der (quantitativ ausgerichteten) Unterrichtsqualitätsforschung besteht darin, dass bei der Beschreibung von Unterricht eine Oberflächen- und eine Tiefenebene unterschieden werden muss (Decristan et al. 2020). Dieser Annahme liegt im Kern eine konstruktivistische Auffassung von Lehren und Lernen zugrunde (vgl. grundlegend Renkl 2020; Reusser 2006), der zufolge das Lernen der Schüler:innen nicht zuvorderst durch den bloßen Einsatz bestimmter Methoden und Sozialformen beeinflusst wird (= Oberflächenmerkmale), sondern durch die zugrundeliegenden lernprozessbezogenen Qualitäten des Unterrichts (= Tiefenmerkmale)¹. Letztere zielen direkt auf die kognitive Verarbeitung der Lernenden oder die dafür notwendigen Voraussetzungen. Hieraus ergeben sich Konsequenzen für unterrichtsbezogene Wirkungsannahmen und darauf aufbauende Forschungsdesigns:

Anstelle einer bloßen Propagierung oder eines reinen Effektivitätsvergleichs bestimmter Methoden (z. B. Wochenplan) oder spezifischer Ablaufmuster (z. B. Direkte Instruktion) wird der Fokus darauf gelenkt, sich mit der Frage auseinanderzusetzen, *warum* der entsprechende Unterricht (oder eine bestimmte Methode) mehr oder weniger förderlich ist für Lernprozesse und deren Ergebnisse und wie Unterricht in diesem Sinne ‚lernwirksam‘ gestaltet werden kann. (Decristan et al. 2020, 103; Herv. im Orig.)

Welche Merkmale des Unterrichts zur Oberflächen- bzw. Tiefenebene des Unterrichts gezählt und folglich in empirischen Untersuchungen verstärkt betrachtet werden, hängt also nicht – wie vielfach verkürzt im Diskurs dargestellt – damit zusammen, dass Oberflächen- oder auch Sichtmerkmale des Unterrichts besser beobachtbar seien als die Tiefenmerkmale. Schließlich müssen auch die Tiefenmerkmale des Unterrichts mithilfe beobachtbarer Indikatoren erschlossen werden. Vielmehr hängt die Zuordnung von Unterrichtsmerkmalen zur einen oder anderen Ebene davon ab, ob vom jeweiligen Merkmal im Rahmen der angelegten Lehr-Lern-Theorie ein direkter Einfluss auf das Lernen der Schüler:innen zu erwarten ist.

2.2 | Modellierung der Tiefenebene des Unterrichts

In der deutschsprachigen Unterrichtsforschung wird bei der Modellierung von Unterrichtsqualitätsdimensionen auf der Tiefenebene in der Regel auf das Modell der *drei Basisdimensionen von Unterrichtsqualität* Bezug genommen. Diesem Modell liegt die Annahme zugrunde, dass mit den drei Dimensionen kognitive Aktivierung, Klassenführung und konstruktive Unterstützung die Qualität von Unterricht über Fächer, Jahrgangsstufen und sogar Kulturen hinweg möglichst sparsam beschrieben werden kann (vgl. Praetorius et al. 2018, 408).

Im aktuellen Diskurs der Unterrichtsqualitätsforschung wird zunehmend hinterfragt, inwiefern die in diesem Modell veranschlagten Dimensionen ausreichen, um Unterrichtsqualität umfassend zu beschreiben. Sowohl aus bildungswissenschaftlicher als auch fachdidaktischer Perspektive wird dabei kritisiert, dass die fachlichen Gegenstände ebenso wenig berücksichtigt werden wie ein fachlich angemessener Umgang mit diesen (vgl. die Beiträge in *Unterrichtswissenschaft* 2020/3 und 2021/2).

Aus *bildungswissenschaftlicher* Sicht lieferten in den letzten Jahren vor allem die Arbeiten von Praetorius und Charalambous (z. B. 2018) entscheidende Impulse für die Weiterentwicklung der Basisdimensionen. Ausgangspunkt war eine Sichtung von vier fachdidaktischen, vier generischen und vier hybriden Frameworks zur Beschreibung von Unterrichtsqualität, die die Autor:innen zu einem sog. „Syntheseframework“ bündelten. Hierbei zeigte sich, dass die drei Basisdimensionen zahlreiche blinde Flecke aufweisen, da beispielsweise Aspekte der Auswahl

¹ Dieser Unterscheidung entsprechend werden im vorliegenden Beitrag die Begriffe ‚Tiefenmerkmal‘ bzw. ‚Tiefendimension‘ und ‚Qualitätsmerkmal‘ bzw. ‚Qualitätsdimension‘ synonym verwendet.

und Präsentation von Lerngegenständen keinerlei Berücksichtigung in diesem Framework finden. Entsprechend werden im Syntheseframework über die drei Basisdimensionen hinaus weitere Dimensionen veranschlagt, die beispielsweise die *Auswahl und Präsentation von Lerngegenständen*, qualitätsvolle *Übungen* oder das *formative Assessment* betreffen (vgl. auch Praetorius, Rogh et al., 2020).

Aus *fachdidaktischer* Sicht haben insbesondere Studien im Bereich der Mathematikdidaktik dazu Anlass gegeben, bestehende Qualitätsmodelle zu überdenken. So konnte etwa Brunner (2017) über die vergleichende Anwendung dreier mathematikdidaktischer Frameworks zur Beurteilung von Unterrichtsqualität auf dieselbe Unterrichtsstunde zeigen, dass die Stunde je nach Gewichtung fachlicher Qualitätskriterien in den unterschiedlichen Frameworks mal sehr gut, mal durchschnittlich und mal unterdurchschnittlich abschneidet. Dies veranlasste die Autorin im Weiteren dazu, in einem eigens entwickelten Qualitätsmodell die fachliche Korrektheit als Voraussetzung für die anderen Qualitätsdimensionen zu modellieren. In ähnlicher Weise konnten Lipowsky et al. (2018) herausarbeiten, dass die fachliche Qualität in den Unterrichtsstunden der Pythagoras-Studie (vgl. für die Operationalisierung Drollinger-Vetter 2011) eine eigenständige Dimension bildet, die sich losgelöst von den übrigen Basisdimensionen beschreiben lässt.

In der deutschdidaktischen Debatte um Unterrichtsqualität haben diese Befunde dazu beigetragen, dass aktuellere Modelle fachliche Unterrichtsqualität in einem umfassenderen Sinne modellieren und nicht mehr nur einzelne Dimensionen, wie bspw. die kognitive Aktivierung, berücksichtigen. So gehen etwa Wiprächtiger-Geppert et al. (2021, 206) in der Studie Profess-R für den Rechtschreibunterricht von fünf Qualitätsdimensionen aus: „1. Effiziente Klassenführung, 2. Auswahl und Präsentation des Lerngegenstands, 3. Sozio-emotionale Unterstützung, 4. Kognitive Aktivierung und 5. Kognitive Unterstützung“. Gegenüber den drei Basisdimensionen wird also eine fachliche Dimension ergänzt und die konstruktive Unterstützung in eine sozio-emotionale und kognitive Dimension aufgespalten (vgl. hierzu Kleickmann et al. 2020).

In ähnlicher Weise geht auch Hesse (2024) vor, der ein Modell von Unterrichtsqualität für den Literaturunterricht entwickelt hat. Eine fachliche Perspektive wird hier in doppelter Weise eingebracht. Zum einen wird die Dimension der kognitiven Aktivierung angesichts der Spezifika literarischer Texte in eine Dimension *kognitiv-emotionaler Aktivierung* überführt. Zum anderen sieht auch Hesse die Notwendigkeit, eine Dimension der *Auswahl und Aufbereitung von Lerngegenständen und Fachmethoden* zu ergänzen, die im Literaturunterricht beispielsweise Fragen der Textauswahl sowie der gegenstandsbezogenen Angemessenheit und Korrektheit berücksichtigt. Insgesamt zeichnet sich also in beiden Modellierungen ab, dass Unterrichtsqualität nicht mehr in Bezug auf isolierte Dimensionen, sondern in einem umfassenderen Sinne gedacht wird.

2.3 | Zusammenhänge zwischen den Tiefenmerkmalen

Tiefenmerkmale von Unterricht wurden bis in die 2000er-Jahre hinein überwiegend in Form von Merkmalslisten guten Unterrichts gebündelt (vgl. die bis heute populären Werke von Brophy 2000; Helmke 2014; Meyer 2016). Diese Listen erwiesen sich insofern als nützlich, als sie eine Vielzahl korrelativer Einzelstudien sowie Ergebnisse von Metaanalysen im Sinne von Forschungsüberblicken bündelten. Kritisieren lassen sich die Merkmalskataloge allerdings dahingehend, dass die angeführten Qualitätsmerkmale vergleichsweise unverbunden nebeneinander stehenbleiben und das Zusammenwirken der Merkmale nicht hinreichend modelliert wird (vgl. Lotz 2016, 15–16).

Als hilfreich haben sich folglich Arbeiten erwiesen, die gemäß dem *Prinzip der Sparsamkeit* versucht haben, einzelne Qualitätsmerkmale zu Qualitätsdimensionen zusammenzufassen. So resultierten ursprünglich auch die bereits angesprochenen drei Basisdimensionen von Unterricht aus einer Faktorenanalyse, die im Rahmen der TIMSS-Studie durchgeführt wurde

(Klieme et al. 2001). Anzumerken ist jedoch auch bezogen auf solche oder vergleichbare Modellierungen (vgl. international z. B. das CLASS-System bei Pianta/Hamre 2009), dass zwar die Wirkung der einzelnen Dimensionen auf die Kompetenzentwicklung oder motivational-affektive Outcomes deutlich genauer beschrieben ist als bei den Merkmalskatalogen, die Zusammenhänge zwischen den einzelnen Dimensionen aber nur selten spezifiziert und theoretisch fundiert werden (vgl. bzgl. der drei Basisdimensionen etwa Praetorius, Klieme et al., 2020, 25).

Erst in jüngerer Zeit finden sich Qualitätsmodellierungen, die das Verhältnis der unterschiedlichen Qualitätsdimensionen explizit beschreiben. In der Regel handelt es sich um *hierarchische Modelle*, die die Realisierung einzelner Qualitätsdimension als Voraussetzung dafür betrachten, dass sich andere Qualitätsdimensionen bestmöglich entfalten können. Angenommen wird beispielsweise in mehreren Modellen, dass generische Merkmale von Unterricht, wie bspw. die Klassenführung oder das Klassenklima bzw. die sozio-emotionale Unterstützung, wichtige Voraussetzungen darstellen, damit sich fachliche Lernprozesse bestmöglich entfalten können (Brunner 2017, 278; Charalambous/Praetorius 2020, 5; Hesse/Winkler 2022, 20; Lotz 2016, 75–76). Ebenso nehmen einige Modelle an, dass die *Adaptivität* des Unterrichts eine übergeordnete Qualitätsdimension darstellt, die gegeben sein muss, damit die anderen Dimensionen bestmöglich an die Voraussetzungen der Schüler:innen anschließen können (Charalambous/Praetorius 2020, 5; Hesse/Winkler 2022, 10–11). Schließlich wird der *fachlichen Korrektheit* bzw. *Angemessenheit* des Unterrichts ein besonderer Stellenwert eingeräumt, da sich bspw. ein kognitiv aktivierender Unterricht nur dort sinnvoll entfalten kann, wo fachliche korrekte Konzepte vermittelt werden (Brunner 2017).

Empirische Befunde für die theoretischen Annahmen liegen mit Blick auf den Deutschunterricht bislang nur in spärlichem Umfang vor. Dies gilt insbesondere für videografische Untersuchungen. So kann Lotz (2016) im Rahmen der PERLE-Studie im Fach Deutsch Zusammenhänge zwischen den 28 von ihr eingesetzten hoch-inferenten Items zur Beurteilung der Unterrichtsqualität in Leseübungsphasen des Grundschulunterrichts mittels Hauptkomponentenanalyse aufzeigen. Konkret ergeben sich fünf Komponenten,

die sich als das „Schaffen einer lernförderlichen Umgebung“, „Konstruktive Lernunterstützung“, „Förderung einer eigenständigen Auseinandersetzung mit dem Lerngegenstand“, „Anregung von Denkprozessen“ und „Motivierende Haltung der Lehrperson“ beschreiben lassen (Lotz 2016, 434).

Lotz schlussfolgert aus diesem Befund, dass ihrem Modell von Unterrichtsqualität zwar ursprünglich keinerlei Hierarchie eingeschrieben war, sich aber aus den Komponenten heraus durchaus hierarchische Zusammenhänge ablesen ließen (vgl. Lotz 2016, 390–391). Dies betreffe insbesondere die Komponente „Schaffen einer lernförderlichen Umgebung“, die nicht nur eine Kombination aus den Basisdimensionen „Klassenführung“ und „entspannte Lernatmosphäre“ darstelle, sondern darüber hinaus die Grundlage für die anderen, eher auf die kognitive Aktivierung ausgerichteten Komponenten zu sein scheint.

Für den Rechtschreibunterricht gehen Wiprächtiger et al. (2023) in einer jüngeren Untersuchung im Rahmen der Studie Profess-R der Frage nach, ob sich „Qualitätsprofile“ innerhalb der 44 videografierten Stunden zur Doppelkonsonantenschreibung aus Deutschland und der Schweiz ausmachen lassen. Hierdurch wollen sie analog zum vorliegenden Beitrag einem bloß variablenzentrierten Ansatz der Unterrichtsforschung entgegen, der unberücksichtigt lässt, „dass die einzelnen Skalen oder Variablen nicht für sich allein stehen und wirken“ (Wiprächtiger et al. 2023, 230). Im Ergebnis zeigen sich mittels clusteranalytischer Verfahren drei verschiedene Qualitätsprofile von Rechtschreibunterricht, die alle ein eher geringes Potenzial zu kognitiver Aktivierung und Unterstützung aufweisen, sich aber z. T. signifikant darin unterscheiden, inwiefern der Unterricht fehlerbehaftet oder der Lerngegenstand verkürzt dargestellt ist. Zudem lassen sich zwei der Profile dadurch differenzieren, dass sie entweder das Vorwissen der Schüler:innen anregen *oder* aktivierende Aufgaben einsetzen. Aus diesen Befunden lässt sich schlussfolgern, dass durchaus differenzielle Zusammenhänge zwischen einzelnen Items kognitiver Aktivierung bzw. Unterstützung sowie einer fachlich angemessenen Auswahl und Aufbereitung von Lerninhalten bestehen, aber nicht per se davon ausgegangen werden kann,

dass eine aktivierende Unterrichtsgestaltung mit einem fokussierten bzw. fehlerfreien Unterrichtsangebot einhergeht. Dies belegen nicht zuletzt auch die geringen Korrelationen der Skalen zur kognitiven Aktivierung, Unterstützung und inhaltlichen Fokussierung (vgl. Wiprächtiger et al. 2023, 234).

Mit Blick auf den Literaturunterricht stehen Untersuchungen, die über Einzelstudien hinaus nach Unterrichtsqualität in alltäglichem (d. h. nicht durch gezielte Interventionen oder Modellerprobungen geprägten) Unterricht betrachten, gegenwärtig weitgehend aus (Wieser 2019). Entsprechend stellen auch Forschungen zum Zusammenhang verschiedener Qualitätsmerkmale und -dimensionen ein wichtiges Desiderat dar.

2.4 | Zusammenhänge zwischen Tiefen- und Oberflächenmerkmalen

Zum Zusammenhang zwischen Tiefen- und Oberflächenmerkmalen des Unterrichts liegen in der Literaturdidaktik bislang nur wenige Studien vor. Vielmehr war die literaturdidaktische Forschung lange Zeit von der Annahme geprägt, dass bestimmte Gestaltungsweisen des Literaturunterrichts direkt mit dem Lernen der Schüler:innen zusammenhängen. Dies zeigt sich beispielsweise in der Studie „Literaturunterricht kontrastiv“ (Fritzsche et al. 2006), in welcher die Wirksamkeit analytischer vs. handlungs- und produktionsorientierter Methoden untersucht wurde, ohne die Prozessqualitäten des Unterrichts zu bestimmen. Aber auch die Forschung zu Unterrichtsgesprächen im Literaturunterricht neigt(e) dazu, eher offene und eher gelenkte Unterrichtsgespräche (z. B. Heidelberger Modell vs. fragend-entwickelnder Unterricht) pauschal gegenüberzustellen, ohne ihr aktivierendes bzw. unterstützendes Potenzial auf der Tiefenebene zu betrachten (vgl. im Überblick Magirius et al. 2022).

In der Bildungsforschung besteht indes eine weit verbreitete Annahme darin, dass bei einer Betrachtung der Unterrichtsqualität zuvorderst die Tiefenmerkmale des Unterrichts zu betrachten sind (siehe Kap. 2.1).² Diese werden in Anlehnung an Oser und Baeriswyls (2001) Theorie der *Choreografies of Teaching* meist als unabhängig von den Oberflächen- bzw. Sichtmerkmalen des Unterrichts betrachtet:

[T]he visible structure is the free part of the choreography – the flexible, the exchangeable, the part that is continually newly adapted by and for learners. It includes methods, social forms, context, representations, teaching styles, learning styles, function rhythms, media, control forms, and so forth. The visible structure represents the free and freely structurable moment in learning. (Oser/Baeriswyl 2001, 1043)

In dieser Perspektive erscheinen die Merkmale der Oberflächenebene des Unterrichts also als nahezu beliebig, solange sie das Lernen der Schüler:innen auf Tiefenebene nicht behindern (vgl. Oser/Baeriswyl 2001, 1048). Demgegenüber nehmen neuere Arbeiten die oben vertretenen Annahmen relativierend an,

dass Oberflächenmerkmale des Unterrichts einen gewissen Rahmen bereitstellen, sodass die Qualität von Unterricht bzw. deren Beurteilung im Rahmen von (videobasierten) Unterrichtsbeobachtungen auch von Aspekten auf Ebene der Oberfläche abhängig sein dürfte (Hess/Lipowsky 2020, 117).

Belege für die letztere Auffassung lassen sich bezogen auf videografische Untersuchungen im Deutschunterricht erneut in der PERLE-Studie (Hess/Lipowsky 2020; Lotz 2016, Kap. 11.2) sowie im Profess-R-Projekt (Riegler et al. 2022) ausmachen. So können Hess und Lipowsky (2020) auf Basis von PERLE-Daten zeigen, dass Art und Anzahl der von Lehrpersonen an die Schüler:innen gerichteten Fragen von der Sozialform abhängt: Lehrpersonen stellen im klassenöffentlichen Unterricht nicht nur mehr Fragen an die Schüler:innen, sondern richten ihre Fragen auch häufiger auf den Unterrichtsgegenstand. In Schüler:innenarbeitsphasen werden demgegenüber häufiger ablaufbezogene Fragen gestellt. Dass jedoch von der Wahl bestimmter Sozialformen nicht automatisch auf das Vorliegen bestimmter Unterrichtsqualitäten geschlossen werden kann, zeigt Lotz (2016) in vorangegangenen Untersuchungen zum PERLE-Korpus. In diesen

² Insofern diese Annahme überwiegend auf Untersuchungen zu mathematisch-naturwissenschaftlichem Unterricht beruht, muss zukünftig noch empirisch abgesichert werden, ob und welche Tiefenmerkmale des Unterrichts kognitive und motivational-affektive Wirkungen hervorrufen (vgl. zum Forschungsstand Hesse/Winkler 2022)

wird nämlich deutlich, „dass es keine Übungsphase oder Sozialform gibt, die mit allen oder der Vielzahl der Komponenten [gemeint sind Qualitätsdimensionen, F. H.] in derselben Richtung zusammenhängt“ (Lotz 2016, 435).

In dieser Hinsicht differenzielle Zusammenhänge finden sich auch in der Profess-R-Studie. Auch hier zeigt sich zunächst, dass für das Gros der sog. Sachstrukturen³ des Unterrichts keine Zusammenhänge mit den Qualitätsmerkmalen des Unterrichts gefunden werden können. Dennoch gehen die Autor:innen davon aus, „dass die verschiedenen Sachstrukturmuster unterschiedliche Anforderungen an die Realisierung fachlich ‚guten‘ Unterrichts stellen“ (Riegler et al. 2022, 20). Dies lässt sich u. a. daran verdeutlichen, dass bestimmte Sachstrukturmuster (z. B. quantitätsbasierter Ansatz) inhaltliche Verkürzungen und fachliche Fehler statistisch eher nahelegen als andere. Entsprechend lässt sich auch im Projekt Profess-R die rahmende Funktion von Oberflächenmerkmalen des Unterrichts punktuell gut erkennen.

3 | Methodisches Vorgehen

3.1 | Forschungsfragen und Hypothesen

Basierend auf den vorangegangenen Darstellungen beschäftigt sich der vorliegende Beitrag mit Zusammenhängen unterschiedlicher Qualitätsmerkmale von Literaturunterricht sowie mit deren Abhängigkeit von der Oberflächengestaltung des Unterrichts. Konkret werden dabei folgende Forschungsfragen gestellt:

- 1| Lassen sich in der betrachteten Stichprobe Zusammenhänge zwischen den unterschiedlichen Tiefenmerkmalen des Literaturunterrichts identifizieren?
- 2| Bestehen in der betrachteten Stichprobe Zusammenhänge zwischen den Tiefen- und Oberflächenmerkmalen von Literaturunterricht?

Hinsichtlich der ersten Forschungsfrage lässt sich unter Berücksichtigung der vorangegangenen Darstellungen begründet annehmen, dass insbesondere fachübergreifend relevante Qualitätsdimensionen (Klassenführung, sozio-emotionale Unterstützung) im Sinne einer breiter gefassten ‚positiven Lernatmosphäre‘ in engem Zusammenhang stehen. Bei den Korrelationen zwischen fachbezogenen bzw. -spezifischen Qualitätsdimensionen ist eine Hypothesenbildung demgegenüber schwieriger, da hier in ähnlich gelagerten Studien vor allem differenzielle Befunde nachgewiesen wurden. Denkbar wäre einerseits im Sinne der oben geschilderten hierarchischen Annahmen, dass ein kognitiv-emotional aktivierender Unterricht insbesondere dort auftritt, wo Rater:innen dem Unterricht auch eine fachlich angemessene Auswahl und Aufbereitung von Lerngegenständen und Fachmethoden attestieren. Andererseits zeigen etwa die Befunde aus Profess-R, dass kognitiv aktivierende Lerngelegenheiten auch mit einer fachlich fehlerhaften Unterrichtsgestaltung einhergehen können. Ähnlich differenzielle Zusammenhangsmuster sind schließlich auch bezüglich der Verbindung von fachübergreifenden und fachbezogenen Qualitätsdimensionen denkbar.

Hinsichtlich der zweiten Forschungsfrage wird entlang der theoretischen Annahmen (Abschn. 2.4) prinzipiell davon ausgegangen, dass keine Zusammenhänge zwischen den untersuchten Oberflächenmerkmalen und den Tiefenmerkmalen des Unterrichts bestehen. Konkret wird dabei hinsichtlich der untersuchten Sozialformen angenommen, dass die angebotsseitige Unterrichtsqualität nicht davon abhängig ist, ob etwa Klassengespräche oder Schüler:innenarbeitsphasen den Hauptbestandteil der Unterrichtsgestaltung ausmachen. Bezüglich der untersuchten inhaltsbezogenen Aktivitäten können aufgrund mangelnder Befunde keine gerichteten

³ Die Ebene der Sachstrukturen des Unterrichts wurde im Rahmen des Projekts Profess-R als eine zusätzliche Ebene zwischen Oberflächen- und Tiefenstrukturen eingezogen. Erfasst werden damit die fachliche Modellierung (quantitätsbasierter Ansatz, silbenbasierter Ansatz usw.) und der didaktische Zugriff (regelerorientiert, strategieorientiert usw.) auf den Gegenstand (Riegler et al. 2020, 56-58).

Hypothesen formuliert werden. Zwar kann davon ausgegangen werden, dass die angebotsseitige Unterrichtsqualität zunächst einmal davon unberührt ist, ob sich die Lernenden im Unterricht schwerpunktmäßig bspw. mit der Handlung des Textes, seinen Figuren oder Textsortenmerkmalen auseinandersetzen, da aufgrund des quantitativen Ausmaßes der Beschäftigung mit (Kon)Textelementen nur schwer auf die die qualitätsvolle Ausgestaltung entsprechender Phasen geschlossen werden kann. Allerdings lässt sich auch nicht ausschließen, dass ein (Nicht-)Vorkommen bestimmter inhaltsbezogener Aktivitäten ‚guten‘ Literaturunterricht zumindest (un-)wahrscheinlicher macht. Zu denken wäre hier etwa an ein Mindestmaß an Aktivitäten, die literarische Wertungsprozesse aufgreifen oder neben der Besprechung inhaltlicher Textaspekte auch die Darstellungsseite des Textes berühren (z. B. sprachliche Gestaltung, Erzählperspektive o.ä. betreffen). Da entsprechende Zusammenhänge in der literaturdidaktischen Unterrichtsforschung bislang kaum untersucht wurden, haben die Analysen gerade in diesem Bereich also vorrangig explorativen Charakter.

3.2 | Datengrundlage

Zur Überprüfung der Forschungsfragen und Hypothesen wird auf ein Korpus von videografieren Literaturstunden zurückgegriffen, die von Jenaer Studierenden im Praxissemester innerhalb des Projekts OVID-PRAX gehalten wurden. Ziel dieser Interventionsstudie war es, die Wirksamkeit onlinebasierter Videofeeds auf die Entwicklung professioneller Kompetenzen zu untersuchen. Zu diesem Zweck sollten insgesamt 38 Studierende einer Videogruppe⁴ eine selbstgewählte Unterrichtsstunde videografieren und einen Ausschnitt der Unterrichtsstunde zu Feedbackzwecken auf einer Moodle-Plattform hochladen. Während im Projekt OVID-PRAX also schwerpunktmäßig die Entwicklung professioneller Kompetenz durch das Feedback zu den Unterrichtsausschnitten im Zentrum stand (Gröschner et al., 2019), konzentrierte sich die vorliegende Untersuchung auf die Analyse der Unterrichtsvideos selbst. Berücksichtigt wurden dabei allerdings nicht nur die Videoausschnitte, sondern die vollständigen Unterrichtsstunden (je ca. 45–90 Minuten), die von insgesamt 22 Studierenden vorlagen (N = 22 Videos).

Inhaltlich konnten die Studierenden weitestgehend frei entscheiden, wie sie die Unterrichtsstunde konzipieren. Die einzige Vorgabe bestand darin, dass die Studierenden einen kurzen epischen Text verhandeln und mindestens ein Plenumsgespräch realisieren sollten. Letztere Vorgabe ist allerdings nicht überzubewerten, da die Studierenden Zweck und Länge der Plenumsphase selbst festlegen durften. Ebenfalls keine Vorgaben bestanden hinsichtlich der zu unterrichtenden Klassenstufe, da die Studierenden sich ihre Klassen in der Regel nicht aussuchen können, sondern diese von den Praktikumsschulen zugewiesen werden. Entsprechend beinhaltet die Stichprobe Unterrichtsstunden in verschiedenen Klassenstufen der Sekundarstufe I und II.⁵

3.3 | Rating der Tiefenmerkmale

Die Einschätzung der Tiefenmerkmale des Unterrichts orientierte sich an den drei Basisdimensionen von Unterrichtsqualität (Praetorius et al. 2018; Praetorius, Klieme et al., 2020), die – wie oben bereits erläutert – allerdings im Rahmen der Studie von Hesse (2024) fachlich profiliert wurde. So wurden neben den beiden generischen Dimensionen „Klassenführung“ und „sozio-emotionale Unterstützung“ auch die beiden fachbezogenen Dimensionen „Auswahl und Aufbereitung von Lerngegenständen und Fachmethoden“ sowie „kognitiv-emotionale Aktivierung“ kriteriengeleitet eingeschätzt. Für die beiden generischen Dimensionen wurden insgesamt 12

⁴ Die Gruppe setzte sich aus allen Teilnehmenden von zwei Praktikumsbegleitseminaren in der Deutschdidaktik zusammen, die im Wintersemester 2017/2018 und im Sommersemester 2018 gehalten wurden. Entsprechend kann davon ausgegangen werden, dass die Stichprobe das gesamte Leistungsspektrum abdeckt.

⁵ Grundschulunterricht kommt in der Stichprobe nicht vor, da in Jena nur Studierende für Gymnasien und Regelschulen ausgebildet werden.

Ratingitems aus bestehenden Studien adaptiert (z. B. Gabriel/Lipowsky 2013a, 2013b; Pianta et al. 2012; Thiel et al. 2013), für die beiden fachlichen Dimensionen wurden 11 Ratingitems neu entwickelt. Im Sinne der Anschaulichkeit werden nachfolgend nur die Namen der jeweiligen Items präsentiert, um eine Vorstellung von der inhaltlichen Ausrichtung der vier Qualitätsdimensionen zu vermitteln (vgl. Tabelle 1). Zusätzlich hingewiesen sei hier lediglich auf den im Folgenden noch relevanten Aspekt, dass es sich bei der kognitiv-emotionalen Aktivierung um ein inhaltlich breit gefächertes Konstrukt handelt, das Elemente der Aktivierung i. e. S. ebenso umfasst wie Aspekte der gegenstandsbezogenen Unterstützung. Eine ausführliche Herleitung der Qualitätsdimensionen und -merkmale inkl. dem zugehörigen Ratingmanual findet sich bei Hesse (2024).

<i>Qualitätsdimension 1: Kognitiv-emotionale Aktivierung</i>	
KEA_1	Anregung zu genauer Textwahrnehmung
KEA_2	Anregung zu subjektiver Involviertheit
KEA_3	Anregung des Wechselspiels genauer Textwahrnehmung und subjektiver Involviertheit
KEA_4	Anregung zur Reflexion des Textverstehensprozesses
KEA_5	Anregung zur Elaboration von Schüler:innenäußerungen
KEA_6	Bereitstellung von aufgaben(set)immanentem Support
KEA_7	Bereitstellung von interaktivem Verstehenssupport
<i>Qualitätsdimension 2: Auswahl + Aufbereitung von Lerngegenständen und Fachmethoden</i>	
AUS_1	Textauswahl und -einsatz
AUS_2	Strukturierung des Unterrichts
AUS_3	fachliche Korrektheit und Angemessenheit
AUS_4	Passung der Methoden und Sozialformen
<i>Qualitätsdimension 3: Klassenführung</i>	
KF_1	Allgegenwärtigkeit der Lehrperson
KF_2	Gruppenmobilisierung
KF_3	Regelklarheit und -anwendung
KF_4	effektiver Umgang mit auftretenden Störungen
KF_5	Störungsmanagement
KF_6	Zeitmanagement
KF_7	Übergangsmanagement
KF_8	Vorbereitung des Klassenraums, der Technik und der Medien
<i>Qualitätsdimension 4: sozio-emotionale Unterstützung</i>	
SEU_1	positive Ausstrahlung der Lehrperson
SEU_2	Wertschätzung der Schüler:innen durch die Lehrperson
SEU_3	positiver Umgang mit Verstehensschwierigkeiten und Fehlern
SEU_4	entspannte Lernatmosphäre

Tabelle 1. Qualitätsdimensionen und -merkmale im Überblick

Bei der Adaption bzw. Konzeption der Items wurde darauf geachtet, dass hohe Ausprägungen eines Items für eine frequente und zugleich adaptive Realisierung des betreffenden Qualitätsmerkmals in der Stunde stehen.⁶ Alle Items wurden vierstufig eingeschätzt, wobei Stufe 4 für

⁶ Dies wird etwa am Beispiel „Anregung zur Reflexion des Textverstehensprozesses“ deutlich. Bei diesem Item geht es sicher darum, dass im Unterricht zu einem gewissen Grad über die Generierung von Interpretationen oder Interpretations-

eine hohe bzw. adaptive Ausprägung des Qualitätsmerkmals steht, während Stufe 1 eine geringe bzw. wenige adaptive Ausprägung des Merkmals abbildet. Pro Qualitätsmerkmal (Item) wurde für jede Stunde exakt ein Urteil gefällt. Dabei wurde jede Stunde von drei unabhängigen Rater:innen (Autor der Studie + zwei geschulte studentische Hilfskräfte) eingeschätzt, um die Reliabilität der Urteile zu prüfen. Abgesehen von einem Item („Anregung zur Elaboration von Schüler:innenäußerungen“) konnten hierbei gemessen an den hohen Anforderungen der Ratingsysteme durchweg zufriedenstellende Inter-Rater-Reliabilitäten für die Items ermittelt werden (.722 > ICC⁷ > .907). Für die weiteren Berechnungen wurde jeweils auf die Einschätzungen des Autors zurückgegriffen.

3.4 | Kodierung der Oberflächenmerkmale

Als Oberflächenmerkmale des Unterrichts werden im vorliegenden Beitrag die inhaltsbezogenen Aktivitäten und die Sozialformen des Unterrichts näher betrachtet.⁸ Erstere geben Auskunft darüber, *was* im Mittelpunkt der unterrichtlichen Betrachtung steht. Kodiert wird hier etwa, ob in der Unterrichtsstunde schwerpunktmäßig die Inhalte des literarischen Textes (Handlungen, Figuren), die Art der Darstellung (Stil, Metaphorik, Symbolik usw.), Kontexte, Wertungen usw. Raum einnehmen. Bei den Sozialformen steht demgegenüber die Frage im Zentrum, *wie* die Lehrpersonen und die Lernenden interagieren (ob im Klassengespräch, Einzelarbeit, Partner:innenarbeit usw.). Die jeweils verwendeten Kategoriensysteme werden in Tabelle 2 und Tabelle 3 überblicksartig vorgestellt. Ausführliche Kategorienbeschreibungen finden sich bei Hesse (2024).

<i>Oberkategorie</i>	<i>ggf. Unterkategorien</i>	<i>Code</i>	<i>Ankerbeispiele (Auswahl)</i>
Auseinandersetzung mit ... dem Textinhalt	der Handlung	IA_1.1	<ul style="list-style-type: none"> ▪ die Handlung eines Textes wiedergeben ▪ zerschnittene Textteile zusammenfügen
	den Figuren	IA_1.2	<ul style="list-style-type: none"> ▪ eine Figur direkt oder indirekt charakterisieren
	unbekannten/nicht verstandenen Wörtern oder Sätzen	IA_1.3	<ul style="list-style-type: none"> ▪ „Gibt es Wörter, die ihr nicht verstanden habt?“ als Frage im Anschluss an die Textlektüre klären
der literarischen Gestaltung des Textes	Form und Stil des Textes	IA_2.1	<ul style="list-style-type: none"> ▪ die Satzstruktur einer Kurzgeschichte beschreiben ▪ über das verwendete sprachliche Register (z. B. Alltagssprache) diskutieren
	der Symbolik, Metaphorik, Parabolik	IA_2.2	<ul style="list-style-type: none"> ▪ verschiedene Bedeutungsvarianten zentraler Symbole (z. B. der Streusel-schnecke) erschließen
	der Fiktionalität des Textes	IA_2.3	<ul style="list-style-type: none"> ▪ darüber diskutieren, ob es sich bei den Figuren um real existierende Personen handelt
	der Erzählperspektive	IA_2.4	<ul style="list-style-type: none"> ▪ die Erzählperspektive nach Stanzl bestimmen ▪ den Einfluss der Erzählperspektive auf den Text untersuchen
Kontexten	Gattungen / Genres	IA_3.1	<ul style="list-style-type: none"> ▪ Gattungsmerkmale nennen und am Text nachweisen

hypothesen nachgedacht wird. Zugleich muss das Ausmaß von Reflexionsgelegenheiten in einem angemessenen Verhältnis zu erarbeitenden Unterrichtsphasen stehen. Darüber hinaus muss ein hohes Rating dieses Items abbilden, dass auch diejenigen Aspekte durch die Lehrkraft und die SuS reflektiert werden, die für die Unterrichtsstunde zentral waren.

⁷ Zur Berechnung der Inter-Rater-Reliabilität wurden zweifaktorielle Intra-Klassen-Korrelationen auf Basis unjustierter Einzelwerte (Wirtz/Caspar 2002) nach 10 und nach 22 Videos berechnet.

⁸ Die im Projekt ebenfalls kodierten methodischen Zugriffsweisen im Literaturunterricht (poetische vs. diskursive Aufgaben) konnten für den vorliegenden Beitrag bedauerlicherweise nicht hinsichtlich ihrer Abhängigkeit von den Tiefenmerkmalen untersucht werden, da insbesondere die poetischen Methoden im Videokorpus in zu geringer Häufigkeit auftraten.

der Epoche	IA_3.2	▪ einen Lehrbuchtext zur Epoche des Textes lesen
den Autor:innen	IA_3.3	▪ wichtige biografische Eckdaten der Autor:innen nennen
dem Thema	IA_3.4	▪ vor der Lektüre des Textes <i>Nacht</i> eine Mindmap zum Thema „Nacht“ oder „Alltag“ anfertigen, um thematisches Wissen zu aktivieren
sonstigem Kontextwissen	IA_3.5	▪ weiterführendes poetologisches, geistes-, sozial- oder kulturwissenschaftliches ▪ Wissen an einen Text herantragen (z. B. Schulz von Thun)
Verfahren der Texterschließung	IA_4	▪ im Unterricht besprechen, wie eine Erörterung nach dem Sanduhren-Prinzip geschrieben werden kann
Wertungen	IA_5	▪ den ersten Leseindruck formulieren ▪ einen Text kriteriengeleitet mittels einer Rezension einschätzen
der Rezeption des Textes	IA_6	▪ einen Text reihum laut vorlesen
sonstigen, nicht auf das literarische Lernen zielenden inhaltsbezogenen Aktivitäten	IA_7	▪ mittels eines Kriterienkataloges das Halten von Referaten üben (KB Sprechen und Zuhören) ▪ an einem Text grammatische Phänomene bestimmen (KB Sprache und Sprachgebrauch untersuchen)
Instruktionen	IA_8	▪ ausgedehnte Aufgabenerklärungen ▪ Erteilung von Hausaufgaben
Übergängen und Organisatorischem	IA_9	▪ Besprechung organisatorischer Angelegenheiten ▪ Übergänge zwischen IAs
einem Mix aus IAs	IA_10	▪ differenzierende Aufgabenstellungen in einer Gruppenarbeitsphase

Tabelle 2. Kategoriensystem: Inhaltsbezogene Aktivitäten

<i>Oberkategorie</i>	<i>Unterkategorie</i>	<i>Code</i>
klassenöffentlicher Unterricht	Lehrer:innenvortrag	SF_1.1
	Schüler:innenvortrag	SF_1.2
	Diktat	SF_1.3
	Klassengespräch	SF_1.4
Schüler:innenarbeitsphasen	Einzelarbeit	SF_2.1
	Partner:innenarbeit	SF_2.2
	Gruppenarbeit	SF_2.3
Sonstige	Mix verschiedener Sozialformen	SF_3.1
	Übergänge und Organisatorisches	SF_3.2

Tabelle 3. Kategoriensystem: Sozialformen

Die Kodierung beider Kategoriensysteme erfolgte im Time-Sampling-Verfahren. Damit ist gemeint, dass die Unterrichtsstunde vor der Kodierung in gleichgroße Intervalle von jeweils zehn Sekunden eingeteilt wurde. Die Aufgabe der Kodierenden bestand dann darin, für jedes Intervall genau *eine* Kategorie zu vergeben, die sich im betrachteten Intervall am deutlichsten zeigt. Auf diese Weise konnten die Anteile bestimmter inhaltsbezogener Aktivitäten bzw. Sozialformen an der Gesamtdauer der Unterrichtsstunde bestimmt werden. Zugleich war mit der Entscheidung für immer genau eine Kategorie insofern ein Informationsverlust verbunden, als kleinste Veränderungen unberücksichtigt bleiben (Seidel 2003, 116).

Wie auch bei den Tiefenmerkmalen des Unterrichts wurde die Güte der entwickelten Kodiersysteme u. a. durch eine Bestimmung der Beobachtungsübereinstimmung bei insgesamt

drei Kodierenden (Autor + zwei geschulte stud. Hilfskräfte) auf Grundlage von ca. 6000 Intervallen pro Kategoriensystem überprüft. Als statistische Kenngröße fungierte hier Lights Kappa (= gemitteltes Cohens Kappa), das als Maß für Beobachtungsübereinstimmung bei mehr als zwei Kodierer:innen empfohlen wird (Wirtz/Caspar 2002). Auch hier zeigten sich durchweg zufriedenstellende Werte ($.89 < \kappa_{\text{Light}} < .93$).

3.5 | Bestimmung der Zusammenhänge

Um im Sinne der ersten Forschungsfrage herauszufinden, wie die Tiefenmerkmale des Unterrichts miteinander zusammenhängen, wurden im vorliegenden Beitrag nach einer Überprüfung der internen Konsistenz mittels Cronbachs Alpha (Döring/Bortz 2016, 467–468) der unterschiedlichen Qualitätsdimensionen Spearman-Korrelationen⁹ zwischen den verschiedenen Dimensionen berechnet. Auf diese Weise sollte sichtbar werden, ob ein gutes/schlechtes Abschneiden einer Qualitätsdimension auch mit einem guten/schlechten Abschneiden in einer anderen Dimension assoziiert ist.

Mittels der *zweiten Forschungsfrage* sollte darüber hinaus überprüft werden, inwiefern sich die Ausprägungen der Qualitätsdimensionen in Abhängigkeit von unterschiedlichen Oberflächenstrukturen unterscheidet. Hierzu wurden die Stunden bezogen auf die Ausprägung der Oberflächenmerkmale agglomerativ geclustert, um zunächst Gemeinsamkeiten von Oberflächenmerkmalen innerhalb einer Kategorie festzustellen (Backhaus et al. 2016; Schendera 2010).¹⁰ Für die inhaltsbezogenen Aktivitäten ergab sich hieraus ein handlungs-, ein figuren- und ein textsortenorientiertes Cluster¹¹, für die Sozialformen ein gesprächs- und vortragsorientiertes, ein an Schüler:innenarbeitsphasen orientiertes und ein gesprächsdominantes Cluster¹² (vgl. Hesse 2024, für eine ausführliche Beschreibung und Diskussion). Um zu überprüfen, inwiefern ein Auftreten bestimmter Tiefenmerkmale des Unterrichts von diesen Clustern abhängen, wurden Kruskal-Wallis-Tests berechnet. Diese Tests nehmen die o. g. Cluster als unabhängige Variable an und prüfen, ob sich die Tiefenmerkmale des Unterrichts in ihrem mittleren Rang je nach Clusterzugehörigkeit unterscheiden. Der Kruskal-Wallis-Test stellt die nicht-parametrische Alternative zur ANOVA dar und wurde hier eingesetzt, da ein Shapiro-Wilk-Test zeigte, dass die Tiefenmerkmale nur in den Dimensionen „kognitiv-emotionale Aktivierung“ und „Auswahl und Aufbereitung der Lerngegenstände und Fachmethoden“ normalverteilt waren ($p > 0.05$) und entsprechend nicht alle Dimensionen die Voraussetzungen zur Durchführung einer ANOVA erfüllten.

⁹ Es wurde auf Spearman-Korrelationen zurückgegriffen, da nicht durchgehend von einer Normalverteilung der Daten und einem linearen Zusammenhang ausgegangen werden konnte, wie ihn beispielsweise die Pearson-Korrelation annimmt (Schober et al. 2018).

¹⁰ Die Anzahl der Cluster wurde mittels Elbow-Kriterium und Test von Mojena bestimmt (Backhaus et al. 2016, 494–497), ihre Intra-Cluster-Homogenität über die Berechnung des F-Werts überprüft (ebd., 507). Diese und alle folgenden Berechnungen wurden mit SPSS 28 ermittelt (IBM Corp 2021).

¹¹ Wie der Name der Cluster bereits vermuten lässt, zeichnen sich diese dadurch aus, dass in den jeweiligen Unterrichtsstunden schwerpunktmäßig die Handlung des literarischen Textes oder dessen Figuren besprochen wurden bzw. Textsortenwissen vermittelt, reaktiviert oder (merkmalslistenartig) auf den Text angewandt wurde.

¹² Diese Cluster zeichnen sich dadurch aus, dass sich Gesprächsphasen und Lehrer:innenvorträge die Waage halten oder Schüler:innenarbeits- bzw. Gesprächsphasen überwiegen.

4 | Ergebnisse

4.1 | Deskriptive Statistiken

In Tabelle 4 werden zunächst die deskriptiven Statistiken sowie die interne Konsistenz der unterschiedlichen Qualitätsdimensionen (Skalen) berichtet.

<i>Dimension</i>	<i>Items</i>	<i>N</i>	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>	<i>α</i>
Kognitiv-emotionale Aktivierung	7	22	2.32	.54	1.29	3.29	.748
Auswahl und Aufbereitung der Lerngegenstände	4	22	2.60	.65	1.50	3.75	.825
Klassenführung	8	22	3.35	.63	2.13	4.00	.930
Sozio-emotionale Unterstützung	4	22	3.48	.48	2.50	4.00	.746

Tabelle 4. Deskriptive Statistiken und interne Konsistenz der Qualitätsdimensionen

Hierbei zeigt sich einerseits, dass die Studierenden im Mittel schlechter in den beiden fachlichen Qualitätsdimensionen abschneiden als in den beiden generischen. Andererseits lassen die hohen Standardabweichungen und Spannweiten den Schluss zu, dass die Qualität der Lernangebote insgesamt als sehr heterogen zu beurteilen ist.¹³ Hervorzuheben ist darüber hinaus, dass alle eingesetzten Skalen zufriedenstellende interne Konsistenzen aufweisen ($\alpha > .70$), so dass nachfolgend Berechnungen auf Skalenebene erfolgen können.

4.2 | Zusammenhänge zwischen den Tiefenmerkmalen

Bezogen auf die erste Forschungsfrage wurde mithilfe von Spearman-Korrelationen überprüft, wie die Tiefenmerkmale des Unterrichts miteinander zusammenhängen (vgl. Tabelle 5).

	1	2	3	4
1. Klassenführung	1.00			
2. sozio-emotionale Unterstützung	.567**	1.00		
3. Auswahl & Aufbereitung von Lerngegenständen	.107	-.073	1.00	
4. kognitiv-emotionale Aktivierung	.213	.136	.502*	1.00

Tabelle 5. Spearman-Korrelationen der Tiefenmerkmale des Unterrichts

Deutlich wird hier einerseits, dass hypothesenkonform ein starker und zugleich hochsignifikanter Zusammenhang zwischen den beiden generischen Dimensionen, Klassenführung und sozio-emotionale Unterstützung, besteht. Andererseits zeigt sich, dass die beiden fachlichen Dimensionen, Auswahl und Aufbereitung der Lerngegenstände und Fachmethoden sowie kognitiv-emotionale Aktivierung, signifikant korrelieren. Demgegenüber bestehen zwischen den generischen und fachlichen Qualitätsdimensionen allenfalls schwache und durchweg nicht signifikante Zusammenhänge.

Da aus fachdidaktischer Perspektive besonders interessant ist, wie die Zusammenhänge zwischen den fachlichen Qualitätsdimensionen im Detail aussehen, wurden hier nochmals feinkörnigere Korrelationen berechnet. Dazu wurde die kognitiv-emotionale Aktivierung in eine Subdimension ‚Aktivierung i. e. S.‘ mit den Items „KEA_1“ bis „KEA_5“ sowie eine Subdimension

¹³ Eine Ausnahme bildet hier allenfalls die sozio-emotionale Unterstützung, bei der keine Stunde den theoretischen Skalenmittelwert von 2.5 unterschreitet.

„Unterstützung“ mit den supportbezogenen Items „KEA 6“ und „KEA 7“ aufgeteilt¹⁴ (vgl. Tabelle 1, oben) und mit den Einzelitems der Dimension „Auswahl und Aufbereitung der Lerngegenstände und Fachmethoden“ korreliert (vgl. Tabelle 6).

Qualitätsdimension	1	2	3	4	5	6
1. Textauswahl und -einsatz	1.00					
2. Strukturiertheit	.502*	1.00				
3. Korrektheit / Angemessenheit	.727*	.460*	1.00			
4. Passung Methoden Sozialformen	.379	.652**	.448*	1.00		
5. kognitiv-emotionale Aktivierung i. e. S.	.316	.367	.416	.544*	1.00	
6. kognitiv-emotionale Unterstützung	.476*	.213	.505*	.435*	.694*	1.00

Tabelle 6. Vertiefende Spearman-Korrelationen zwischen den fachlichen Qualitätsdimensionen

Sieht man von den (hoch-)signifikanten Korrelationen zwischen den Items derselben Dimension ab (grau hinterlegt), die angesichts der hohen Cronbachs Alpha-Werte zu erwarten waren, zeigen sich signifikante Zusammenhänge mittlerer Größe vor allem zwischen der kognitiv-emotionalen Unterstützung und den Einzelitems der „Auswahl und Aufbereitung der Lerngegenstände und Fachmethoden“ (orange hinterlegt). Hier lässt sich feststellen, dass die Rater:innen den Unterricht umso unterstützender wahrnehmen, desto besser die Besonderheiten des literarischen Textes berücksichtigt, der Unterricht fachlich korrekt bzw. angemessen gestaltet und Methoden und Sozialformen in Passung zu Lernzielen, Lernvoraussetzungen und Text gebracht werden (und umgekehrt). Das Potenzial zu kognitiv-emotionaler Aktivierung i. e. S. korreliert hingegen ausschließlich mit der Passung der Methoden und Sozialformen signifikant (blau hinterlegt).

4.3 | Zusammenhänge zwischen den Oberflächen- und Tiefenmerkmalen

Ob sich die Ausprägung der Tiefenmerkmale in Abhängigkeit von bestimmten Oberflächenstrukturen des Unterrichts unterscheidet, wurde mithilfe des Kruskal-Wallis-Tests geprüft. Als unabhängige Variablen wurden dabei die durch Clusteranalysen ermittelten Cluster im Bereich der inhaltsbezogenen Aktivitäten und Sozialformen auf Oberflächenebene angenommen; als abhängige Variablen fungierten die Skalenmittelwerte der vier Qualitätsdimensionen auf Tiefenebene. Für die Cluster beider Variablen wurde argumentiert, dass Zusammenhänge mit angebotsseitigen Qualitätsmerkmalen des Unterrichts zwar nicht ausgeschlossen, aber doch eher unwahrscheinlich sind. Vielmehr wurde davon ausgegangen, dass sowohl die Sozialform-Cluster (z. B. an Schüler:innenarbeitsphasen orientierter Unterricht) als auch die Cluster der inhaltsbezogenen Aktivitäten (z. B. an der Handlung des Textes ausgerichteter Unterricht) auf der Tiefenebene des Unterrichts je unterschiedlich ausgestaltet werden können.

In Tabelle 7 zeigt sich, dass die angenommenen Hypothesen bestätigt werden können. Sowohl für die Cluster der Sozialformen als auch für die Cluster der inhaltsbezogenen Aktivitäten zeigt sich, dass jeweils über alle vier Qualitätsdimensionen hinweg keine signifikanten Zusammenhänge bestehen. Zumindest mit Blick auf die hier gefundenen Cluster scheint die angebotsseitige Unterrichtsqualität also unabhängig von der Oberflächengestaltung zu sein.

¹⁴ Limitierend muss angemerkt werden, dass die internen Konsistenzen mit $\alpha = .708$ für die Skala kognitiv-emotionale Aktivierung i.e.S. sowie $\alpha = .623$ für die Skala kognitiv-emotionale Unterstützung gerade noch ausreichend bis eher unzureichend ausfallen. Dabei ist zu bedenken, dass Cronbachs Alpha bei einer (hier z.T. sehr starken) Reduzierung der Itemanzahl trotz einer inhaltlichen Fokussierung auf eine Subdimension tendenziell niedrigere Werte annehmen kann (Döring & Bortz 2016, 481). Insbesondere mit Blick auf die Skala kognitiv-emotionale Unterstützung mit 2 Items kann die geringe Konsistenz überdies am ehesten damit erklärt werden, dass zwar beide Items auf die Unterstützung der Lernenden bezogen sind, allerdings mit dem Fokus auf aufgabenbezogenen sowie interaktiven Support zwei unterschiedliche Unterstützungsarten vorliegen, die zwar auf mittlerem Niveau, aber nicht sehr gut korreliert sind.

<i>Oberflächencluster (UV)</i>	<i>Qualitätsdimension (AV)</i>	<i>Kruskal-Wallis H</i>	<i>df</i>	<i>Sig.</i>
Cluster der inhalts- bezogenen Aktivitä- ten	Klassenführung	0.208	2	$p = .901$
	Sozio-emotionale Unterstützung	3.678	2	$p = .159$
	Auswahl & Aufbereitung der Lerngegenstände und Fach- methoden	4.640	2	$p = .098$
	Kognitiv-emotionale Aktivierung	1.744	2	$p = .418$
Cluster der Sozialfor- men	Klassenführung	1.743	2	$p = .418$
	Sozio-emotionale Unterstützung	0.151	2	$p = .927$
	Auswahl & Aufbereitung der Lerngegenstände und Fach- methoden	3.012	2	$p = .222$
	Kognitiv-emotionale Aktivierung	0.984	2	$p = .611$

Tabelle 7. Kruskal-Wallis-Tests

5 | Diskussion

Ziel des vorliegenden Artikels war es, anhand einer Stichprobe von Unterrichtsvideos aus dem Praxissester einerseits Zusammenhänge zwischen den Tiefenmerkmalen des Unterrichts zu untersuchen und andererseits deren Abhängigkeit von der Oberflächengestaltung des Unterrichts zu prüfen.

5.1 | Forschungsfrage 1: Wie hängen die Tiefenmerkmale des Unterrichts miteinander zusammen?

Bezogen auf die erste Fragestellung konnten mittels korrelativer Analysen unterschiedliche Zusammenhänge exploriert werden. So zeigt sich zum Beispiel, dass die generischen Qualitätsdimensionen von Unterricht, Klassenführung und sozio-emotionale Unterstützung, erwartungskonform stark positiv und hochsignifikant miteinander korrelieren. Konkret heißt das, dass Unterrichtsstunden, die sich durch eine effektive Klassenführung auszeichnen, in der vorliegenden Stichprobe tendenziell auch durch Freundlichkeit und Wertschätzung der Lehrperson sowie eine insgesamt entspannte Lernatmosphäre geprägt sind. Dieses Ergebnis leuchtet einerseits theoretisch unmittelbar ein, da gerade die Störungsintervention und -prävention potenziell konfliktbehaftet ist und sich immer auch auf das Unterrichtsklima auswirken kann, ebenso wie umgekehrt ein gutes Unterrichtsklima eine störungspräventive Funktion erfüllen kann (Seidel 2020, 126). Andererseits ist die Korrelation gut anschlussfähig an die Befunde von Lotz (2016), die im Rahmen der PERLE-Studie mittels einer Hauptkomponentenanalyse eine Gruppe von Items herausarbeiten konnte, die sich grundlegend auf das „Schaffen einer lernförderlichen Atmosphäre“ beziehen.

Dass die generischen Qualitätsdimensionen nicht nur hoch korrelieren, sondern zugleich auch im Mittel hoch ausgeprägt waren, deutet darauf hin, dass es sich hierbei um Qualitätsdimensionen handelt, die von Studierenden im Praxissester im Durchschnitt bereits gut umgesetzt werden. Limitierend muss allerdings berücksichtigt werden, dass beide Qualitätsdimensionen aufgrund der Erhebungssituation starken positiven Verzerrungen unterliegen könnten. Naheliegend ist hier der Einfluss der Kamera auf das Geschehen, den Studien zur Reaktivität von Kameras in der Unterrichtsforschung allerdings eher relativieren (vgl. z. B. Maak/Ricart Brede 2014; Petko et al. 2003). Stärker dürften demgegenüber Verzerrungen wiegen, die durch die Positivselektion der gewählten Klassen sowie ‚Vererbungseffekte‘ seitens der Mentor:innen im Praktikum entstehen: So liegt zum einen die Annahme auf der Hand, dass

die Studierenden eher solche Klassen für die Videografie auswählen, in denen sie sich wohlfühlen. Zum anderen dürften Studierende von den Verhaltensregeln und Routinen profitieren, die ihre Mentor:innen in den Klassen bereits eingeführt haben (vgl. Junker et al. 2021, 354).

Das Problem erhebungsbedingter positiver Verzerrungen trifft auch auf die fachlichen Qualitätsdimensionen zu, wie bereits Stigler et al. (1999, 6) vermuten: „[T]eachers will obviously try to do an especially good job, and may do some extra preparation, for a lesson that is to be videotaped“. Vor diesem Hintergrund sind dann auch die Ratings der Unterrichtsstunde zu interpretieren, die im Mittel den theoretischen Skalenmittelwert nicht erreichen oder nur knapp überschreiten und somit trotz mutmaßlich guter Vorbereitung noch erhebliches Steigerungspotenzial aufweisen. Zwar kann Hesse (2024) mittels der bereits erwähnten Clusteranalysen zeigen, dass sich auch unter den Studierenden Profile sowohl im Bereich der kognitiv-emotionalen Aktivierung als auch mit Blick auf die Auswahl und Aufbereitung von Lerngegenständen und Fachmethoden identifizieren lassen, die auf eine qualitativ hochwertige Durchführung des Unterrichts schließen lassen. Dennoch machen die deskriptiven Ergebnisse darauf aufmerksam, dass die Studierenden gerade eine fachlich angemessene und zugleich kognitiv-emotional aktivierende Unterrichtsgestaltung vor große Herausforderungen stellt (vgl. dazu auch White/Klette 2023).

Die im Artikel vorgenommenen Zusammenhangsanalysen erweitern diesen Befund nun insofern, als hohe signifikante Korrelationen zwischen den beiden fachlichen Dimensionen zu beobachten sind. Anders ausgedrückt: Je angemessener die literarischen Gegenstände ausgewählt und aufbereitet wurden, desto eher boten die Lerngelegenheiten im Unterricht auch Potenzial zu kognitiv-emotionaler Aktivierung (und umgekehrt).

Dieser Befund mag zunächst trivial erscheinen, könnte man doch argumentieren, dass eine fachlich korrekte, strukturierte und das Potenzial der Gegenstände adressierende Unterrichtsstunde mit hoher Wahrscheinlichkeit auch Gelegenheit bietet, dass Schüler:innen kognitiv-emotional aktiviert werden. Mit Blick auf den Forschungsstand muss dies aber mitnichten der Fall sein. So wurden beispielsweise im Projekt Profess-R eine ganze Reihe von Lehrpersonen in Lektionen zur Doppelkonsonantenschreibung beobachtet, die einen in vielen Punkten fehlerbehafteten Unterricht gestalten, aber dennoch anregende Aufgaben stellen (Wiprächtiger et al. 2023).

Im Literaturunterricht der vorliegenden Untersuchung zeigt sich demgegenüber ein in Teilen anderes Bild: Hier ist zumindest auf dimensionaler Ebene ein deutlicher Zusammenhang zwischen der fachlich angemessenen Aufbereitung der Lerngegenstände und dem Potenzial zu kognitiver Aktivierung erkennbar, der einen für das Lernen der Schüler:innen doppelt positiven bzw. doppelt negativen Effekt nach sich ziehen könnte. Denn wenn ein starker Zusammenhang zwischen beiden Dimensionen besteht, liegt es nahe, dass Lernende entweder mit einem fachlich unangemessenen und zugleich wenig aktivierenden *oder* einem fachlich angemessenen und zugleich aktivierenden Literaturunterricht in Kontakt kommen und dementsprechend ein besonders gutes oder ungünstiges Lernangebot vorfinden.

Konkretisieren lässt sich dieser Befund, wenn man die Korrelationen zwischen den beiden Dimensionen auf Ebene der Subdimensionen bzw. Einzelitems betrachtet. Hier zeigt sich, dass vor allem jene Items der kognitiv-emotionalen Aktivierung, die auf die gegenstandsbezogene Unterstützung der Lernenden (durch Aufgaben und in der Interaktion) zielen, mit beinahe allen Einzelitems der Dimension ‚Auswahl und Aufbereitung von Lerngegenständen und Fachmethoden‘ signifikant positiv korrelieren (wobei die geringen Konsistenzwerte der Skala limitierend zu berücksichtigen sind). Der Unterricht enthält also umso adaptivere Unterstützungsangebote, desto angemessener er gestaltet ist (und umgekehrt). Dieser Befund lässt sich insofern plausibilisieren, als adaptiver Support in der Regel nur dann gewährleistet werden kann, wenn die Lehrperson die Verstehensanforderungen des literarischen Textes genau durchdrungen (Item ‚Textauswahl und -einsatz‘) und sich Gedanken über (un)angemessene Deutungen (Item ‚fachliche Korrektheit und Angemessenheit‘) sowie geeignete Verstehenswege (Item ‚Passung der Methoden und Sozialformen‘) gemacht hat. Damit liefern die Befunde empirische Indizien

für die in der Literaturdidaktik weitgehend unstrittige Annahme, dass sich eine genaue didaktische Analyse positiv auf die Qualität des Unterrichts auswirkt und deshalb in der Professionalisierung verstärkt in den Blick zu nehmen ist (vgl. jüngst z. B. Zabka et al. 2022). Inwiefern allerdings tatsächlich ein Zusammenhang zwischen Fähigkeiten im Bereich der didaktischen Analyse einerseits und dem Unterrichtshandeln andererseits besteht, gilt es in zukünftigen Untersuchungen gesondert zu untersuchen. Da im Projekt von Hesse (2024) zu den videografierten Unterrichtsstunden didaktische Analysen im Rahmen von Prüfungsleistungen angefertigt wurden, scheint es gewinnbringend, die Qualität der Analysen mit der beobachteten Unterrichtsqualität zu relationieren.

Zu diskutieren bleibt abschließend der Befund, dass sich zwischen den generischen und den fachspezifischen Qualitätsdimensionen keinerlei Korrelationen gezeigt haben. Dieses Resultat scheint auf den ersten Blick der Annahme hierarchischer Qualitätsmodelle zuwiderzulaufen, dass generische Dimensionen wie Klassenführung und sozio-emotionale Unterstützung Grundvoraussetzungen darstellen, damit sich die fachlichen Qualitätsdimensionen bestmöglich entfalten können. Bei genauerer Betrachtung lässt sich der Widerspruch allerdings auflösen, wenn man den Zusammenhang generischer und fachlicher Qualitätsdimensionen nicht streng linear denkt, sondern wie Reusser (2020, 243) davon ausgeht, dass die generischen Dimensionen lediglich einen gewissen „Schwellenwert“ nicht unterschreiten dürfen, damit fachliches Lernen möglich wird. Da in der vorliegenden Studie in nahezu allen Klassen sehr hohe und den Schwellenwert vermutlich überschreitende Werte sowohl in der Klassenführung als auch in der sozio-emotionalen Unterstützung erreicht wurden, ist folglich nicht überraschend, dass sich keine Korrelationen mehr zwischen den generischen und den fachlichen Qualitätsdimensionen zeigen.

5.2 | Forschungsfrage 2: Wie hängen die Tiefenmerkmale mit den Oberflächenmerkmalen des Unterrichts zusammen?

Mit Blick auf die zweite Forschungsfrage ist das Ergebnis für die beiden betrachteten Oberflächen- und Tiefenmerkmale eindeutig ausgefallen: Sowohl die fachlichen als auch die generischen Qualitätsdimensionen von Unterricht hängen gemäß den zuvor angenommenen Hypothesen weder von der Sozialform noch von bestimmten inhaltsbezogenen Aktivitäten ab, die im Unterricht fokussiert werden. Damit liefern die vorliegenden Analysen für den Literaturunterricht erste empirische Evidenz für die in der allgemeinen Unterrichtsforschung verbreitete Annahme, der zufolge es bezogen auf die Qualität des Unterrichts weniger darauf ankommt, wie der Unterricht auf der Oberflächenebene gestaltet ist. Vielmehr muss querliegend zu bestimmten sozialorganisatorischen Gestaltungsmustern des Unterrichts oder inhaltlichen Schwerpunktsetzungen die Frage eruiert werden, wie Lern- und Verstehensprozesse des Unterrichts angeregt und die Voraussetzungen dafür geschaffen werden können (Decristan et al. 2020).

Gleichwohl dürfen die Befunde nicht derart verstanden werden, dass es fortan nicht mehr lohnend wäre, über bestimmte methodische Zugriffsweisen oder die Wahl von Sozialformen sowie inhaltsbezogenen Aktivitäten nachzudenken. Schließlich schaffen die verschiedenen didaktischen Settings einen jeweils unterschiedlichen Rahmen, in dem sich literarisches Lernen mit unterschiedlichen Akzentuierungen und Zielsetzungen vollziehen kann (vgl. analog zur Rechtschreibdidaktik auch Riegler et al. 2022). Angesichts der überaus großen Breite an literarischen Formen und der Fülle an denkbaren Lernzielen ist dies vermutlich sogar unabdingbar, um Lehrpersonen in der gering strukturierten Domäne Halt zu geben. Die hier formulierten Schlussfolgerungen beziehen sich vielmehr darauf, eine differenziertere Vorstellung über die Wirkannahmen didaktischer Konzeptionen zu entwickeln, und zwar sowohl in der Forschung als auch im Rahmen der Ausbildung angehender Lehrpersonen. Hierzu geben auch Studien aus anderen Lernbereichen des Deutschunterrichts Anlass, die differenzielle Zusammenhänge zwi-

schen Sach- bzw. Oberflächenmerkmalen des Unterrichts und der angebotsseitigen Tiefenstruktur finden (Lotz 2016; Riegler et al. 2022). Solchen Zusammenhängen gilt es auch in der literaturdidaktischen Unterrichtsforschung zukünftig nachzuspüren.

Eine Frage, die auf Basis der vorliegenden Untersuchung nicht empirisch beantwortet werden konnte, bezieht sich auf Wirksamkeitsannahmen methodischer Zugriffsweisen des Literaturunterrichts. Hier war angesichts des geringen Vorkommens poetischer (vs. diskursiver) Zugriffsweisen im betrachteten Korpus keine aussagekräftige Analyse von Abhängigkeiten der Oberflächen- von der Tiefenebene des Unterrichts möglich. Ursächlich hierfür könnte sein, dass im untersuchten Videokorpus nur Einzelstunden untersucht wurden, in denen mindestens eine diskursive Plenumsphase realisiert werden sollte. Gleichwohl darf diese Vermutung relativierend nicht übersehen werden, dass auch Fragebogenerhebungen zum Deutsch- bzw. Literaturunterricht nahelegen, dass poetische handlungs- und produktionsorientierte Methoden auch bei erfahrenen Lehrpersonen seltener als diskursive bzw. analytische Methoden zum Einsatz kommen (Gattermaier 2003; Klieme et al. 2008). Angesichts des Umstands, dass in der Literaturdidaktik in der Vergangenheit immer wieder poetische und diskursive (bzw. handlungs- und produktionsorientierte und analytische) Zugriffe in ihrer Wirksamkeit gegenübergestellt wurden (z. B. Fritzsche et al. 2006), stellt die Frage nach der Abhängigkeit von methodischen Zugriffsweisen von den Tiefenstrukturen des Unterrichts zukünftig also weiterhin ein zentrales Desiderat literaturdidaktischer Unterrichtsforschung dar.

6 | Fazit und Ausblick

Indem der vorliegende Beitrag Zusammenhänge zwischen Tiefenmerkmalen des Literaturunterrichts untersucht und deren Abhängigkeit zu Oberflächenmerkmalen in den Blick genommen hat, konnten Einsichten in die Wirkweise von Literaturunterricht gewonnen werden. Zuvor sei hier der enge Zusammenhang zwischen der Dimension der kognitiv-emotionalen Aktivierung (und hier vor allem der kognitiv-emotionalen Unterstützung) sowie der Auswahl und Aufbereitung von Lerngegenständen und Fachmethoden angesprochen, da dieser die Bedeutung von Fähigkeiten der didaktischen Analyse bei (angehenden) Lehrpersonen betont. Darüber hinaus konnte der Beitrag zudem zeigen, dass die Ausprägung der Tiefenmerkmale nicht signifikant von der Oberflächengestaltung des Unterrichts abhängt. Dies erübrigt nicht die Auseinandersetzung mit Sozialformen und inhaltsbezogenen Aktivitäten, zeigt aber dennoch auf, dass diese immer auch mit Blick auf die Tiefenebene des Unterrichts durchdacht werden müssen. Entsprechende Fähigkeiten müssen im Rahmen der Professionalisierung von Seiten der Fachdidaktik angebahnt werden, da eine rein bildungswissenschaftliche Perspektive auf Unterrichtsqualität die hier als wichtig markierten fachlichen Spezifika des Unterrichts nur unzureichend berücksichtigt.

Die im Beitrag berichteten Befunde sind mit einigen Limitationen behaftet. Neben den bereits erwähnten Verzerrungen, die im Zuge videografischer Erhebungen entstehen können, betreffen diese u. a. die Tatsache, dass in der vorliegenden Arbeit Zusammenhänge lediglich zwischen theoretisch hergeleiteten Dimensionen berechnet wurden. Diese weisen zwar eine hohe interne Konsistenz auf; ausgeschlossen werden kann aber nicht, dass die Items sich auch in anderen Konstellationen zusammensetzen könnten. Zur Überprüfung der Dimensionalität der Items wäre deshalb zukünftig eine Faktorenanalyse gewinnbringend, die im vorliegenden Datensatz aufgrund der geringen Stichprobengröße allerdings nicht sinnvoll durchgeführt werden konnte.

Darüber hinaus gilt es, in zukünftigen Arbeiten dem Einfluss unterschiedlicher professioneller Kompetenzen und Überzeugungen auf die angebotsseitige Qualität (und perspektivisch auch das Lernen der Schüler:innen) nachzuspüren. Im Rahmen des Projekts OVID-PRAX wurden neben den Videodaten der Studierenden auch Testungen fachdidaktischer Kompetenz (Winkler/Seeber 2020) sowie der professionellen Unterrichtswahrnehmung (Hesse/Seeber i. V.) vorgenommen. Darüber hinaus liegen didaktische Analysen in Form von Hausarbeiten zu den

gezeigten Stunden vor, die aktuell ausgewertet werden. Den Einfluss der in diesen Daten vorkommenden Kompetenzen auf den Unterricht zu untersuchen, dürfte wichtige Aufschlüsse dahingehend vermitteln, welche Fähigkeiten im Rahmen der Professionalisierung zukünftig verstärkt zu vermitteln sind. Darüber hinaus muss neben den Prädiktoren für die Ausprägung der angebotsseitigen Unterrichtsqualität auch deren Wirkung auf unterschiedliche Schüler:innen-Outcomes (literarisches Verstehen, Motivation etc.) näher betrachtet werden. Neuere Modellierungen professioneller Kompetenz, wie etwa das sog. Kaskadenmodell (Krauss et al., 2020), können hierfür einen gewinnbringenden Forschungsrahmen bereitstellen, der Wirkweisen des Unterrichts von den Dispositionen der Lernenden bis hin zu den Wirkungen aufseiten der Schüler:innen betrachtet.

Abschließend sei angemerkt, dass eine zentrale Limitation des vorliegenden Beitrags darin besteht, dass die Analysen auf einem Korpus von Einzelstunden beruhen, in denen insbesondere in den methodisch-didaktischen Zugriffsweisen eine geringe Varianz bestand. Zukünftig ist deshalb in anderen Korpora Abhängigkeiten von Oberflächen- und Tiefenmerkmalen des Unterrichts in dieser Hinsicht weiter nachzugehen. Interessante Einblicke kann hier das aktuell laufende Projekt QuaLITY¹⁵ geben, das Unterrichtsreihen erfahrener Lehrpersonen zum Roman *Tschick* untersucht.

¹⁵ <https://didaktikdeutsch.de/quality/>

7 | Literaturverzeichnis

- Backhaus, Klaus / Erichson, Bernd. / Plinke, Wulff / Weiber, Rolf (2016): *Multivariate Analysemethoden. Eine anwendungsorientierte Einführung*. Berlin: Springer. Online unter <<https://doi.org/10.1007/978-3-662-46076-4>>
- Brophy, Jere (2000): *Teaching*. Genf: International Bureau of Education.
- Brunner, Esther (2017): Qualität von Mathematikunterricht. Eine Frage der Perspektive. In: *Journal für Mathematik-Didaktik*, 39, H. 2, S. 257–284. Online unter <<https://doi.org/10.1007/s13138-017-0122-z>>
- Charalambous, Charalambos Y. / Praetorius, Anna-Katharina (2020): Creating a forum for researching teaching and its quality more synergistically. In: *Studies in Educational Evaluation*, 67, H. 100894, S. 1–8. Online unter <<https://doi.org/10.1016/j.stueduc.2020.100894>>
- Decristan, Jasmin / Hess, Miriam / Holzberger, Doris / Praetorius, Anna-Katharina (2020): Oberflächen- und Tiefenmerkmale. Eine Reflexion zweier prominenter Begriffe der Unterrichtsforschung. In: *Zeitschrift für Pädagogik*, 66, 66. Beiheft, S. 102–116. Online unter <<https://doi.org/10.25656/01:25867>>
- Depner, Simone / Kernen, Nora / Pieper, Irene. (2020): Gegenstandskonstitution und literarisches Lernen im Unterrichtsgespräch: Die Videostudie im Projekt TAMoLi – Texte, Aktivitäten und Motivationen im Literaturunterricht der Sekundarstufe I. In: Dawidowski, Christian / Hoffmann, Anna Rebecca / Stolle, Angelika R. / Witte, Jennifer (Hg.): *Schulische Literaturvermittlungsprozesse im Fokus empirischer Forschung*. Frankfurt am Main: Peter Lang, S. 141–160.
- Döring, Nicola / Bortz, Jürgen (2016): *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*. Wiesbaden: Springer.
- Drollinger-Vetter, Barbara (2011): *Verstehenselemente und strukturelle Klarheit. Fachdidaktische Qualität der Anleitung von mathematischen Verstehensprozessen im Unterricht*. Münster: Waxmann.
- Fritzsche, Joachim / Krempelmann, Anita / Tosun, Claudia / Zaborowski, Katrin (2006): *Literaturunterricht kontrastiv. Der handlungs- und produktionsorientierte Literaturunterricht auf dem Prüfstand*. Baltmannsweiler: Schneider Verlag Hohengehren.
- Gabriel, Kathrin / Lipowsky, Frank (2013a): Hoch inferentes Rating. Klassenführung in Deutsch, Kunst und Mathematik. In: Lotz, Miriam / Lipowsky, Frank / Faust, Gabriele (Hg.): *Dokumentation der Erhebungsinstrumente des Projekts „Persönlichkeits- und Lernentwicklung von Grundschulkindern“ (PERLE): 3. Technischer Bericht zu den PERLE-Videostudien (Bd. 23/3)*. Frankfurt am Main: GPPF, S. 145–168.
- Gabriel, Kathrin / Lipowsky, Frank (2013b): Hoch inferentes Rating. Unterrichtsklima in Deutsch, Kunst und Mathematik. In: Lotz, Miriam / Lipowsky, Frank / Faust, Gabriele (Hg.): *Dokumentation der Erhebungsinstrumente des Projekts „Persönlichkeits- und Lernentwicklung von Grundschulkindern“ (PERLE): 3. Technischer Bericht zu den PERLE-Videostudien (Bd. 23/3)*. Frankfurt am Main: GPPF, S. 169–190.
- Gattermaier, Klaus (2003): *Literaturunterricht und Lesesozialisation. Eine empirische Untersuchung zum Lese- und Medienverhalten von Schülern und zur lesesozialisatorischen Wirkung ihrer Deutschlehrer*. Regensburg: Vulpes.
- Gröschner, Alexander / Klaß, Susi / Winkler, Iris (2019): Lernbegleitung von Langzeitpraktika—Konzeption und Designelemente einer hochschuldidaktischen Intervention mittels Unterrichtsvideos. In: Košinár, Julia / Gröschner, Alexander / Weyland, Ulrike (Hg.), *Langzeitpraktika als Lernräume. Historische Bezüge, Konzeptionen und Forschungsbefunde (Bd. 4)*. Münster: Waxmann, S. 85–101.
- Harwart, Miriam / Scherf, Daniel (2018): „Vielleicht muss man aber auch so damit leben können und es aushalten.“: Zur Bedeutung des Lehrerhandelns in schulischen ästhetischen Rezeptionsprozessen. In: Bertschkaufmann, Andrea / Scherf, Daniel (Hg.): *Ästhetische Rezeptionsprozesse in didaktischer Perspektive*. Weinheim: Beltz Juventa, S. 149–163.
- Heins, Jochen (2017): *Lenkungsgrade im Literaturunterricht: Zum Einfluss stark und gering lenkender Aufgabensets auf das Textverstehen*. Berlin: Springer Fachmedien. Online unter <<https://doi.org/10.1007/978-3-658-17890-1>>
- Helmke, Andreas (2014): *Unterrichtsqualität und Lehrerprofessionalität: Diagnose, Evaluation und Verbesserung des Unterrichts*. Stuttgart: Klett | Kallmeyer.
- Hess, Miriam / Lipowsky, Frank (2020): Zur (Un-)Abhängigkeit von Oberflächen- und Tiefenmerkmalen im Grundschulunterricht. Fragen von Lehrpersonen im öffentlichen Unterricht und in Schülerarbeitsphasen im Vergleich. In: *Zeitschrift für Pädagogik*, 66, 66. Beiheft, S. 117–131.
- Hesse, Florian (2024): *Qualitäten von Literaturunterricht. Eine Videostudie im Praxissemester*. Berlin: Springer / J. B. Metzler.
- Hesse, Florian / Seeber, Anna (i. V.): *Professionelle Unterrichtswahrnehmung Studierender im Praxissemester anhand von Peer-Feedbacks zu videografiertem Unterricht*.

- Hesse, Florian / Winkler, Iris (2022): Fachliche Qualität im Literaturunterricht. In: SLLD-Z, 2, S. 1–29. Online unter <<https://doi.org/10.46586/SLLD.Z.2022.9605>>
- IBM Corp. (2021): SPSS Statistics for Windows [Software]. IBM Corp.
- Junker, Robin / Gold, Bernadette / Holodynski, Manfred (2021): Classroom management of pre-service and beginning teachers. From dispositions to performance. In: *International Journal of Modern Education Studies*, 5, H. 2, S. 339–363. Online unter <<https://doi.org/10.51383/ijonmes.2021.137>>
- Kleickmann, Thilo / Steffensky, Mirjam / Praetorius, Anna-Katharina (2020): Quality of Teaching in Science Education: More Than Three Basic Dimensions? In: *Zeitschrift für Pädagogik*, 66, 66. Beiheft, S. 37–55. Online unter <<https://doi.org/10.25656/01:25862>>
- Klieme, Eckhardt / Jude, Nina / Rauch, Dominique / Ehlers, Holger / Helmke, Andreas / Eichler, Wolfgang / Thomé, Günther / Willenberg, Heiner (2008): Alltagspraxis, Qualität und Wirksamkeit des Deutschunterrichts. In: *DESI-Konsortium (Hg.): Unterricht und Kompetenzerwerb in Deutsch und Englisch: Ergebnisse der DESI-Studie*. Weinheim: Beltz, S. 319–344.
- Klieme, Eckhardt / Schümer, Gundel / Knoll, Steffen (2001): Mathematikunterricht in der Sekundarstufe I: „Aufgabekultur“ und Unterrichtsgestaltung. In: Klieme, Eckhardt / Baumert, Jürgen (Hg.): *TIMSS – Impulse für Schule und Unterricht*. Bonn: BMBF Publik, S. 43–57.
- Krauss, Stefan / Bruckmaier, G. / Lindl, Alfred / Hilbert, Sven / Birkner, Karin / Steib, N. / Blum, Werner (2020): Competence as a continuum in the COACTIV study: The “cascade model”. In: *ZDM Mathematics Education*, 52, H. 2, S. 311–327. Online unter <<https://doi.org/10.1007/s11858-020-01151-z>>
- Lipowsky, Frank / Drollinger-Vetter, Barbara / Klieme, Eckhardt / Pauli, Christine / Reusser, Kurt (2018): Generische oder fachdidaktische Dimensionen von Unterrichtsqualität—Zwei Seiten einer Medaille? In: *Martens, Matthias / Rabenstein, Kerstin / Bräu, Karin / Fetzer, Marei / Gresch, Helge / Hardy, Ilonca / Schelle Carla (Hg.): Konstruktionen von Fachlichkeit. Ansätze, Erträge und Diskussionen der empirischen Unterrichtsforschung*. Bad Heilbrunn: Klinkhardt, S. 183–202.
- Lotz, Miriam (2016): Kognitive Aktivierung im Leseunterricht der Grundschule: Eine Videostudie zur Gestaltung und Qualität von Leseübungen im ersten Schuljahr. Berlin: Springer VS. Online unter <<https://doi.org/10.1007/978-3-658-10436-8>>
- Maak, Diana / Ricart Brede, Julia (2014): Empirische Erfassung von Invasivität in videografierten Lehr-Lernsituationen: Entwicklung und Erprobung eines Beobachtungssystems. In: *Neumann, Astrid / Mahler, Isabelle (Hg.): Empirische Methoden der Deutschdidaktik: Audio- und videografierte Unterrichtsforschung*. Schneider Verlag Hohengehren, S. 151–173.
- Magirius, Magirius / Scherf, Daniel / Steinmetz, Michael (2022): Lernunterstützung im Literaturgespräch: Nachweis und Wirkung eines potenziellen Qualitätsaspekts gesprächsförmigen Literaturunterrichts. In: *SLLD*, 2, S. 1–29. Online unter <<https://doi.org/10.46586/SLLD.Z.2022.9552>>
- Meyer, Hilbert (2016): Was ist guter Unterricht? (11. Aufl.). Berlin: Cornelsen.
- Oser, Fritz, / Baeriswyl, Franz J. (2001): *Choreographies of Teaching: Bridging Instruction to Learning*. In: *Richardson, Virginia (Hg.), Handbook of Research on Teaching*. Washington, D. C.: American Educational Research Association, S. 1031–1065.
- Petko, Dominik / Waldis, Monika / Pauli, Christine / Reusser, Kurt (2003): Methodologische Überlegungen zur videogestützten Forschung in der Mathematikdidaktik: Ansätze der TIMSS 1999 Video Studie und ihrer schweizerischen Erweiterung. In: *ZDM*, 35, H. 6, S. 265–280.
- Pianta, Robert C. / Hamre, Bridget K. (2009): Conceptualization, Measurement, and Improvement of Classroom Processes: Standardized Observation Can Leverage Capacity. In: *Educational Researcher*, 38, H. 2, S. 109–119. Online unter <<https://doi.org/10.3102/0013189X09332374>>
- Pianta, Robert C. / Hamre, Bridget K. / Mintz, Susan (2012): *Classroom Assessment Scoring System: Secondary Manual*. Charlottesville, V.A.: Teachstone.
- Praetorius, Anna-Katharina / Charalambous, Charalambos Y. (2018): Classroom observation frameworks for studying instructional quality: Looking back and looking forward. In: *ZDM*, 50, H. 3, S. 535–553. Online unter <<https://doi.org/10.1007/s11858-018-0946-0>>
- Praetorius, Anna-Katharina / Klieme, Eckhardt / Herbert, Benjamin / Pinger, Petra (2018): Generic dimensions of teaching quality: The German framework of Three Basic Dimensions. *ZDM Mathematics Education*, 50, H. 3, S. 407–426. Online unter <<https://doi.org/10.1007/s11858-018-0918-4>>
- Praetorius, Anna-Katharina / Klieme, Eckhardt / Kleickmann, Thilo / Brunner, Esther / Lindmeier, Anke / Taut, Sandy / Charalambous, Charalambos (2020): Towards Developing a Theory of Generic Teaching Quality. Origin, Current Status, and Necessary Next Steps Regarding the Three Basic Dimensions Model. In: *Zeitschrift für Pädagogik*, 66, 66. Beiheft, S. 15–36. Online unter <<https://doi.org/10.25656/01:25861>>

- Praetorius, Anna-Katharina / Rogh, Wida / Kleickmann, Thilo (2020): Blinde Flecken des Modells der drei Basisdimensionen von Unterrichtsqualität? Das Modell im Spiegel einer internationalen Synthese von Merkmalen der Unterrichtsqualität. In: *Unterrichtswissenschaft*, 48, H. 3, S. 303–318. Online unter <<https://doi.org/10.1007/s42010-020-00072-w>>
- Renkl, Alexander (2020): Wissenserwerb. In: Wild, Elke / Möller, Jens (Hg.): *Pädagogische Psychologie*. Berlin: Springer, S. 3–24.
- Reusser, Kurt (2006): Konstruktivismus – vom epistemologischen Leitbegriff zur Erneuerung der didaktischen Kultur. In: Baer, Matthias / Fuchs, Michael / Füglistner, Peter / Reusser, Kurt / Wyss, Heins (Hg.): *Didaktik auf psychologischer Grundlage: Von Hans Aebli's kognitionspsychologischer Didaktik zur modernen Lehr- und Lernforschung*. Bern: h.e.p., S. 151–168.
- Reusser, Kurt (2020): Unterrichtsqualität zwischen empirisch-analytischer Forschung und pädagogisch-didaktischer Theorie. Ein Kommentar. In: *Zeitschrift für Pädagogik*, 66, 66. Beiheft, S. 236–253. Online unter <<https://doi.org/10.25656/01:25877>>
- Riegler, Susanne / Wiprächtiger-Geppert, Maja / Kusche, Dorothea / Schurig, Michael (2020): Wie Primarlehrpersonen Rechtschreiben unterrichten. Zur Praxis des Rechtschreibunterrichts in Deutschland und der Schweiz. In: *Didaktik Deutsch*, 25, H. 49, S. 50–69.
- Riegler, Susanne / Wiprächtiger-Geppert, Maja / Kusche, Dorothea / Schurig, Michael (2022): Wie professionelles Wissen und gegenstandsbezogene Sachstruktur die Qualität von Rechtschreibunterricht beeinflussen. In: *SLLD*, 2, S. 1–25. Online unter <<https://doi.org/10.46586/SLLD.Z.2022.9040>>
- Schendera, Christian F. G. (2010): *Clusteranalyse mit SPSS. Mit Faktorenanalyse*. München: Oldenbourg.
- Schober, Patrick / Boer, Christa / Schwarte, Lothar A. (2018): Correlation Coefficients. Appropriate Use and Interpretation. In: *Anesthesia & Analgesia*, 126, H. 5, S. 1763–1768. Online unter <<https://doi.org/10.1213/ANE.0000000000002864>>
- Schwahl, Markus (2019): „... Aber jetzt zerbreche ich mir den Kopf darüber!“—Kognitive Aktivierung im Literaturunterricht. In: Gawatz, Andreas / Stürmer, Kathleen (Hg.): *Kognitive Aktivierung im Unterricht. Befunde der Bildungsforschung und fachspezifische Zugänge*. Braunschweig: Westermann, S. 86–103.
- Seidel, Tina (2003): Sichtstrukturen – Organisation unterrichtlicher Aktivitäten. In: Seidel, Tina / Prenzel, Manfred / Duit, Reiner / Lehrke, Manfred (Hg.): *Technischer Bericht zur Videostudie „Lehr-Lern-Prozesse im Physikunterricht“*. Kiel: IPN, S. 113–128.
- Seidel, Tina (2020): Klassenführung. In: Wild, Elke / Möller, Jens (Hg.): *Pädagogische Psychologie*. Berlin: Springer, S. 119–132.
- Stahns, Ruven (2021): Zur Operationalisierung der kognitiven Aktivierung in Studien zum Deutschunterricht. In: *Didaktik Deutsch*, 26, H. 51, S. 64–77.
- Steinmetz, Michael (2020): *Verstehenssupport im Literaturunterricht. Theoretische und empirische Fundierung einer literaturdidaktischen Aufgabenorientierung*. Berlin: Springer VS. Online unter <<https://doi.org/10.1007/978-3-658-28378-0>>
- Stigler, James W. / Gonzales, Patrick / Kawanaka, Takako / Knoll, Steffen / Serrano, Ana (1999): *The TIMSS Videotape Classroom Study: Methods and Findings from an Exploratory Research Project on Eighth-Grade Mathematics Instruction in Germany, Japan, and the United States*. Washington, D. C.: National Center for Education Statistics.
- Thiel, Felicitas / Ophardt, Diemut / Piwowar, Valentina (2013): Abschlussbericht des Projekts Kompetenzen des Klassenmanagements (KODEK). Entwicklung und Evaluation eines Fortbildungsprogramms für Lehrkräfte zum Klassenmanagement. Online unter <<https://www.ewi-psy.fu-berlin.de/einrichtungen/arbeitsbereiche/schulentwicklungsforschung/downloads/Abschlussbericht-KODEK.pdf>>.
- White, Marc / Klette, Kirsti (2023): What's in a score? Problematizing interpretations of observation scores. In: *Studies in Educational Evaluation*, 77, H. 101238, S. 1–8. Online unter <<https://doi.org/10.1016/j.stueduc.2023.101238>>
- Wieser, Dorothee (2019): Gegenwärtiger Stand der empirischen Unterrichtsforschung zum Literaturunterricht. In: Kämper-van den Boogaart, Michael / Spinner, Kaspar H. (Hg.): *Lese- und Literaturunterricht. Teil 2: Kompetenzen und Unterrichtsziele, Methoden und Unterrichtsmaterialien: Gegenwärtiger Stand der empirischen Unterrichtsforschung (Bd. 11/2)*. Schneider Verlag Hohengehren, S. 353–384.
- Winkler, Iris (2017): Potenzial zu kognitiver Aktivierung im Literaturunterricht. Fachspezifische Profilierung eines prominenten Konstrukts der Unterrichtsforschung. In: *Didaktik Deutsch*, 22, H. 43, S. 78–97.
- Winkler, Iris (2020): Cognitive Activation in L1 Literature Classes. A content-specific framework for the description of teaching quality. In: *L1 – Educational Studies in Language and Literature*, 20, S. 1–32. Online unter <<https://doi.org/10.17239/L1ESLL-2020.20.01.03>>

- Winkler, Iris / Seeber, Anna (2020): Facetten literaturdidaktischer Kompetenz bei Deutschstudierenden vor und nach dem Praxissemester. Eine Interventionsstudie zur Wirksamkeit videobasierter Lernbegleitung. In: *Didaktik Deutsch*, 25, H. 49, S. 23–49. Online unter <<https://doi.org/10.25656/01:22286>>
- Wiprächtiger, Maja / Stahns, Ruven / Schurig, Michael / Riegler, Susanne / Harcsa, Romana (2023): Fachspezifische Qualitätsprofile von Rechtschreibunterricht in Deutschland und der Schweiz. In: *Swiss Journal of Educational Research*, 45, H. 3, S. 228–241. Online unter <<https://doi.org/10.24452/sjer.45.3.1>>
- Wiprächtiger-Geppert, Maja / Stahns, Ruven / Riegler, Susanne (2021): Fachspezifität von Unterrichtsqualität in der Deutschdidaktik. In: *Unterrichtswissenschaft*, 49, H. 2, S. 203–209. Online unter <<https://doi.org/10.1007/s42010-021-00109-8>>
- Wirtz, Markus, & Caspar, Franz (2002): *Beurteilerübereinstimmung und Beurteilerreliabilität*. Göttingen: Hogrefe.
- Zabka, Thomas / Winkler, Iris / Wieser, Dorothee / Pieper, Irene (2022): *Studienbuch Literaturunterricht. Unterrichtspraxis analysieren, reflektieren und gestalten*. Stuttgart: Klett | Kallmeyer.

Florian Hesse

Universität Jena
florian.hesse@uni-jena.de