



Kriterien naturwissenschaftsdidaktischer Unterrichtsqualität – ein systematisches Review videobasierter Unterrichtsforschung

Benjamin Heinitz · Andreas Nehring

Online publiziert: 4. Mai 2020
© Der/die Autor(en) 2020

Zusammenfassung Im Kontext des Themenheftes „Die Verortung von Merkmalen der Unterrichtsqualität zwischen Generik und Fachspezifik“ bestehen die Ziele des vorliegenden Beitrags darin, einen systematischen Überblick über die in quantitativen, naturwissenschaftsdidaktischen Videostudien angewendeten Kriterien zur Konzeptualisierung und Messung von Unterrichtsqualität zu geben und inhaltliche Beziehungen zu generisch konzeptualisierten und formulierten Kriterien der allgemeinen Unterrichtsforschung herzustellen und zu reflektieren. Im Rahmen eines systematischen Reviews wurden dazu die Kriterien, Operationalisierungen und Beobachtungssitems von 28 Videostudien analysiert, ordnend zusammengefasst und in Beziehung zu einem generisch orientierten Syntheseframework zur Unterrichtsqualität gesetzt, das drei Ebenen unterschiedlicher Auflösungsgrade umfasst. Es wurde ein Kodierverfahren umgesetzt, das wiederum zwischen drei Graden der Übereinstimmung differenziert. Die Objektivität der Kodierungen wurde mit einer Doppelkodierung abgesichert ($0,66 < \kappa < 0,78$). Die Ergebnisse zeigen, dass aus naturwissenschaftsdidaktischen Videostudien insgesamt 388 Kriterien extrahiert und in Beziehung zum Syntheseframework gesetzt werden konnten. Insgesamt lässt sich, auf der allgemeinsten Ebene des Frameworks, ein hoher Grad an Überschneidungen zwischen naturwissenschaftsdidaktischen und eher generisch formulierten Perspektiven herstellen. Auf detaillierteren Ebenen zeigt sich, dass zahlreiche Kriterien spezifisch operationalisiert und in Form von „fachspezifischen Perspektivierungen“ dem Framework zugeordnet werden können. Auf dieser Grundlage präsentiert der Beitrag eine Systematisierung generischer und naturwissenschaftsdidaktischer Perspektiven

Zusatzmaterial online Zusätzliche Informationen sind in der Online-Version dieses Artikels (<https://doi.org/10.1007/s42010-020-00074-8>) enthalten.

B. Heinitz (✉) · A. Nehring
Institut für Didaktik der Naturwissenschaften, Leibniz Universität Hannover, Am Kleinen
Felde 30, 30167 Hannover, Deutschland
E-Mail: heinitz@idn.uni-hannover.de

auf Unterrichtsqualität. In der Diskussion wird aufgezeigt, dass diese Systematisierung disziplinenübergreifende Kommunikation ermöglicht und neue Impulse für die weitere naturwissenschaftsdidaktische Unterrichtsqualitätsforschung setzt.

Schlüsselwörter Unterrichtsqualität · Generik · Fachspezifik · Videostudien · Systematisches Review

Quality of instruction in science education—a systematic review including goals, contents, and methods of science education

Abstract Aiming at systematizing science education and rather generic perspectives on the quality of instruction, this article presents a systematic review of quantitative video studies from science education. Based on a description of the specifics of goals, contents and methods of science education and a literature research, using the PRISMA statement, the items and criteria of 28 video studies were systematized and compared to an interdisciplinary-oriented generic framework. That framework originated from generic classroom research and distinguishes between dimensions and subdimensions of quality of instruction. A coding method was implemented that differentiates between three degrees of congruence. The objectivity of the encoding was verified with a double encoding ($0.66 < \kappa < 0.78$). The results show that a total of 388 criteria could be extracted and compared to the framework. Overall, we found a high degree of comparability between science education and generic perspectives, even though terminologies and content-related emphases can turn out to be different. The article discusses the extent to which the extended framework of generic and science education dimensions of teaching quality can facilitate interdisciplinary communication and which research perspectives emerge from the mutual reflection of the disciplines as well as from the interplay of subject-specific goals, contents and methods.

Keywords Quality of instruction · Generic framework · Content-specificity · Video studies · Systematic review

1 Einleitung

1.1 Forschungsbedarf

Die Bemühungen um ein tieferes und systematisches Verständnis von Unterrichtsqualität gehören zweifelsohne zum Kern naturwissenschaftsdidaktischer Forschungen¹ (Brovelli 2018; Rehm 2018; Wilhelm 2018). Dabei konnte sich seit den TIMSS-Videostudien aus den Jahren 1995 (Stigler et al. 1999) eine For-

¹ Um deutlich zu machen, dass es sich bei den in diesem Beitrag berücksichtigten Disziplinen (Didaktik der Biologie, der Chemie, der Physik und des Sachunterrichts sowie Science Education) um kein homogenes Forschungsfeld ist, wird der Terminus „naturwissenschaftsdidaktische Forschungen“ im Plural verwendet.

schungslinie empirisch-fundierter Unterrichtsforschung etablieren, die insbesondere auf die Analyse videographierter Unterrichtsstudien fokussiert. Darin werden Unterrichtsqualitätsmerkmale in Form von Kriterien und konkreten Beobachtungspunkten operationalisiert und zur Beurteilung von Unterricht angewendet. Facetten von Unterrichtsqualität werden anhand dieser Kriterien explizierbar (Dorfner et al. 2017).

Wenngleich in naturwissenschaftsdidaktischen Forschungen durchaus Bezüge zur allgemeinen Unterrichtsforschung hergestellt werden, ist ein systematischer und studienübergreifender Abgleich von Kriterien aus naturwissenschaftsdidaktischen Studien und generischen bzw. generisch formulierten Kriterien bisher weitgehend ausgeblieben. Dabei hat ein solcher Abgleich das Potenzial, einen Beitrag zu einem umfassenderen Bild von Unterrichtsqualität zu leisten und Kommunikation zwischen den eher naturwissenschaftsdidaktisch und den eher generisch orientierten Disziplinen auf eine gemeinsame theoretische Grundlage zu stellen.

Vor diesem Hintergrund präsentiert der vorliegende Beitrag Einblicke in ausgewählte Spezifika der Ziele, Inhalte und Methoden der naturwissenschaftlichen Unterrichtsfächer sowie ein systematisches Review der, in quantitativen Videostudien, genutzten Kriterien für die Beschreibung und Beurteilung von naturwissenschaftlichem Unterricht. Die in diesen Studien verwendeten Kriterien werden in Beziehung zu einem Syntheseframework gesetzt, das generische und fachspezifisch zu operationalisierende Kriterien der Unterrichtsqualität abbildet (Praetorius und Charalambous 2018; Praetorius et al. [eingereicht](#); Tab. 4 dieses Beitrags). Damit soll herausgearbeitet werden, in welchem Umfang sich Beziehungen zwischen Kriterien der naturwissenschaftsdidaktischen Forschungen sowie dem Syntheseframework herstellen lassen, welche naturwissenschaftsdidaktischen Spezifikationen notwendig sind und auf welchen Ebenen Ergänzungen des Syntheseframeworks zielführend sein können.

1.2 Verortung des Beitrags im Themenheft

Der vorliegende Beitrag ist Teil des Themenheftes „Die Verortung von Merkmalen der Unterrichtsqualität zwischen Generik und Fachspezifik“. Darin werden Abgleiche für weitere Fächer vorgenommen und in einem gemeinsamen Synthesebeitrag zusammengeführt (Praetorius et al. [eingereicht](#)). Das Syntheseframework dient als kommunikative und konzeptuelle Gelenkstelle dieser Abgleiche. Es wurde im Rahmen eines vorangegangenen Reviewprozesses erarbeitet (Praetorius und Charalambous 2018), wobei wiederum generische, fachspezifische und „hybride“, d. h. sowohl fachspezifisch als auch generisch konzeptualisierte, Frameworks zum Mathematikunterricht eingeflossen sind. Bedeutsam für diesen Beitrag ist dabei, dass die Formulierung der Unterrichtsqualitätskriterien nicht allein mathematikspezifisch, sondern auf den Ebenen der Dimensionen, Subdimensionen und Indikatoren auf einem übergreifenden Level ohne fachspezifische Operationalisierungen vorgenommen wurden. Somit bietet sich die Möglichkeit, die formulierten Dimensionen, Subdimensionen und Indikatoren auf weitere Fächer zu übertragen und zu überprüfen, inwiefern Überschneidungen der Konzeptualisierung von Unterrichtsqualität vorliegen und inwiefern Zuordnungen möglich sind. Im Rahmen des Einleitungsbeitrages des The-

menhefts (Praetorius et al. [eingereicht](#)) wird eine deutsche Übersetzung des Syntheseframeworks vorgestellt, welche explizit generische Formulierungen verwendet, sodass eine Übertragung des Syntheseframeworks auf andere Fächer erleichtert wird. Diese deutsche Übersetzung des Syntheseframeworks wird für den abschließenden Vergleich zwischen den Fächern, aber auch zur Darstellung der Ergebnisse dieses Beitrags verwendet. Für die grundlegende Auswahl des Syntheseframeworks gegenüber anderen Möglichkeiten zur strukturierten Erfassung von Unterrichtsqualitätskriterien sei an dieser Stelle auf den Einleitungsbeitrag verwiesen, in dem diese Frage ausführlich erläutert wird.

1.2.1 Vergleich von Kriterien für Unterrichtsqualität

Bei der Abbildung von Kriterien zur Erfassung der Unterrichtsqualität wird in den meisten Ansätzen eine hierarchische Form der Strukturierung genutzt. Kriterien werden dabei unter bestimmten Begriffen zusammengefasst, wobei die Anzahl der Strukturierungsebenen zwischen einzelnen Ansätzen variieren kann. Ein Kriterium für Unterrichtsqualität kann wiederum selbst als Sammelbegriff für einzelne Beobachtungselemente verwendet werden, sodass dieses Kriterium nicht automatisch die unterste Ebene der Hierarchie bilden muss. Für den Vergleich von Kriterien der naturwissenschaftsdidaktischen Forschung mit dem Syntheseframework, muss deshalb zunächst eine gemeinsame Bezugsebene gefunden werden, auf der eine zielführende Gegenüberstellung möglich ist. Die Ebenen, die zur strukturierten Abbildung von Unterrichtsqualität genutzt werden, unterscheiden sich in der Anzahl der jeweils untergeordneten Elemente und somit in ihrem Grad der Abstraktion. Hierbei ist es grundsätzlich so, dass eine „höhere“ Ebene einen stärkeren Grad der Abstraktion aufweist, da eine größere Anzahl unterschiedlicher Konzepte unter einem Begriff versammelt ist (geringerer Auflösungsgrad). Die „unteren“ Ebenen werden dagegen, durch eine abnehmende Anzahl untergeordneter Konzepte, zunehmend konkretisiert und dadurch weniger abstrakt (höherer Auflösungsgrad). Verallgemeinert kann zunächst festgehalten werden, dass ein Beobachtungselement, das sich auf genau einen Aspekt einer Unterrichtsbeobachtung bezieht, die „unterste“ und konkreteste mögliche Ebene der hierarchischen Struktur darstellt. Die darüber liegende Ebene, die mehrere dieser Items zu einem didaktischen Konzept zusammenfasst, wird im Folgenden als ein „Kriterium“ für Unterrichtsqualität bezeichnet. Zwischen den einzelnen Ebenen kann es durch unterschiedliche theoretische Grundlagen dazu kommen, dass terminologische Übereinstimmung vorliegen, auch wenn diese inhaltlich nicht gegeben ist. Andersherum kann es vorkommen, dass eine unterschiedliche Terminologie verwendet wird, auch wenn inhaltlich dasselbe Konzept beschrieben wird. Aus diesem Grund stellt ein Vergleich auf einer möglichst differenzierten „unteren“ Ebene für eine Gegenüberstellung unterschiedlicher Ansätze zur Beschreibung von Unterrichtsqualität den besten Ansatz dar, wobei die inhaltliche Ausrichtung eines Kriteriums auch im Fall einer einheitlichen Terminologie beachtet werden muss. Um eine Handhabbarkeit dieser Vergleiche zu gewährleisten, erfolgt die Darstellung in diesem Beitrag auf „höheren“, d. h. allgemeineren Ebenen. Über Online-Supplements sind die Vergleiche auf den untersten, d. h. am stärksten detailliertesten Ebenen zugänglich gemacht.

1.3 Verortung des Beitrags im Feld der Unterrichtsqualitätsforschung

Die Fragen danach, was Unterrichtsqualität ist, sind keinesfalls einfach zu beantworten und in hohem Maße davon abhängig, mit welchem Ziel, vor welchem theoretischen Hintergrund, mit welchem Grad der Ausdifferenzierung konzeptualisiert und operationalisiert oder inwiefern auf Tiefen- und Oberflächenstrukturebene analysiert wird (Givvin et al. 2005; Kunter und Ewald 2016). Wenngleich mit der Unterscheidung zwischen „good teaching“, „effective teaching“ und „quality teaching“ eine grundlegende Unterscheidung für die Ableitung von Zielkriterien vorliegt (Berliner 2005; Brunner 2018), bestehen nach wie vor zahlreiche Herausforderungen für die Forschung. Dazu gehören nicht nur methodische für eine valide Messung von Unterrichtsqualität durch externe Beobachtende (z. B. Praetorius et al. 2014), sondern auch theoretische Herausforderungen. Die lerntheoretische Verortung von Unterrichtsqualitätskriterien, die Frage der Hierarchisierung verschiedener Kriterien oder die Frage des Verhältnisses von fachspezifischen oder generischen Kriterien von Unterrichtsqualität lassen sich dazu zählen (Schlesinger und Jentsch 2016).

Auf einer strukturellen Ebene ist Unterrichtsforschung ein disziplinenübergreifendes Unterfangen, das sowohl im Rahmen der erziehungswissenschaftlichen, pädagogisch-psychologischen (Kunter und Ewald 2016) als auch der fachdidaktischen Forschung (Sumfleth und Fischer 2013) umgesetzt wird. So zeigen Dorfner et al. (2017) in ihrem systematischen Review auf, dass sich in zahlreichen Videostudien aus dem mathematisch-naturwissenschaftsdidaktischen Bereich Bezüge zu den drei Basisdimensionen von Unterrichtsqualität finden lassen (siehe auch Einleitungsbeitrag in diesem Themenheft), jedoch die konkreteren Konzeptualisierungen und Operationalisierungen sehr verschieden sein können. Ein Grund hierfür kann darin bestehen, dass naturwissenschaftsdidaktische Forschungen vor dem Hintergrund der Spezifik des jeweils beforschten Fachunterrichts oder gar Themenfeldes stattfinden, jedoch kooperativ und interdisziplinär ausgerichtet sein können. Das Modell der drei Basisdimensionen stellt in der Forschung zur Unterrichtsqualität einen häufig verwendeten Ansatz der Strukturierung dar, wird jedoch im Rahmen dieses Themenheftes nicht als Bezugsrahmen für den Vergleich der fachspezifischen Qualitätskriterien verwendet. Trotz der vielfachen Verweise auf das Modell der drei Basisdimensionen in der Forschung zur Unterrichtsqualität wird häufig der fachspezifische (aber teilweise auch der generische) Ergänzungsbedarf der Dimensionen um weitere Aspekte herausgestellt. Im Einleitungsbeitrag dieses Themenheftes wird der Bedarf nach der Ergänzung ausführlich dargestellt, weshalb an dieser Stelle darauf verwiesen sei. Exemplarisch sei jedoch die Untersuchung von Szogs et al. (2017) genannt, in deren Rahmen die *fachliche Korrektheit* und die *fachliche Transparenz* als Ergänzungen der drei Basisdimensionen genannt werden oder die Untersuchung von Brunner (2018) in deren Rahmen die *fachliche Korrektheit* ergänzt wird. Das verwendete Syntheseframework von Praetorius und Charalambous (2018) beinhaltet die zentralen Aspekte der drei Basisdimensionen und erweitert diese um weitere Punkte, die bereits in der empirischen Forschung Anwendung finden, wie die genannten Punkte zur fachlichen Ergänzung. Es erscheint für diesen Beitrag als auch für das Themenheft zielführender, eine solche Erweiterung zu nutzen und heraus-

zuarbeiten, ob bei dieser erweiterten Systematisierung von Unterrichtsqualität nach wie vor Ergänzungsbedarf besteht.

1.3.1 Abgrenzung von Generik und Fachspezifik

Da im Folgenden eine Unterscheidung zwischen generischen und fachspezifischen Qualitätskriterien vorgenommen wird, soll diese Abgrenzung zunächst näher erläutert werden. Grundlegend orientiert sich diese Unterscheidung, wie auch im Einleitungsbeitrag beschrieben, daran, ob ein Kriterium auf andere Fächer übertragen werden kann, oder ob es spezifisch nur in einem Fach angewendet werden kann. Hierbei muss jedoch bedacht werden, dass die Formulierung einer Dimension oder Subdimension durchaus generisch erfolgen kann, auch wenn die darunterliegenden Ebenen erst durch eine Konkretisierung der Operationalisierung einen fachspezifischen Fokus erhalten. Aus diesem Grund sei auf die Unterscheidung zwischen Generik und Fachspezifik von Wüsten (2010) verwiesen, die auch im Rahmen des Reviews zu mathematisch-naturwissenschaftsdidaktischen Videostudien von Dorfner et al. (2017) angewendet wurde, woran dieser Beitrag anschließt und die somit auch im Folgenden Anwendung findet. Nach dieser Unterscheidung setzt ein generisches bzw. fachunabhängiges Kriterium weder fachliches, noch fachdidaktisches Wissen für eine Beurteilung voraus. Fachspezifische Kriterien hingegen sind abhängig vom unterrichteten Fachinhalt und setzen somit für eine Beurteilung ein fachliches und/oder fachdidaktisches Wissen voraus. Durch diese Unterscheidung wird berücksichtigt, dass eine Dimension, Subdimension oder eine darunterliegende Organisationsebene fachunspezifisch formuliert und somit auf andere Fächer übertragen werden kann, aber trotzdem als fachspezifisch wahrgenommen wird, wenn es um die konkrete Operationalisierung geht. Auf das Syntheseframework bezogen bedeutet dies, dass die Dimensionen und Subdimensionen grundsätzlich generisch formuliert sind und somit auf andere Fächer übertragen werden können. Sobald jedoch ein Aspekt beurteilt wird, der fachspezifisches Wissen voraussetzt, wie z. B. die „Akkuratheit und Korrektheit der thematisierten Inhalte sowie Fachmethoden“, wird die Subdimension als fachspezifisch betrachtet, auch wenn die grundlegende Konzeption eines korrekten Inhalts problemlos auf andere Fächer übertragen werden kann. Andere Dimensionen setzen wiederum kein fachspezifisches Wissen voraus und sind somit fächerübergreifend formuliert und auch inhaltlich generisch, wie z. B. das „Verhaltensmanagement“. Während einige Kriterien als klar fachspezifisch oder generisch beschrieben werden können, lassen sich andere als Konglomerat von fachspezifischen und generischen Anteilen beschreiben, so z. B. das Kriterium der „Klarheit“ (Brunner 2018) oder der „kognitiven Aktivierung“ (Praetorius und Charalambous 2018), da zu einer Beurteilung dieser Kriterien sowohl fachliches oder fachdidaktisches Wissen notwendig ist, als auch ein grundsätzliches pädagogisches Wissen.

Die in Studien genutzten Terminologien, Konzeptualisierungen und Operationalisierungen können sowohl generische als auch naturwissenschaftsdidaktische Anteile tragen (siehe Beispiele im Ergebnisteil). Das zieht es nach sich, dass Kriterien in Videostudien angewendet werden können, die eine terminologische Nähe zu generischen Ansätzen tragen, aber auf der Ebene der Konzeptualisierung und Ope-

rationalisierung unterschiedlich sind. Um diesen Zustand zu bezeichnen, nutzt der vorliegende Beitrag den Begriff der „Perspektivierung“ und meint damit, dass generisch orientierte Konzepte und Terminologien mit besonderem naturwissenschaftsdidaktischem Fokus verknüpft sein und interpretiert werden können. Dabei ist eine Fokussierung von Kriterien in generischer Terminologie auf spezifische Elemente des naturwissenschaftlichen Unterrichts möglich, wie z.B. im Falle der „Klarheit und Strukturiertheit des Experiments“ (Schulz 2011). Aber auch eine Entwicklung von fachspezifischen Kriterien und Terminologien, die wiederum eine konzeptuelle Passung zu generischen Merkmalen enthalten, erscheint möglich.

1.3.2 Explizite und implizite Bezüge zur Unterrichtsqualität

Im Kontext naturwissenschaftsdidaktischer Forschungen können die Bezüge zwischen konkreten Studien und Diskursen zur Unterrichtsqualität sowohl expliziter als auch impliziter Natur sein. So existieren empirische und theoretische Arbeiten, die den expliziten Anspruch erheben, naturwissenschaftsspezifische Unterrichtsqualität abzubilden. Zur Konkretisierung seien hier exemplarisch die Videostudie von Börlin (2012) oder die Sammelbände „Wirksamer Fachunterricht“ (Brovelli 2018; Wilhelm 2018; Rehm 2018) genannt. In diesen Arbeiten werden entweder auf einer kriterienorientiert-empirischen oder auf einer diskursiv-theoretischen Ebene Merkmale von Unterrichtsqualität expliziert.

Für eine Zusammenführung von Unterrichtsqualitätskriterien wäre es jedoch unproduktiv anzunehmen, dass naturwissenschaftsdidaktische Arbeiten – ohne expliziten Bezug zum Diskurs der Unterrichtsqualitätsforschung – keine Kriterien formulierten und anwendeten und damit nicht anschluss- und aussagefähig zur Unterrichtsqualität wären. Anders formuliert: Vermutlich behaupteten nur sehr wenige bis keine Naturwissenschaftsdidaktikerinnen oder Naturwissenschaftsdidaktiker, ihre Arbeiten leisteten keinen Beitrag zu einem Verständnis von Unterrichtsqualität – auch wenn darin Studien zur Unterrichtsqualität nur randständig zitiert werden. Exemplarisch seien die Videostudien zur Basiskonzeptorientierung im Biologieunterricht (Förtsch et al. 2018) oder zur Umsetzung naturwissenschaftlicher Denk- und Arbeitsweisen erwähnt (Nehring et al. 2016). Auch wenn diese Studien sich weitgehend außerhalb von expliziten Diskursen zur Unterrichtsqualität verorten, bestehen implizite Bezüge zur Unterrichtsqualität. Im konkreten Fall dieser beiden Studien lässt sich ableiten, dass ein „guter“ naturwissenschaftlicher Unterricht an den fachspezifischen Basis Konzepten orientiert ist, die im Rahmen der Nationalen Bildungsstandards etabliert worden sind, und naturwissenschaftliche Denk- und Arbeitsweisen epistemologisch adäquat in den Unterricht integriert werden sollten. Für das Inbezugsetzen naturwissenschaftsdidaktischer Kriterien mit dem Syntheseframework ist es dabei wichtig, dass eine Aussage zur Unterrichtsqualität abgeleitet werden kann. Die genaue Abgrenzung, ab wann ein Kriterium eine vergleichbare Aussage zur Unterrichtsqualität macht und somit für den Abgleich mit dem Syntheseframework verwendet werden konnte, wird im Rahmen der methodischen Beschreibung des Reviews dargestellt (Abschn. 3.3.3).

2 Ziele und Fragestellungen

Da generisch und naturwissenschaftsdidaktisch orientierte Unterrichtsforschungen bisher selten systematisch in Bezug gesetzt wurden, kann disziplinenübergreifende Kommunikation und die Vergleichbarkeit von Studien erschwert werden. Dies behindert nicht nur Kooperationen in zukünftigen Forschungen, sondern auch die Aggregierung des Standes der Forschung in Reviews oder Meta-Analysen.

An diesem Punkt verortet sich der vorliegende Beitrag. Seine Ziele bestehen darin, eine Übersicht über Kriterien für die Beschreibung und Beurteilung von naturwissenschaftlichem Unterricht zu erarbeiten und in Bezug auf das, auf Praetorius und Charalambous (2018) zurückgehende, Syntheseframework zu setzen. Dazu werden exemplarische Einblicke in ausgewählte Spezifika der Ziele, Inhalte und Methoden der naturwissenschaftlichen Unterrichtsfächer und ein systematisches Review quantitativer Videostudien dargestellt und herausgearbeitet, welche naturwissenschaftsdidaktischen Perspektivierungen und Ergänzung des Syntheseframeworks gewinnbringend sein können.

Angesichts der konzeptuellen Verzahnung naturwissenschaftsdidaktischer Forschungen zur allgemeinen Unterrichtsforschung bei gleichzeitiger fachspezifischer Konkretisierung wird eine vergleichsweise hohe terminologische Passung zwischen naturwissenschaftsdidaktischen Kriterien und dem Syntheseframework erwartet. Diese Erwartung ist insofern begründet, als das Syntheseframework grundsätzlich viele generische Anteile aufweist und für die Beschreibung fachspezifischer Kriterien ebenfalls generische Formulierungen verwendet werden, da diese auf einer eher abstrakten Ebene betrachtet werden. Wird im Bereich naturwissenschaftsdidaktischer Forschung auf einer ebenso abstrakten Ebene gearbeitet, ist eine ähnlich generische Formulierung zu erwarten. Sofern die gereviewten Studien ihre Kriterien auch in der konkretesten fachspezifischen Operationalisierung, z. B. in Form von Beobachtungssitems, präsentieren, wird davon ausgegangen, dass Fachspezifika auftreten können, die auf allgemeiner terminologischer Ebene nicht deutlich werden. Diese Fachspezifika sollen bei der Gegenüberstellung zwischen dem Syntheseframework und der naturwissenschaftsdidaktischen Forschung berücksichtigt werden.

Folgende konkrete Forschungsfragen werden dabei fokussiert:

1. Welche Kriterien von Unterrichtsqualität werden in quantitativen Videostudien naturwissenschaftsdidaktischer Forschungen verwendet und in welchem Umfang lässt sich ein Vergleich der empirisch verwendeten Kriterien zum eher generischen Syntheseframework herstellen?
2. Inwiefern lassen sich aus einem Einblick in Spezifika der Ziele, Inhalte und Methoden der naturwissenschaftlichen Unterrichtsfächer Hinweise dafür ableiten, dass eine weitere Spezifikation des Syntheseframeworks vorgenommen werden kann?

3 Methoden

3.1 Das Vorgehen im Überblick

Zur Erfassung der Qualitätskriterien naturwissenschaftsdidaktischer Unterrichtsforschung wird der Untersuchung ein systematisches Review quantitativer Videostudien vorangestellt. Mit der Fokussierung auf Videostudien wird eine Eingrenzung vorgenommen, die folgendermaßen zu begründen ist: Bei den quantitativen Videostudien handelt es sich um einen dominanten und vielfach umgesetzten Bereich deutschsprachiger naturwissenschaftsdidaktischer Unterrichtsforschungen. Insbesondere die fachlichen Anteile naturwissenschaftlicher Lehr-Lern-Prozesse lassen sich aufgrund ihrer Komplexität und Vielschichtigkeit (siehe Ausführungen in 4.2 „Ausgewählte Spezifika naturwissenschaftlicher Lernprozesse“) nur eingeschränkt in Befragungen von Schülerinnen und Schülern beurteilen, da für diese Beurteilung ein Wissen notwendig wäre, das über die Stundeninhalte hinausgeht. So existieren Befragungen von Schülerinnen und Schülern im naturwissenschaftsdidaktischen Bereich, z. B. in Kontexten des Schwimmens und Sinkens, die sich auch als prädiktiv valide für den Lernerfolg der Schülerinnen und Schüler erwiesen haben (z. B. Fauth et al. 2014). Jedoch sind die Items dieser Befragungen, hier im Falle der kognitiven Aktivierung dargestellt, die als die fachspezifische Dimension der drei Basisdimensionen beschrieben wird, vergleichsweise allgemein („In our science class, we are working on tasks that I have to think about very thoroughly.“, Fauth et al., S. 8), so dass sie sich für die Zwecke des vorliegenden Beitrags eher weniger eignen. Aspekte, wie die Sachstruktur des Unterrichts oder der epistemologisch adäquate Modelleinsatz, lassen sich trivialerweise erst beurteilen, wenn Lehr-Lern-Prozesse durchschritten sind, die weit über die Inhalte einzelner Stunden hinausgehen, was einer Beurteilung durch Schülerbefragungen entgegensteht. Darüber hinaus operationalisieren quantitative Videostudien Unterrichtsqualitätsmerkmale in Form objektiv vergleichbarer Kriterien. Dabei werden Unterrichtsmerkmale in kurzer, teilweise gar stichpunktartiger Form in Listen oder Katalogen zusammengefasst und für die Beurteilung von Unterricht aufbereitet. Teilweise eignen sich die Kriterien bereits für eine Beurteilung von Unterricht, z. B. in einem Ratingverfahren, teilweise werden sie auch in Form von Beobachtungssitems weiter operationalisiert und konkretisiert. Diese Kriterien sind besonders anschlussfähig an das Syntheseframework von Praetorius und Charalambous (2018), das selbst eine Systematisierung von objektivierbaren Kriterien darstellt. Schließlich wird, für die naturwissenschaftsdidaktische Forschung wertvoll, eine Weiterführung der bisherigen Systematisierung des Standes der naturwissenschaftsspezifischen Unterrichtsforschung erreicht. Um die Erfassung der Qualitätskriterien aus dem Review naturwissenschaftsdidaktischer Videostudien theoretisch abzusichern, werden in einem ersten Schritt aus den beschriebenen Spezifika der Ziele, Inhalte und Methoden naturwissenschaftsspezifische Unterrichtsqualitätsmerkmale abgeleitet (siehe Abschn. 3.2). Aufbauend auf das systematische Review zur „[...] Ausrichtung quantitativer Videostudien [...]“ von Dorfner et al. (2017), werden dann in einem zweiten Schritt die Kriterien quantitativer naturwissenschaftsdidaktischer Videostudien herausgearbeitet und diese Kriterien in eine vergleichende Beziehung zum, auf Praetorius und Charalambous (2018) zurückgehenden, Synthe-

seframework gesetzt (siehe 3.3). Diese Gegenüberstellung der Kriterien mit dem Syntheseframework wird zusätzlich mit einer Doppelkodierung abgesichert (siehe 3.4).

Zur abschließenden Gegenüberstellung der Kriterien aus naturwissenschaftsdidaktischen Videostudien und dem Syntheseframework wird der Begriff der „Perspektivierung“ eingeführt (siehe 3.5). Damit soll deutlich gemacht werden, dass eine Abgrenzung zwischen Generik und Fachspezifik erst auf einer Ebene der stärkeren Operationalisierung deutlich wird, auch wenn bei einer abstrakteren Betrachtung auf Ebene der Dimensionen und Subdimensionen eine terminologische Überschneidung auftritt. Diese Perspektivierung stellt somit den Übergang von der fächerübergreifenden Beschreibung generischer und fachspezifischer Merkmale zu einer fachspezifischen Auslegung derselben dar. Die Übereinstimmungen, die im Rahmen des Vergleichs der Spezifika und empirisch angewendeten Kriterien mit dem Syntheseframework untersucht werden, stellen somit keine Gleichheit der beiden Perspektiven dar, sondern verdeutlichen, dass ein Vergleich beider Perspektiven auf Basis eines gemeinsamen Konzeptes zur Unterrichtsqualität möglich ist. Bei dieser Übereinstimmung ist jedoch nicht auszuschließen, dass bei einer differenzierteren Betrachtung Fachspezifika deutlich werden, was durch die Perspektivierung als Bindeglied zwischen Generik und Fachspezifik deutlich gemacht werden soll.

3.2 Berücksichtigung der Spezifika der Ziele, Inhalte und Methoden der naturwissenschaftlichen Fächer

Bei der Identifikation und Ableitung von Qualitätskriterien aus den Spezifika der Ziele, Inhalte und Methoden handelt es sich um ein theoriebasiertes, deduktives Vorgehen, das in den konkreten Fällen der Ergänzung von Kriterien begründet wird. Dabei werden die Spezifika beschrieben, aus denen sich ein Kriterium für Unterrichtsqualität ableiten kann und vor dem Hintergrund ihres Potentials zur Ermöglichung von Lernerfolg im naturwissenschaftlichen Unterricht beurteilt. Anschließend werden sie mit dem Syntheseframework abgeglichen.

Dabei sei explizit darauf hingewiesen, dass es sich – allein schon aus Platzgründen – um ein exemplarisches Vorgehen handelt. Ein theoriebasierter Abgleich von Kriterien mit dem Syntheseframework kann Gegenstand einer eigenen Arbeit sein. Die Einblicke in die Spezifika der Ziele, Inhalte und Methoden der naturwissenschaftlichen Fächer dienen vielmehr dazu, exemplarisch herauszuarbeiten, inwiefern eine naturwissenschaftsdidaktische Theoriearbeit mit dem Syntheseframework weitere Unterrichtsqualitätskriterien gewinnbringend zu Tage fördern könnte. Darüber hinaus soll den Lesenden des vorliegenden Themenheftes, die nicht im Bereich der Naturwissenschaftsdidaktiken tätig sind, eine terminologische Grundlage für das Verständnis des systematischen Reviews gegeben werden. Aus pragmatischen Gründen und um unnötige Redundanzen zu vermeiden, werden die Merkmale beschrieben, die geeignet erscheinen, Dimension oder Subdimension im Syntheseframework zu ergänzen und die nicht schon durch Videostudien benannt werden.

3.3 Systematisches Review

3.3.1 Literatursuche

Das systematische Review wurde unter Berücksichtigung der strukturellen Schritte des PRISMA Statements (Preferred Reporting Items for Systematic Reviews and Meta-Analyses; Moher et al. 2009) durchgeführt. Hierzu wurde die Checkliste zum PRISMA Statement (Ziegler et al. 2011; Anhang 1) abgearbeitet und die für dieses Review relevanten Punkte systematisch expliziert, wobei für die vollständige Darstellung des Vorgehens auf diesen Beitrag verwiesen wird. Als eine erste Grundlage für die Literaturrecherche diente das systematische Review von Dorfner et al. (2017), welches Videostudien im mathematisch-naturwissenschaftlichen Fachbereich erfasste und deren methodische und inhaltliche Ausrichtung untersuchte. Das Review konnte als Grundlage für das in diesem Artikel beschriebene systematische Review genutzt werden, da sich durch den Fokus auf quantitative Videostudien im (mathematisch-)naturwissenschaftlichen Fachbereich eine gemeinsame Untersuchungsbasis ergab, auch wenn in der weiteren Analyse mit einem anderen Fokus gearbeitet wurde. Die im Artikel von Dorfner et al. (2017) berichteten Publikationen zu Videostudien bildeten somit den ersten Datensatz für das hier beschriebene Review. Weiterhin wurde, um anschlussfähig an diesen Stand der Forschung zu bleiben, dasselbe Suchraster wie im Review von Dorfner et al. (2017) genutzt und die darin berichteten Studien um Studien aus den Jahren 2016 bis 2019 ergänzt (Tab. 1).

Publikationen aus dem Jahr 2016 wurden ebenfalls in der Suche zur Erweiterung des Reviews berücksichtigt, um eine lückenlose Abbildung des Erhebungszeitraumes zu ermöglichen, da das Review von Dorfner et al. (2017) bereits Mitte 2016 erstmalig eingereicht wurde. Studien aus dem Bereich der Mathematik, die bei Dorfner et al. (2017) berücksichtigt waren, wurden für die weitere Analyse ausgeschlossen. Zusätzlich wurde eine Einschränkung auf Forschungsarbeiten aus dem deutschsprachigen Raum vorgenommen, wodurch insgesamt 22 Videostudien bzw. 85 Publikationen zu diesen Videostudien aus dem Review von Dorfner et al. (2017) in die weitere Analyse aufgenommen werden konnten.

Die Einschränkung auf den deutschsprachigen Raum wurde vorgenommen, da die teils unterschiedlichen fachdidaktischen Konzeptionen in den verschiedenen Ländern einen einheitlichen Vergleich mit dem Syntheseframework und auch zwischen den in diesem Themenheft berichteten anderen Fachbereichen erschweren würde. Das kann sich u. a. in unterschiedlichen Schwerpunktsetzungen in den Curricula ausdrücken, was wiederum andere Schwerpunktsetzungen in den Qualitätskriterien bzw. andere Orientierungsrahmen für Qualitätskriterien nach sich ziehen kann, die wiederum in einem Beitrag berücksichtigt werden müssten. Es müsste zunächst auch abgesichert werden, dass durch Übersetzungen keine terminologischen Abweichungen entstehen, die Fehlinterpretationen hervorbringen würden. Diese Perspektive kann in Anschlussarbeiten durchaus weiterverfolgt werden, würde aber den Umfang des vorliegenden Beitrags übersteigen.

Auch wenn die Einbettung des Beitrags in das vorliegende Themenheft eine Fokussierung auf den deutschsprachigen Stand der Forschung nahelegt, wurden ebenso internationale Studien in die erweiterte Suche eingeschlossen und erst in einem letz-

ten Schritt aus dem Suchergebnis entfernt. Dies ermöglicht es auf Basis des Suchergebnisses weiterführende Untersuchungen durchzuführen, bedeutet jedoch auch, dass diese in der quantitativen Darstellung der Suchergebnisse ebenfalls abgebildet sind, was bei einer Replikation der Suche beachtet werden muss. Eine Darstellung der Ergebnisse des internationalen Raums kann in diesem Themenheft, aus Gründen der Vergleichbarkeit zwischen den Beiträgen, zwar nicht umgesetzt werden, ist aber somit perspektivisch möglich.

3.3.2 Eingrenzung des Suchergebnisses

Beim ersten Suchdurchlauf zur Erweiterung des Reviews konnten insgesamt 1174 Publikationen in den in Tab. 1 aufgeführten Datenbanken gefunden werden. Dieses umfangreiche Ergebnis konnte jedoch beim ersten Screening stark eingeschränkt werden. Hierbei wurden zunächst Publikationen aussortiert, die auf Grund ihres Titels oder des Abstracts ausgeschlossen werden konnten. Durch die Übernahme des Suchrasters von Dorfner et al. (2017) wurden beispielsweise auch Publikationen gefunden, die in anderen Fächern als dem für diesen Artikel gesetzten naturwissenschaftsidaktischen Rahmen durchgeführt wurden (z. B. Mathematik). Um die Suche nicht durch zusätzliche Suchbegriffe zu beeinflussen, wurden diese Artikel erst nach dem ersten Suchdurchlauf aussortiert, sodass nach dem ersten Screening 457 Publikationen in Hinblick auf ihre angewendete Methodik näher betrachtet wur-

Tab. 1 Suchraster aus Dorfner et al. (2017). Für das Review auf den Zeitraum 2016–2019 angepasst

Recherchequelle	Suchbegriffe	Trefferzahl
Pedocs	Frei-/Volltext= Videostudie ODER Frei-/Volltext= Videoanalyse ODER Frei-/Volltext= Unterrichtsbeobachtung Seit 2016	133
ERIC	Publication Date: „since 2016“ abstract: classroom video analysis OR abstract: video study OR abstract: video tape classroom OR abstract: video data classroom Peer reviewed	97
PsycINFO	Publication Date: „2016–2019“ AB classroom video analysis OR AB video study OR AB video tape classroom Peer reviewed	11
Taylor & Francis On- line	Publication Date: „2016–2019“ Publication Title: „video study“ OR Abstract: „classroom video analysis“ OR Ab- stract: „video study“ OR Abstract: „video tape classroom“ OR Abstract: „video data classroom“	20
Wiley Online Library	Publication Date: „2016–2019“ Abstract: „classroom video analysis“ OR Abstract: „video study“ OR Abstract: „video tape classroom“ OR Abstract: „video data classroom“	60
Springer Link	Find Resources with all of the words: „video“ AND „study“ AND „classroom“ with at least one of the words: „tape“ OR „data“ OR „analysis“ Show documents published between „2016“ and „2019“ Within „Science Education“	853

den. Neben der bereits genannten Fokussierung auf naturwissenschaftliche Fächer wurden weitere Kriterien herangezogen, um die gefunden Publikationen auf ihre Passung für das Review zu überprüfen: Hierzu gehörte die Bedingung, dass es sich um eine quantitative Videostudie handelt, die in einer Schule durchgeführt wurden (Ausschluss von Exkursionen, Laborsettings, universitärer Lehrer, Kindergärten und Vorschulen). Das Verhalten der Lehrkraft sollte fokussiert werden und es sollte sich um eine vollständig ausgebildete Lehrkraft handeln (Ausschluss von Studierenden und ReferendarInnen, sofern sie nicht die Rolle einer vollständig ausgebildeten Lehrkraft einnahmen und eine reale Unterrichtsstunde bewertet werden konnte). Weiterhin wurden Videostudien ausgeschlossen, in denen das Unterrichtsvideo nicht der Fokus der Untersuchung war, sondern lediglich als unterstützende Quelle zur Bestätigung von Interviewdaten o. ä. genutzt wurde. Viele dieser Kriterien konnten bereits beim zweiten Screening mit einem Blick auf die jeweiligen angewandten Methoden überprüft werden, sodass 111 Publikationen für ein abschließendes Screening des Gesamttextes verblieben. Bei diesem abschließenden Screening wurden die bereits genannten Kriterien erneut angelegt und der Gesamttext der Publikation betrachtet, falls sich bei der Untersuchung der Methoden noch keine konkrete Aussage über die Eignung für das Review treffen ließ. Weiterhin gewährte dieses Vorgehen eine Absicherung der zuvor getroffenen Auswahl. Nach der Betrachtung des Gesamttextes verblieben 51 Publikationen zu quantitativen Videostudien, die den zuvor genannten Kriterien entsprachen. Wie bereits beschrieben wurde das Suchergebnis für die Erhebung der Qualitätskriterien für diese Untersuchung auf den deutschsprachigen Raum begrenzt, was 27 Publikationen einschloss. Durch die Überschneidung des Zeitraums des Reviews von Dorfner et al. (2017) und der erweiterten Suche dieses Beitrags mussten in einem letzten Schritt gefundene Duplikate ausgeschlossen werden. Für den Vergleich zwischen den Kriterien der Videostudien und dem Syntheseframework von Praetorius und Charalambous (2018) verblieben somit 10 Publikationen für den Zeitraum 2016–2019, in denen über 6 zuvor nicht erfasst Videostudien berichtet wurde. In zwei dieser Fälle wird von einer Anbindung an bereits erfasst Videostudien berichtet, da die Publikationen jedoch über den Rahmen einer einzelnen Videostudie hinaus berichten, wurden sie gesondert aufgeführt. Für eine detaillierte Auflistung aller Videostudien und der dazugehörigen Publikationen, sowie der fachlichen Ausrichtung – siehe Anhang 2. Unter Berücksichtigung des Reviews von Dorfner et al. (2017) ergab dies eine Gesamtzahl von 28 Videostudien mit 95 Publikationen, die in das Review eingingen.

3.3.3 Herausarbeitung der Qualitätskriterien

Für die Beantwortung der ersten Fragestellung nach den empirisch angewendeten Qualitätskriterien mussten zunächst die in den jeweiligen Publikationen angegebenen Kriterien identifiziert und herausgearbeitet werden. Hierbei wurden alle Kriterien berücksichtigt, die für die Datengenerierung im Rahmen der Beurteilung von Unterricht Anwendung fanden (z. B. in Form von Kodiermanualen) und somit eine direkte Anbindung an die empirische Forschung aufweisen. Kriterien, die sich auf die Oberflächenstruktur von Unterricht beziehen und keine Bezüge zur Unterrichtsqualität oder zur Nutzung von Unterrichtsangeboten aufwiesen, z. B. „Redeanteile

innerhalb einer Stunde“ oder die Art der „Sachbegegnung“ (Alltagsgegenstände, Computersimulation, Chemische Geräte, Gedankenexperimente, Mischung; Schulz 2011), wurden nicht berücksichtigt. Bei den Beispielen werden Aspekte der Untersuchung benannt, die in der genannten Publikation ohne eine weitere Aussage zur Unterrichtsqualität aufgeführt wurden. Kriterien mit einer inhaltlichen Überschneidung oder ähnlichen Terminologie können durchaus in anderen Publikationen oder in einem anderen fachlichen Diskurs eine Anbindung zur Unterrichtsqualität aufweisen. Wenn dies jedoch nicht in der untersuchten Publikation geschah, wurden sie für das Review nicht als Qualitätskriterien erfasst. Die Erfassung von Oberflächenstruktur ohne eine Anbindung zur Unterrichtsqualität konnte nicht zielführend mit Kriterien verglichen werden, die eine Anbindung zur Unterrichtsqualität aufweisen. Beispielsweise trifft ein Kriterium wie die „Verwendung von Materialien aus dem Alltag“ (Widodo und Duit 2004) eine eindeutige Aussage, welche Art der von Schulz (2011) beschriebenen „Sachbegegnung“ zu bevorzugen ist. Auch für den späteren Vergleich mit dem Syntheseframework, welches auf Unterrichtsqualität fokussiert, ist eine eindeutige Aussage zur Unterrichtsqualität notwendig, damit ein Kriterium beispielsweise mit der Beschreibung „Content is explicitly presented (e.g., by selecting appropriate examples)“ vergleichbar ist. Eine einfache Auflistung der verwendeten Zugänge weist zwar eine inhaltliche Ähnlichkeit auf, es wird jedoch hierbei keine Aussage in Bezug auf einen „guten“ Unterricht getroffen. Wurden Aspekte der Oberflächenstruktur in einem Kontext der Unterrichtsqualität oder Angebotsnutzung verortet, wurden sie – wie z. B. „Organisationformen im Unterricht“ mit einem Hinweis auf die Lernförderlichkeit von kooperativem Lernen (Schulz 2011) – einbezogen. Wichtig für eine Berücksichtigung eines Kriteriums ohne einen expliziten Bezug zur Unterrichtsqualität ist, dass sich eine Handlungs- oder Verhaltensempfehlung für den Unterricht daraus ableiten lässt.

Wurden in den Publikationen einer Studie keine Kodiermanuale zugänglich gemacht, wurden die Kriterien berücksichtigt, die im Text der Publikationen benannt wurden. Zur Aufnahme in die Kodierung war es jedoch notwendig, dass die genannten Qualitätskriterien im Rahmen einer Unterrichtsaufzeichnung Anwendung fanden. Im Rahmen des Reviews wurde jede gefundene Publikation zu naturwissenschaftsdidaktischen Videostudien erfasst und die darin berichteten Qualitätskriterien tabellarisch erfasst (siehe Anhang 3²). Da die im Rahmen des Reviews gefunden Videostudien eine sehr unterschiedliche Anzahl an Publikationen aufweisen, wurde die Einschränkung vorgenommen, dass ein Qualitätskriterium für den an das Review anschließenden Vergleich mit dem Syntheseframework lediglich einmalig pro Videostudie erfasst wurde. Sollten Qualitätskriterien innerhalb einer Videostudie mit einer ähnlichen Terminologie, aber inhaltlicher Abweichung in unterschiedlichen Publikationen auftreten, wurden sie für die Videostudie vollständig erfasst. Beispielsweise weisen ein „motivational unterstützender Unterricht“ (Seidel et al. 2006a) und eine „Exploration der Interessen [...] der Schüler“ (Widodo und Duit 2004) eine inhaltli-

² Anhang 3 kann auf Grund des Umfangs der Tabelle (111 × 124 Felder, inklusive Kommentare) nicht in einer angemessenen Darstellung an diesen Artikel angehängt werden. Um die vollständige Tabelle dennoch für die Lesenden verfügbar zu machen, kann sie auf Nachfrage von den Autoren individuell zur Verfügung gestellt werden.

che Überschneidung auf, nähern sich diesem Aspekt der Unterrichtsqualität jedoch aus unterschiedlichen Richtungen. Beide Beispiele wurden in diesem Fall Publikationen entnommen, die im Rahmen der IPN-Videostudie veröffentlicht wurden und beide Qualitätskriterien wurden für den Vergleich mit dem Syntheseframework genutzt. Sollte hingegen in unterschiedlichen Publikationen zu derselben Videostudie unterschiedliche Terminologien für dasselbe inhaltliche Konstrukt verwendet werden, wurde lediglich eine der Ausführungen erfasst – dieser Fall ist vor allem dann eingetreten, wenn die Ergebnisse einer Videostudie mehrsprachig veröffentlicht wurden. Als Beispiel hierfür kann die „lebensweltliche Einbettung“ (Börlin 2012) genannt werden, welche in einer englischen Übersetzung als „embedding of everyday-life“ (Börlin und Labudde 2014) ebenfalls mit Bezug zur QuIP-Studie beschrieben wird, aber auch in der Variation „everyday-life context“ (Beerenwinkel und Arx 2016) im Rahmen derselben Videostudie verwendet wird. In diesem Fall wurde lediglich die „lebensweltliche Einbettung“ als Qualitätskriterium der QuIP-Studie mit dem Syntheseframework verglichen. Die Anzahl der Veröffentlichungen zu einer Videostudie unterschied sich in einem Rahmen von einer bis 26 Publikationen. Innerhalb einer Videostudie weisen die Publikationen häufige Überschneidungen und Querverweise zu Publikationen oder Kodiermanualen derselben Videostudie auf. Bei einer quantitativen Darstellung der erhobenen Qualitätskriterien und deren Verteilung auf das Syntheseframework würde das Gesamtbild somit durch Videostudien mit besonders vielen Publikationen sehr stark geprägt. Um zu vermeiden, dass eine einzelne Videostudie das Gesamtbild beim Vergleich der Videostudien mit dem Syntheseframework zu stark beeinflusst, wurde die Einschränkung vorgenommen, dass pro Videostudie lediglich ein Qualitätskriterium aufgeführt wird, wenn es eine vollständige inhaltliche Überschneidung zwischen mehreren Qualitätskriterien gibt.

3.3.4 Vergleich der Qualitätskriterien mit dem Syntheseframework

Der Beitrag steht vor der grundlegenden Herausforderung Kriterien zur Erfassung der Unterrichtsqualität aus unterschiedlichen Bereichen gegenüberzustellen. Für einen Vergleich von Qualitätskriterien aus der naturwissenschaftsdidaktischen Forschung mit dem Syntheseframework ist jedoch zunächst eine Analyse der jeweiligen Hierarchisierungen notwendig. Das Syntheseframework besteht aus Dimensionen („höchste und allgemeine Ebene“), Subdimensionen („mittlere Ebene“) und einer Indikatorebene („unterste und detaillierteste Ebene“). Im Fall des Syntheseframeworks bedeutet dies, dass ein Vergleich auf der Indikatorebene den besten Zugang bietet und für die naturwissenschaftsdidaktische Forschung ist dies, abhängig von der jeweiligen Publikation, würde jedoch im Optimalfall die Betrachtung von Beobachtungssitems bedeuten. Wie bereits in Abschn. 1.2.1 angedeutet, können bei diesem Vergleich jedoch Schwierigkeiten auftreten. Zunächst kann es vorkommen, dass in einer Publikation zu einer Videostudie keine Beobachtungssitems, sondern direkt Kriterien zur Erfassung der Unterrichtsqualität genannt werden. In diesem Fall wurde mit der Beschreibung in der jeweiligen Publikation gearbeitet, um die inhaltliche Ausrichtung des Kriteriums herauszustellen und einen Vergleich mit dem Syntheseframework zu ermöglichen. Damit eine einheitliche Darstellung auf der Seite der Naturwissenschaftsdidaktik erfolgen konnte, wurde der Vergleich zum Synthese-

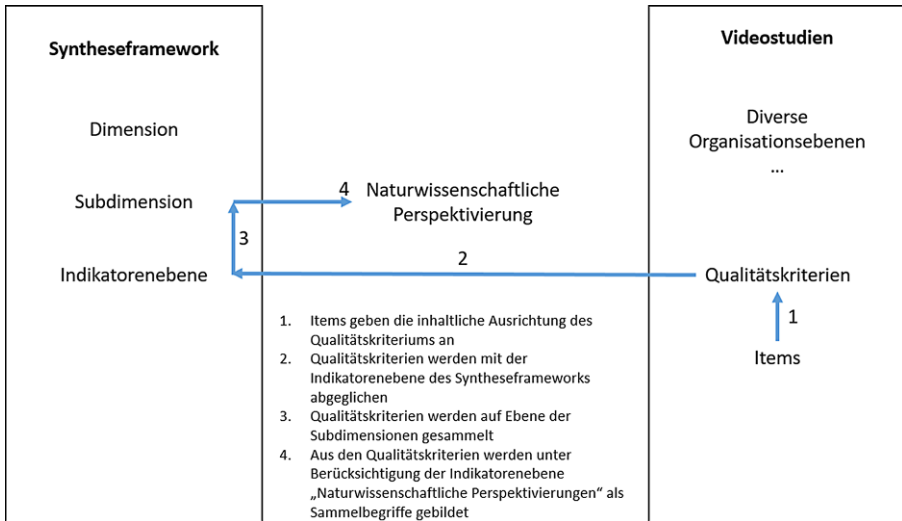


Abb. 1 Ebenen des Vergleichs zwischen Syntheseframework und Qualitätskriterien naturwissenschafts-didaktischer Videostudien

framework grundsätzlich mit den jeweils verwendeten Kriterien dargestellt. Sofern Beobachtungssitem vorlagen, gaben sie die inhaltliche Ausrichtung eines Qualitätskriteriums an und wurden somit direkt berücksichtigt.

Eine weitere mögliche Schwierigkeit besteht dann, wenn ein Kriterium inhaltlich einen größeren Bereich abdeckt als die Indikatoren des Syntheseframeworks. Das Syntheseframework wird deshalb zwar grundlegend auf der Indikatorebene betrachtet, es wird jedoch die Möglichkeit geboten Kriterien aus der naturwissenschafts-didaktischen Forschung auch auf Ebene der Subdimensionen zu vergleichen, sofern dies durch den verwendeten Grad der Abstraktion notwendig ist. Um abschließend einen einheitlichen Vergleich der zugeordneten Kriterien zu ermöglichen, werden die Ergebnisse auf Ebene der Subdimensionen zusammengefasst, welche auch für den fächerübergreifenden Vergleich genutzt werden.

Insgesamt wird somit eine einheitliche Ebene der Abstraktion zur Beschreibung von Unterrichtsqualität geboten, die einen möglichst detaillierten Vergleich zwischen den unterschiedlichen Ansätzen erlaubt, ohne dabei zwischen den hierarchischen Ebenen zu wechseln.

Für den Vergleich der in den Videostudien verwendeten Qualitätskriterien mit dem generischen Syntheseframework wurde, ausgehend von der Beschreibung des Syntheseframeworks (Praetorius und Charalambous 2018) gearbeitet. Die herausgearbeitete inhaltliche Ausrichtung eines naturwissenschafts-didaktischen Qualitätskriteriums wurde mit den Subdimensionen bzw. der Indikatorebene des Syntheseframeworks abgeglichen und zunächst tabellarisch gegenübergestellt (siehe Anhang 3).

Abb. 1 stellt die Schritte zum Vergleich der Kriterien in ihrer Gesamtheit dar. Im Fall von fehlenden Beobachtungssitem entfiel Schritt 1 und im Fall eines höheren Abstraktionsgrades eines Qualitätskriteriums beinhaltet Schritt 2 eine direkte Zu-

ordnung auf Subdimensionsebene. Da bei der Beschreibung von Qualitätskriterien der naturwissenschaftsdidaktischen Forschungen davon ausgegangen werden kann, dass unter anderem fachspezifische Formulierungen verwendet werden, sollen diese abschließenden in Schritt 4 ebenfalls berücksichtigt werden. Die Beschreibung zur Bildung dieser naturwissenschaftlichen Perspektivierung wird in Abschn. 3.5 dargestellt.

Für den Vergleich wurde die englische Version des Syntheseframeworks von Praetorius und Charalambous (2018) verwendet, weshalb diese auch in Tab. 3 dargestellt wird. Eine deutsche Übersetzung lag zu Beginn der Arbeit nicht vor, wird jedoch in einer erweiterten Darstellung in Tab. 4 und im Rahmen des Einleitungsbeitrags berichtet. Da die deutsche Übersetzung bereits eine inhaltliche Erweiterung der Dimensionen und Subdimensionen beinhaltet, wird für die Darstellung des ersten Vergleichs mit dem Syntheseframework die englische Version verwendet. Bei der Übersetzung und inhaltlichen Erweiterung sind bereits Ergebnisse des ersten Vergleichs mit dem Syntheseframework eingeflossen und Abweichungen, die zuvor zwischen dem Syntheseframework und den naturwissenschaftsdidaktischen Qualitätskriterien bestanden, wurden dadurch teilweise behoben, wodurch die in Tab. 3 dargestellten Abweichungen nicht vollständig auf das Erweiterte Syntheseframework übertragen werden können.

Die Gegenüberstellung der Qualitätskriterien mit dem Syntheseframework erfolgte anhand eines dreistufigen Kategoriesystems. Hierbei wurde der Grad der Vergleichbarkeit zum Syntheseframework bei jedem Qualitätskriterium analysiert. Die drei Stufen der Vergleichbarkeit wurden als „vollständige Übereinstimmung“, „teilweise Übereinstimmung“ und „starke Abweichung“ festgelegt.

Eine „vollständige Übereinstimmung“ schließt alle Qualitätskriterien ein, die inhaltlich im Syntheseframework wiederzufinden sind oder einen Teil der Indikatorebene einer Subdimension des Syntheseframeworks erfüllen. Somit werden in dieser Kategorie auch Qualitätskriterien erfasst, die ein weniger komplexes Konstrukt beschreiben, als es in einer Subdimension des Syntheseframeworks der Fall ist – z. B. weist „goal-clarity“ (Börlin und Labudde 2014) eine „vollständige Übereinstimmung“ mit „Presenting the content in a structured way“ auf, da die Indikatorebene der Subdimension des Syntheseframeworks „Lesson objectives are clear“ (siehe Anhang 3) enthält. Das Syntheseframework und die Beschreibung des Qualitätskriteriums stimmen hierbei in ihrer inhaltlichen Auslegung einer Zielklarheit als Teil der guten Struktur einer Unterrichtsstunde überein, auch wenn im Syntheseframework in diesem Bereich noch weitere Kriterien benannt werden.

Eine „teilweise Übereinstimmung“ umfasst alle Fälle, in denen Qualitätskriterien aus Studien entweder durch mehrere unterschiedliche Subdimensionen des Syntheseframeworks abgedeckt werden, und damit nicht klar zuzuordnen sind oder zusätzliche Aspekte beinhalten, die im Syntheseframework in dieser Form nicht abgedeckt sind. So könnten in einigen Fällen Items oder der beschriebene Inhalt eines Qualitätskriteriums auch auf mehrere Subdimensionen des Syntheseframeworks verteilt werden, so dass mehrfache Zuordnungen vorgenommen werden müssten. Als Beispiel können hierfür die „Anker und Integrationshilfen“ (Herweg 2008) genannt werden, die eine Übereinstimmung zu der Subdimension „Presenting the content in a structured way“ (Syntheseframework) durch die Verknüpfung zu „Lernzielen“

und dem „roten Faden“ aufweisen. Aber auch eine Übereinstimmung zu „Teacher facilitation of students’ cognitive activity“ (Syntheseframework) aufweisen, da eine „Verknüpfung mit Vorwissen“ stattfinden soll. In diesen Fällen wurde das Qualitätskriterium lediglich der Subdimension zugeordnet, bei der die meisten Items eine Passung aufwiesen bzw. in der die Beschreibung innerhalb der Publikation am ehesten treffend schien. Im zweiten Fall der „teilweisen Übereinstimmung“ kann ein Qualitätskriterium zwar prinzipiell einer Subdimension zugeordnet werden, enthält jedoch weitere Aspekte, die nicht im Syntheseframework abgebildet sind. Diese können hierbei sowohl fachspezifisch, als auch generisch sein, geben jedoch in beiden Fällen einen Hinweis auf eine mögliche Ergänzungsstelle des Syntheseframeworks. Als Beispiel für diesen Fall kann das Qualitätskriterium „Vernetzung“ (Glennitz 2007) genannt werden, bei dem sich zwar eine Überschneidung zu „Presenting the content in a structured way“ (Syntheseframework) ergibt, jedoch zusätzlich zur einfachen Verbindung der Inhalte auch das Niveau der Vernetzung untersucht wird, was im Syntheseframework nicht weiter ausgeführt wird.

Um keine artifizielle Erhöhung der Anzahlen „teilweiser Übereinstimmung“ zu erzeugen, wurde für jedes Qualitätskriterium der Videostudien nur eine Zuordnung auf Subdimensionsebene vorgenommen. Das betreffende Qualitätskriterium wurde dann innerhalb der Kodiertabelle mit einem Hinweis auf nicht enthaltene Inhalte oder auf Inhalte aus anderen Subdimensionen versehen. Die Frage nach der Fachspezifik von Kriterien wurde im Rahmen der „Perspektivierung“ von Kriterien weiterverfolgt.

Die dritte Abstufung „starke Abweichung“ weist wiederum auf explizite Ergänzungsstellen hin. Kriterien, die dieser Abstufung zugeordnet werden, können in keiner der Subdimensionen des Syntheseframeworks verortet werden und wurden lediglich auf Dimensionsebene eingeordnet. Als ein Beispiel hierfür kann der „Evolutionärer Umgang mit Schülervorstellungen“ (Cauet 2015) genannt werden. Hierbei wird im Gegensatz zum Syntheseframework nicht eine Korrektur bestehender Vorstellung beschrieben, sondern ein aufbauender Prozess, der an die bestehenden Vorstellungen anknüpft. Dieser fachspezifische Prozess ist in dieser Form nicht im Framework enthalten, wodurch sich die inhaltliche Abweichung ergibt.

Dieser Vergleich auf den Ebenen der „Übereinstimmung“ bildet somit ab, welche Qualitätskriterien in Videostudien der naturwissenschaftlichen Fächer angewendet werden und welche sich bereits inhaltlich im Syntheseframework wiederfinden lassen.

3.4 Absicherung der Objektivität durch Doppelkodierung im systematischen Review

Um sicherzustellen, dass – angesichts des interpretativen Anteils der Zuordnungen – die Verortung der Qualitätskriterien im Syntheseframework mit einer ausreichenden Objektivität erfolgte, wurde eine Doppelkodierung durch einen zweiten Rater durchgeführt. Die Doppelkodierung umfasste 17% der insgesamt erfassten Kriterien, was 65 Kriterien entspricht. Die Auswahl der Kriterien erfolgte hierfür zufällig, wobei jeweils ganzheitliche Publikationen zu verschiedenen Videostudien aus verschiedenen Fachbereichen gewählt wurden. Dies umfasste 3 Publikationen aus den Fachbereichen Chemie, Physik und Sachunterricht, welche in Anhang 2 markiert

Tab. 2 Ergebnis der Doppelkodierung

	Anzahl der Übereinstimmungen	Anzahl der Abweichungen	Anzahl der Kriterien	Prozentuale Übereinstimmung (%)	Anzahl der möglichen Zuordnungs-Dimensionen	Cohens Kappa
Dimension	54	11	65	83,1	7	0,78
Sub-Dimension	45	20	65	69,2	20	0,66
Abstufung	45	20	65	69,2	47	0,67

sind. Als Maße der Objektivität wurden sowohl die relative Beurteilerübereinstimmung als auch Cohens Kappa betrachtet. Nach Wirtz und Caspar (2002) weisen die Richtwerte für Cohens Kappa zwar je nach Autor unterschiedliche Werte auf, können jedoch auf die Werte ($0,60 < \kappa < 0,75$) für den Bereich guter Übereinstimmung und ($\kappa > 0,75$) für den Bereich sehr guter Übereinstimmung zusammengefasst werden. Die Ergebnisse der Doppelkodierung (Tab. 2) lassen darauf schließen, dass die Einordnung der Kriterien in das Syntheseframework mit einer angemessenen Objektivität erfolgte. Die Doppelkodierung wurde für eine präzisere Analyse auf drei Facetten der Zuordnungen betrachtet. Zuerst wurde die Übereinstimmung bezüglich der Zuordnung von Kriterien auf Ebene der sieben Dimensionen des Syntheseframeworks überprüft, anschließend die Zuordnung auf Ebene der Subdimensionen und abschließend die Übereinstimmung bei den zuvor beschriebenen Abstufungen der Zuordnungen (3.3.4). Bei den beschriebenen Ebenen ist zu beachten, dass eine gemeinsame Zuordnung zu einer tieferen Ebene z. B. „Abstufung in der Übereinstimmung“ voraussetzt, dass eine gemeinsame Zuordnung in den darüber liegenden Ebenen (Dimension & Subdimension) stattgefunden hat. Die Rater können nicht dieselbe „Abstufung der Übereinstimmung“ gewählt haben, wenn sie zuvor nicht dieselbe Subdimension gewählt haben und ebenso nicht dieselbe Subdimension, wenn sie zuvor nicht dieselbe Dimension gewählt haben.

Der Wert für Cohens Kappa lag im ersten Fall der Zuordnung der Qualitätskriterien (Dimensionen) bei 0,78 woraus geschlossen werden kann, dass die Rater die Dimensionen in weiten Teilen gleich interpretierten. Im zweiten (Subdimensionen) und dritten Fall (Abstufungen) ließ sich zunächst feststellen, dass die absolute Anzahl der Übereinstimmungen in beiden Fällen gleich groß war. Hieraus folgt, dass die Rater bei einer einheitlichen Zuordnung zu einer Subdimension auch dieselbe Abstufung wählten, da eine gemeinsame Abstufung voraussetzt, dass zuvor dieselbe Subdimension gewählt wurde.

Die Zuordnung zu einer Subdimension liegt mit einem Wert für Cohens Kappa von 0,66 in einem Bereich guter Übereinstimmung, woraus hierfür ebenfalls eine weitgehend einheitliche Interpretation der Subdimensionen abgeleitet werden kann. Dies kann ebenfalls für die Zuordnung zu einer Abstufung mit einem Wert für Cohens Kappa von 0,67 angenommen werden. Die Abweichungen zwischen den Ratern wurden für eine bessere Analyse der Doppelkodierung detailliert betrachtet. Hierbei stellte sich heraus, dass in 18 von 20 Fällen einer Abweichung durch mindestens einen der beiden Rater die Abstufung „teilweise Übereinstimmung“ oder „starke Abweichung“ gewählt wurde. Aus den Definitionen dieser Abstufungen (3.3.4) geht

hervor, dass diese einen interpretativen Anteil enthalten, welcher von den Ratern dementsprechend unterschiedlich ausgelegt werden kann. Trotz dieser interpretativen Anteile lässt das Ergebnis der Doppelkodierung auf eine grundlegende Objektivität der Zuordnung schließen. Abweichungen zwischen den Ratern wurden im Anschluss an die Doppelkodierung diskutiert, wobei ein besonderer Fokus auf den interpretativen Anteil der Abstufungen und das Verständnis der Subdimensionen des Syntheseframeworks gelegt wurde. Abweichende Interpretationen wurden hierbei herausgearbeitet und in einem Konsensverfahren vereinheitlicht, sodass die gemeinsame Interpretation beider Rater in nachfolgenden Kodierungen berücksichtigt werden konnte.

3.5 Bildung „naturwissenschaftsdidaktischer Perspektivierungen“ von Unterrichtsqualität

Die Bildung von „naturwissenschaftsdidaktischen Perspektivierungen“ wurde, angesichts der in 1.3.1 beschriebenen möglichen terminologischen und konzeptuellen Unterschiede und Gemeinsamkeiten, als abschließender Schritt der Gegenüberstellung der Qualitätskriterien aus den Videostudien mit dem Syntheseframework durchgeführt. Die grundlegende Idee der Perspektivierung ist es, ein Bindeglied zwischen den Kriterien der naturwissenschaftsdidaktischen Unterrichtsforschung und der generisch formulierten Systematisierung des Syntheseframeworks zu bilden. Durch diese Verbindung soll eine bessere Vergleichbarkeit beider Perspektiven hergestellt und die Kommunikation erleichtert werden. Für die Bildung der naturwissenschaftlichen Perspektivierung wurde die Indikatorebene des Syntheseframeworks genutzt und die Formulierungen überarbeitet, um die fehlenden Aspekte der naturwissenschaftsdidaktischen Qualitätskriterien zu erfassen. Dies terminologische Überarbeitung berücksichtigte sowohl die Formulierungen der Qualitätskriterien aus den Videostudien, als auch die Spezifika der naturwissenschaftlichen Fächer, die in Abschn. 4 näher erläutert werden. Die Formulierungen wurden durch einen ersten Rater erstellt und anschließend durch einen zweiten Rater überarbeitet. Die abschließend gewählte Formulierung der naturwissenschaftlichen Perspektivierung wurde dann in einem Konsensverfahren festgelegt.

Für das genaue Vorgehen bedeutet dies, dass an das systematische Review angeschlossen wurde, indem die Zuordnung der Qualitätskriterien zur Indikatorebene des Syntheseframeworks (siehe Anhang 3) als Basis für die inhaltliche Sammlung der Kriterien genutzt wurde. In dieser Tabelle wurden die Qualitätskriterien aus den Studien bereits unter bestimmten Sammelbegriffen des Syntheseframeworks zusammengefasst, wie z. B. „Logische Sequenzierung der Unterrichtsinhalte“ (Syntheseframework), die teilweise bereits eine Überschneidung zu den (generischen) Terminologien der in den Videostudien verwendeten Qualitätskriterien aufwiesen. Für eine möglichst umfangreiche Abbildung der Qualitätskriterien aus den Videostudien, wurden auch im Fall eher generischer Kriterien neue Formulierungen für die Perspektivierung gewählt, damit alle Aspekte der Qualitätskriterien abgedeckt werden können, wie z. B. „Strukturierter Ablauf und Sequenzierung der Stunde“ (Perspektivierung). Diese eher generischen Perspektivierungen können somit von der Terminologie des Syntheseframeworks abweichen, auch wenn die dazugehörigen Kriterien

zuvor mit einer „vollständigen Übereinstimmung“ in das Syntheseframework eingeordnet wurden. Darüber hinaus wurden die Qualitätskriterien, die zuvor mit einer „teilweisen Übereinstimmung“ oder „starken Abweichung“ kodiert wurden, genauer betrachtet, um die fehlenden Aspekte des Syntheseframeworks terminologisch in die Perspektivierung aufzunehmen. Es wurden somit Begriffe gewählt, die naturwissenschaftsdidaktische Aspekte hervorheben, die zuvor nicht ausreichend durch die generische Formulierung des Syntheseframeworks abgedeckt wurden, wie z. B. „Auswahl naturwissenschaftlicher Denk- und Arbeitsweisen“ (Perspektivierung), im Vergleich zu „Auswahl von bedeutungsvollen, dem Lernstand angemessenen Inhalten sowie Fachmethoden“ (Syntheseframework) oder „Konstruktive Einbindung von eigenen Ideen und Schülervorstellungen in den Unterricht“ (Perspektivierung) im Vergleich zu „Entwicklung und Revision von Konzepten“ (Syntheseframework).

Hierarchisch liegt die Indikatorebene, auf der die naturwissenschaftsdidaktischen Qualitätskriterien zunächst gesammelt wurden, unter der Subdimensionsebene. Die Perspektivierung wird jedoch bewusst neben der Subdimensionsebene dargestellt, um zu verdeutlichen, dass die Qualitätskriterien der naturwissenschaftsdidaktischen Forschung zwar eine grundsätzliche Vergleichbarkeit zum Syntheseframework aufweisen, aber nicht generell vollständig durch dasselbe abgebildet werden können und somit nicht einfach eine fachliche Auslegung eines generischen Aspekts darstellen. Die naturwissenschaftsdidaktische Perspektivierung hat hierbei nicht den Anspruch das Syntheseframework zu ersetzen oder eine Erweiterung der Indikatorebene vorzunehmen, sondern Bezüge zwischen der generischen und der naturwissenschaftsdidaktischen Perspektive herzustellen.

Um abzubilden, in welchem Umfang die naturwissenschaftsdidaktischen Perspektivierungen fachspezifische Aspekte beinhalten, wurde für jede Perspektivierung eine Markierung des Verhältnisses zwischen Fachspezifik und Generik vorgenommen. Diese Markierung erfolgte vergleichbar zum Syntheseframework von Praetorius und Charalambous (2018) und enthält die Kategorien:

1. Generisch (G): Die Aspekte der Perspektivierung sind eher generisch und können ohne fachspezifisches Wissen umgesetzt oder bewertet werden (z. B. Allgegenwärtigkeit).
2. Fachspezifisch (S): Die Aspekte der Perspektivierung sind eher fachspezifisch und setzen ein fachspezifisches Wissen zur Umsetzung oder Bewertung voraus (z. B. Auswahl von Fachinhalten).
3. Generisch mit fachspezifischen Anteilen (G+S): Die Aspekte der Perspektivierung sind großteils generisch, fachspezifische Anteile unterstützen die Umsetzung oder Bewertung jedoch (z. B. Zielklarheit).
4. Verknüpfung generischer und fachspezifischer Aspekte (G*S): Die Aspekte der Perspektivierung haben sowohl generische, als auch fachspezifische Anteile und eine Berücksichtigung beider Perspektiven ist zur Umsetzung oder Bewertung notwendig (z. B. Auswahl herausfordernder Lerngelegenheiten)

Zur Zuordnung der Perspektivierungen zu den unterschiedlichen Kategorien der Fachspezifik wurde zunächst die Kodierung der Subdimensionen aus dem Syntheseframework von Praetorius und Charalambous (2018) auf die Perspektivierung übertragen. Die Perspektivierungen einer Subdimension erhielten somit in diesem Schritt

jeweils dieselbe Kategorie der Fachspezifik. Um die Perspektivierungen innerhalb einer Subdimension präziser zuzuordnen und ggf. Abweichungen zur Zuordnung im Syntheseframework herauszustellen, wurde in einem nächsten Schritt, ausgehend von den zugeordneten Kriterien zu einer Perspektivierung der Grad der Fachspezifik ermittelt. Dies erfolgt auf Basis der Beschreibungen oder Items, die zu den jeweiligen Kriterien innerhalb der Videostudien vorlagen. Für den Fall, dass auf Basis der Kriterien keine genaue Zuordnung zu einer der 4 Kategorien der Fachspezifik getroffen werden konnte, wurde mit Hilfe der theoretischen Ausarbeitung zu den Spezifika der naturwissenschaftlichen Fächer eine Auswahl getroffen. Als Grundlage für die Unterscheidung zwischen Generik und Fachspezifik diente die in Abschn. 1.3.1 beschriebene Abgrenzung. Die Zuordnung zu den Kategorien der Fachspezifik wurde zunächst durch einen Rater getroffen und anschließend mit einem zweiten Rater besprochen. Die abschließende Zuordnung der Perspektivierungen zu diesen Kategorien erfolgte dann in einem Konsensverfahren zwischen beiden Ratern.

Bei der Zuordnung und Bestimmung des fachspezifischen Anteils der Perspektivierungen ergaben sich unterschiedliche Typisierungen des Verhältnisses von Fachspezifik und Generik. Diese werden im Diskussionsteil (6.2) weiter ausgeführt, wurden jedoch für die Darstellung in Tab. 4 und Anhang 4 nicht berücksichtigt.

Damit der Prozess der Bildung und Kategorisierung der Perspektivierung nachvollziehbar und transparent bleibt, werden in Anhang 4 sämtliche Kriterien berichtet, die unter einer Perspektivierung zusammengefasst wurden. Die naturwissenschafts-didaktischen Perspektivierungen, sowie die genaue Anzahl der Qualitätskriterien, die zu einer Perspektivierung zusammengefasst wurden sind Tab. 4 zu entnehmen.

4 Einblicke in ausgewählte Spezifika der Ziele, Inhalte und Methoden der naturwissenschaftlichen Unterrichtsfächer

4.1 Zieldimensionen naturwissenschaftlichen Unterrichts

Zentral für aktuelle nationale und internationale Diskurse über die Ziele schulischen naturwissenschaftlichen Lernens ist der Begriff „scientific literacy“ (Bybee 1997; Gräber et al. 2002; Osborne 2007). Zu dessen Kern gehören das Verfügen und Anwenden fachlich adäquater Vorstellungen, die aktive und auf naturwissenschaftlichem Verständnis beruhende Teilhabe an gesellschaftlich relevanten Diskussionen und Entscheidungsprozessen, Kenntnisse epistemologischer Merkmale der Naturwissenschaften, die Fähigkeit, zu einer ökonomisch bedeutsamen Innovationsfähigkeit eines Landes beizutragen wie auch positive Einstellungen gegenüber den Naturwissenschaften selbst und naturwissenschaftlich-technischen Berufen (siehe dazu insb. Kind 2013; oder Gräber et al. 2002).

Eine grundlegende Kategorisierung von vier Zieldimensionen naturwissenschaftlichen Unterrichts wurde, vor diesem Hintergrund, durch Hodson (2014, S. 2537) vorgelegt:

1. *Learning science*: Der Erwerb von Wissen und Entwicklung naturwissenschaftlich adäquater Vorstellungen über Theorien und Modelle auf einer theoretischen Ebene als Zieldimension von Unterricht.
2. *Learning about science*: Die Entwicklung von epistemologisch adäquaten Vorstellungen über das Wesen der Naturwissenschaften als Zieldimension von Unterricht. Dazu gehören Einsichten in die Sicherheit, aber auch die Grenzen und Veränderbarkeit naturwissenschaftlichen Wissens und dessen Generierung und Überprüfung in naturwissenschaftlicher Forschung. Im internationalen Diskurs hat sich hierfür der Begriff „nature of science“ weitgehend durchgesetzt.
3. *Doing science*: Der Erwerb von Fähigkeiten zur Umsetzung naturwissenschaftlicher Denk- und Arbeitsweisen. Dazu zählt die aktive und eigenständige Umsetzung von Experimenten, Beobachtungen oder die Nutzung von Modellen als Forschungswerkzeuge. Schülerinnen und Schüler sollen danach in die psychomotorischen Aktivitäten eingebunden werden und selbst naturwissenschaftlich untersuchend tätig sein („hands on“), aber auch damit verbundene Denkprozesse (z. B. beim Bilden von Hypothesen oder bei der Interpretation von Daten) nachvollziehen („minds on“).
4. *Addressing socio-scientific issues (SSIs)*: Die Entwicklung einer naturwissenschaftlich fundierten Kritikfähigkeit an gesellschaftlichen, ökonomischen, ökologischen, politischen und persönlichen Problemstellungen und Entscheidungsprozessen als Zieldimension von Unterricht. Die Schülerinnen und Schüler sollen insbesondere an Schnittstellen zwischen Naturwissenschaften, Technik und Gesellschaft die Bedeutung von Wissen kennenlernen und für fundierte Entscheidungen nutzbar machen.

In Anlehnung an die Diskussionen um „scientific literacy“ hat sich in Deutschland der Begriff der naturwissenschaftlichen Grundbildung weitgehend etabliert und wurde in den nationalen Bildungsstandards der KMK (2005a, 2005b, 2005c) verankert, wobei sich eine weitgehende Kongruenz zwischen den vier Zieldimensionen und den deutschen Kompetenzbereichen feststellen lässt, auf die aus Platzgründen nicht eingegangen werden kann.

4.2 Ausgewählte Spezifika naturwissenschaftlicher Lernprozesse

Zwar lassen sich Inhalte der naturwissenschaftlichen Fächer anhand konkreter, lebensweltlicher oder fachlicher Phänomene motivieren und ganzheitlich erlebbar machen (Muckenfuss 2001; Wagenschein 1976). Das Wesen und der Erfolg der Naturwissenschaften als wissenschaftliche Disziplinen bestehen ja gerade aber darin, von diesen Phänomenen zu abstrahieren (Talanquer 2011; Treagust et al. 2003) und eine Nutzung von naturwissenschaftlichen Modellvorstellungen anzubahnen, die diese Phänomene beschreiben, erklären oder vorhersagen können. Für diese Unterscheidung existieren eine Fülle von Begrifflichkeiten in den verschiedenen naturwissenschaftsdidaktischen Disziplinen, wie z. B. die der „Erfahrungswelt“, den Sinnen zugänglichen Bereichen der Naturphänomene, und der „Modellwelt“, den Denkbereichen der naturwissenschaftlichen Theorien, wie sie von Mikelski-Seifert und Fischler (2003) vorgeschlagen wurden.

Häufig steht naturwissenschaftlicher Unterricht vor der Anforderung, an phänomenorientierten Erfahrungen der Schülerinnen und Schüler anzuknüpfen oder diese Erfahrungen, z. B. in Versuchen oder Experimenten, gezielt zu stiften („Erfahrungswelt“), gleichzeitig aber so davon zu abstrahieren, dass fachlich tragfähiges Theoriewissen („Modellwelt“) auf- und ausgebaut werden kann – wobei dieser Auf- und Ausbau wiederum den Umgang mit fachspezifischen und teilweise komplexen Repräsentationen einfordert. Mit Blick auf die Frage der exemplarischen Erweiterung des Syntheseframeworks sei auf eine Auswahl von weiteren Charakteristika verwiesen, die angesichts der Fülle an Facetten und Forschungsarbeiten nicht abschließend sein kann:

1. Die Schülerinnen und Schüler kommen mit zahlreichen Erfahrungen und individuellen Vorstellungen über die Phänomene der Natur in den Unterricht. Diese Erfahrungen bilden den Ausgangspunkt ihrer Lernprozesse (Gropengießer und Kattmann 2013) und bleiben bei Nichtberücksichtigung im Unterricht weiterbestehen. Verstärkt werden diese Prozesse durch die Problematik der Polysemie, bei der Fachwörter andere Bedeutungen haben als im Alltag („Wärme“, „Reaktion“, „Teilchen“). In der Erforschung und Weiterentwicklung von Schülervorstellungen finden naturwissenschaftsdidaktische Forschungen eines ihrer zentralen Themenfelder; für jede der drei naturwissenschaftlichen Fächer liegen Überblickswerke zu Schülervorstellungen vor (Kattmann 2015; Barke 2006; Schecker et al. 2018). Vor dem Hintergrund eines konstruktivistischen Lernverständnisses greift guter naturwissenschaftlicher Unterricht diese Schülervorstellungen auf und macht sie, wie z. B. im Rahmen der Arbeiten zur didaktischen Rekonstruktion (Duit et al. 2012) beschrieben, zum Gegenstand der Planung und Durchführung.
2. Phänomene bilden häufig den Ausgangspunkt für die Erarbeitung von fachlich adäquaten Vorstellungen der „Modellwelt“, jedoch können verschiedene Phänomene in den Unterricht eingebunden werden, die diesen Zugang ermöglichen. Mit der Entscheidung für oder gegen ein spezifisches Phänomen ergeben sich Konsequenzen für die Sachstruktur des Unterrichts, für mögliche Schülervorstellungen oder gar den weiterführenden Verlauf des Unterrichts über einen längeren Zeitraum hinweg. Bestimmte Experimente eignen sich besser oder schlechter, um z. B. spezifische Fragestellungen aufzuwerfen oder einen spezifischen Effekt in den Unterricht zu integrieren (Sommer und Pfeifer 2018). Die im Syntheseframework beschriebene „Auswahl und Thematisierung von Inhalten und Fachmethoden“ ist in besonderer Weise auch durch die Auswahl von (z. B. experimentellen) Zugängen auf Phänomenebene gekennzeichnet.
3. Die „Erfahrungswelt“ und die „Modellwelt“ lassen sich in verschiedene Ebenen differenzieren, deren Durchdenken und Verknüpfung vertieftes fachliches Verständnis und naturwissenschaftliche Kompetenz ermöglichen (Klein et al. 2018):
 - Physik (nach Mikelski-Seifert und Fischler 2003): Alltagserfahrungen, Experimente, ikonische Modelle, verbale Modellaussagen, simulative Modelle
 - Chemie (nach Johnstone 1991): Makro-, Submikro-, Symbolebene (für einen Überblick der zahlreichen Ausdifferenzierungen: Sjöström und Talanquer 2014; Talanquer 2011)

- Biologie (nach Hammann 2019): Biosphäre, Ökosystem, Lebensgemeinschaft, Organismus, Organsysteme und Organe, Gewebe, Zellen, Organellen, Moleküle.

Der Blick für Unterrichtsqualität erhält auf diesen Ebenen wesentliche Impulse dadurch, dass ein Phänomen der Erfahrungswelt aus unterschiedlichen Ebenen der Modellwelt erklärt und vorhergesagt werden kann. Vertieftes Lernen und Verständnis eines Phänomens vollziehen sich häufig in einem Durchschreiten dieser Ebenen. Die Fähigkeit, fachlich sinnvolle Bezüge zwischen diesen Ebenen herstellen zu können, ist zentral für das Verständnis. Auch Beschreibung von zunehmender naturwissenschaftlicher Expertise durch den Umgang zunehmend vernetzter und komplexer Sachverhalte (Bernholt et al. 2009; Kauertz et al. 2010) basiert auf dieser grundlegenden Idee. Im Sinne von Spiralcurricula kann das Unterrichtsangebot so aufgebaut sein, dass Phänomene zu mehreren Zeitpunkten thematisiert werden, aber jeweils auf neuen Ebenen oder auf Grundlage zunehmender Komplexität betrachtet werden können. Naturwissenschaftsdidaktische Forschungen verweisen dabei auch darauf, dass Zusammenhänge zwischen dem Komplexitätsniveau des Unterrichtsangebotes und der Schülerinnen und Schüler existieren können (Podschuweit et al. 2016) und gerade die Passung dazwischen ein bedeutsamer Faktor ist (Lau 2011).

4. Gleichzeitig lassen sich jedoch in verschiedenen naturwissenschaftlichen Themenfeldern und Theorien grundlegende Ideen bzw. Denkweisen identifizieren, die als häufig wiederkehrend und inhaltlich fundamental für naturwissenschaftliches Verständnis angesehen werden. Zu solchen „central ideas“ oder „core ideas“ gehören z. B. der diskontinuierliche Aufbau der Materie, die Erhaltung von Energie und der Masse (Talanquer 2015). Die Bildungsstandards der KMK (2005a, 2005b, 2005c) definieren Basiskonzepte, zu denen in verschiedenen Inhalten der Curricula Bezüge hergestellt werden sollen (z. B. Biologie: Struktur und Funktion; Chemie: Struktur-Eigenschaft; Physik: Wechselwirkung). Die oben angesprochenen Spiralcurricula können anhand derartiger zentraler Ideen oder Basiskonzepten organisiert sein.
5. Häufig diskutiert ist das im Mittel abfallende Interesse an Naturwissenschaften mit zunehmender Schulungsdauer, das bereits in Deutschland in den 1990er Jahren systematisch nachgewiesen wurde (Gräber 1992) und auch aktuell zu bestehen scheint (Reiss et al. 2016). Während aktuelle Forschung auf ein komplexes Wechselspiel zwischen Wissen und Interesse hinweist (Höft et al. 2019), kann auf affektiver Ebene nicht nur das Interesse, sondern auch emotionale Reaktionen wie Angst, Respekt oder Ekel eine Rolle spielen, so dass eine inhaltsbezogene emotionale Unterstützung notwendig wird. Beispielhaft sei hier auf das Sezieren von Organen oder die Arbeit mit lebendigen Tieren, wie Mäusen oder Schaben, in der Biologie verwiesen (Polte und Wilde 2018).

4.3 Ausgewählte Spezifika naturwissenschaftlicher Methoden

Die Stiftung von Erfahrungen an Phänomenen verläuft häufig im Kontext von Beobachtungen oder Experimenten. Während mit dem Begriff „Experimentieren“ in

einer fachdidaktisch unbedarften Verwendung jedwede Aktivität an konkreten naturwissenschaftlichen Phänomenen (Gyllenpalm und Wickman 2011) und auch unter „Beobachten“ ein „Schauen“, „Gucken“ oder „Betrachten“ gemeint sein kann, werden im naturwissenschaftsdidaktischen Diskurs diese Begriffe vor allem genutzt, um eigenständige naturwissenschaftliche Arbeitsweisen zu bezeichnen (Gropengießer 2009; Nehring et al. 2016; Wellnitz und Mayer 2013).

Das bedeutet, dass mit den Experimenten oder Beobachtungen ein wissenschaftstheoretisch fundiertes Vorgehen zur Generierung neuen Wissens bezeichnet wird. Schülerinnen und Schüler sollen an Experimenten nicht nur Merkmale naturwissenschaftlicher Phänomene erlernen oder naturwissenschaftliches Wissen aus der „Modellwelt“ erarbeiten und anwenden („knowing science“/in den Standards der KMK: Kompetenzbereich Fachwissen), sondern auch die Umsetzung dieser Arbeitsweisen nachvollziehen und verstehen („knowing about science“/in den Standards der KMK: Kompetenzbereich Erkenntnisgewinnung) sowie befähigt werden, diese Arbeitsweisen selbstständig und problemlösend umzusetzen („doing science“/in den Standards der KMK: Kompetenzbereich Erkenntnisgewinnung). Das Experimentieren kann daher mit dem Ziel des „knowing science“ implementiert werden – es wäre dann eine Methode zum Erwerb von Fachwissen – aber auch zum Erlernen der Arbeitsweise des Experimentierens an sich. Es kann aber auch derart im Mittelpunkt von Unterrichtsphasen stehen, dass es selbst zum genuinen Inhalt des Unterrichts wird (Gyllenpalm und Wickman 2011). Das Experimentieren ist dann Ziel des Unterrichts.

Zahlreiche Studien verweisen darauf, dass für eine Förderung des „doing science“ explizit reflexive Unterrichtsphasen mit Fokus auf die Denk- und Arbeitsweisen besonders lernwirksam sind (Schwichow et al. 2016b; Vorholzer et al. 2018). Gleichzeitig sollten die Denk- und Arbeitsweisen, auch bei einer didaktischen Reduktion wissenschaftstheoretisch so angemessen sein, dass kein stark vereinfachtes oder verzerrtes Bild der Naturwissenschaften vermittelt wird („knowing about science“). Inwiefern das mit welchen Ergebnissen der Fall ist und was wissenschaftstheoretisch adäquat ist, ist wiederum Gegenstand eigenständiger und umfangreicher naturwissenschaftsdidaktischer Debatten auf nationaler (z. B. Höttecke und Rieß 2015) und internationaler Ebene (z. B. Chinn und Malhotra 2002). Dass die grundlegende Voraussetzung hierfür die Beachtung von Sicherheitsregeln, sowie die Durchführung einer Gefahrstoffprüfung im Falle eines Einsatzes von Geräten und Chemikalien, ist, sei an dieser Stelle selbstverständlich auch erwähnt (Unfallkasse NRW 2018).

5 Ergebnisse

5.1 Ergebnisse des systematischen Reviews, Verortung von Kriterien im Syntheseframework und Erweiterung des Syntheseframeworks um naturwissenschaftsdidaktische Subdimensionen

Zur Beantwortung der ersten Forschungsfrage sei an dieser Stelle zunächst auf die vollständige Übersicht der Qualitätskriterien in Anhang 3 verwiesen, welche die Grundlage für die weitere Auswertung darstellt. Um den Umfang der Zuordnung

in ein übersichtliches Format zu überführen, wurde eine vollständige Auszählung der verorteten Kriterien vorgenommen, die im Folgenden näher erläutert wird und der Frage nach dem Vergleich der empirisch verwendeten Kriterien mit dem Syntheseframework nachgeht. Für eine exemplarische Verortung einzelner Kriterien mit Hilfe der Abstufungen der Übereinstimmung siehe Abschn. 3.4.

Bei der Verortung der Qualitätskriterien aus den Videostudien im Syntheseframework ergab sich eine Verteilung der naturwissenschaftsspezifischen Kriterien auf das Syntheseframework (Tab. 3) mit einer „vollständigen Übereinstimmung“ bei 72 %, einer „teilweisen Übereinstimmung“ bei 21 % und einer „starken Abweichung“ bei 7 % der Kriterien. Besonders auffällig ist hierbei, dass der Großteil der Kriterien in den Dimensionen *I. Content selection and presentation*, *II. Cognitive activation* und *VII. Classroom and time management*, verortet werden konnten.

Auf Ebene der Subdimensionen treten einige Stellen hervor, an denen ein besonders hoher Anteil „teilweiser Übereinstimmung“ beobachtet werden kann. *Time management*, *Presenting the content in mathematically accurate and correct ways*, *Teacher supports students solidify their procedural knowledge/skills*, *Forming an environment that nurtures productive habits* und *Enhancing participation and the active engagement of all students* besitzen zwar einen hohen Anteil „teilweiser Übereinstimmung“, beinhalten jedoch nur einen verhältnismäßig geringen Anteil der insgesamt verorteten Kriterien, wodurch diese Abweichung etwas relativiert wird. Dennoch bieten diese Subdimensionen einen wichtigen Ansatzpunkt für eine Erweiterung des Syntheseframeworks. Auffälliger sind die Subdimensionen *Potential for cognitive activation (a & b)* und *Teacher facilitation of students' cognitive activity*, die jeweils 12 % der im Syntheseframework verorteten Kriterien beinhalten und somit bereits knapp ein Viertel der insgesamt verorteten Kriterien abbilden. Die Subdimension *Presenting the content in a structured way* bietet einen weiteren interessanten Ansatz, da hier ein großer Anteil der Kriterien in einer einzelnen Subdimension verortet wurde (19 %). Der Anteil „teilweise Übereinstimmung“ aller drei genannten Subdimensionen befindet sich jedoch im erwarteten Bereich von ca. 20 %. Der Anteil der Kriterien mit einer „starken Abweichung“ konnte vollständig der Dimension *II. Cognitive activation* zugeordnet werden und bietet einen direkten Hinweis für eine fachspezifische Ergänzungsstelle.

Wie in 3.5 beschrieben, erfolgte eine Bildung von Perspektivierungen zur Abbildung von Qualitätskriterien aus naturwissenschaftsdidaktischen Videostudien. Tab. 4 stellt den Vergleich zwischen dem erweiterten generischen Syntheseframework und der naturwissenschaftsdidaktischen Perspektive dar.

Die Definition der Sammelbegriffe in der naturwissenschaftsdidaktischen Perspektive erlaubte es, – in einer erneuten Kodierung der Kriterien aus den Videostudien in die Sammelbegriffe – einen Teil der Kriterien mit „teilweise Übereinstimmung“ oder „starker Abweichung“ zuzuordnen. Aus pragmatischen Gründen wird der Vergleich zum Syntheseframework direkt in dieser Version dargestellt, eine ausführliche Darstellung der Verortung der Qualitätskriterien in den Sammelbegriffen befindet sich in Anhang 4. Insgesamt stieg die Anzahl der vollständigen Übereinstimmungen damit auf 95 % Prozent und die Anzahl der Kriterien in den unterschiedlichen Subdimensionen kann damit zwischen Tab. 3 und 4 differieren.

Tab. 3 Ergebnis des Vergleichs naturwissenschaftsdidaktischer Qualitätskriterien mit dem Syntheseframework nach Praetorius und Charalambous (2018)

Dimensionen und Subdimensionen der Unterrichtsqualität nach Praetorius und Charalambous (2018)	Anzahl der kodierten naturwissenschaftsdidaktischen Kriterien			Σ	Verteilung		
	Vollständige Übereinstimmung	Teilweise Übereinstimmung	Starke Abw		RH T. Ü. (%)	Anteil T. Ü. gesamt (%)	RH Ges (%)
<i>I. Content selection and presentation</i>							
Selecting mathematically worthwhile and developmentally appropriate content	25	4	–	29	13,8	4,9	7,5
Motivating the content	20	3	–	23	13,0	3,7	5,9
Presenting the content in a structured way	57	16	–	73	21,9	19,5	18,8
Presenting the content in mathematically accurate and correct ways	7	5	–	12	41,7	6,1	3,1
<i>II. Cognitive activation</i>							
Potential for cognitive activation through:			26	26			
a) Teacher's selection of challenging tasks which respond to students' cognitive level	19	3	–	22	13,6	3,7	5,7
b) Teacher's use of mathematically rich practices	18	5	–	23	21,7	6,1	5,9
Teacher facilitation of students' cognitive activity	42	8	–	50	16,0	9,8	12,9
Teacher supports students' meta-cognitive learning from cognitively activating tasks	16	4	–	20	20,0	4,9	5,2
<i>III. Practicing</i>							
Teacher supports students solidify their procedural knowledge/skills	3	2	–	5	40,0	2,4	1,3
Teacher procedural remediation of students' difficulties and errors in practicing	0	0	–	0	0,0	0,0	0,0
<i>IV. (Formative) Assessment</i>							
Assessment is aligned with learning objectives	1	0	–	1	0,0	0,0	0,3
Teacher regularly checks for understanding	2	0	–	2	0,0	0,0	0,5

Tab. 3 (Fortsetzung)

Dimensionen und Subdimensionen der Unterrichtsqualität nach Praetorius und Charalambous (2018)	Anzahl der kodierten naturwissenschaftsdidaktischen Kriterien			Σ	Verteilung		
	Vollständige Übereinstimmung	Teilweise Übereinstimmung	Starke Abw		RH T. Ü. (%)	Anteil T. Ü. gesamt (%)	RH Ges (%)
Quality of feedback for students	7	0	–	7	0,0	0,0	1,8
Teacher capitalizes on formative assessment information to guide next instructional steps	1	0	–	1	0,0	0,0	0,3
<i>V. Cutting-across instructional aspects aiming to maximize student learning</i>							
Forming an environment that nurtures productive habits	8	6	–	14	42,9	7,3	3,6
Differentiation and adaptation	9	1	–	10	10,0	1,2	2,6
Enhancing participation and the active engagement of all students	6	8	–	14	57,1	9,8	3,6
<i>VI. Socio-emotional support</i>							
Teacher-student relationships	7	2	–	9	22,2	2,4	2,3
Student-student relationships	0	0	–	0	0,0	0,0	0,0
<i>VII. Classroom and time management</i>							
Behavior management	22	5	–	27	18,5	6,1	7,0
Time management	10	10	–	20	50,0	12,2	5,2
Σ	280	82	26	388			
Relative Häufigkeit	72 %	21 %	7 %	–	–	–	–

T.Ü. teilweise Übereinstimmung, RH Relative Häufigkeit

Neben der theoriebasierten Ergänzung von Kriterien durch die Reflexion der Ziele, Inhalte und Methode der naturwissenschaftlichen Fächer ergeben sich aus der Befundlage im Bereich der videostudienbasierten Kriterien mit teilweiser Übereinstimmung, zusätzlich zur fachspezifischen Auslegung der Subdimensionen, weitere Ergänzungen. Dabei handelt es sich um:

II. Kognitive Aktivierung

- *Kooperatives Arbeiten zur zielführenden Aktivierung der Schüler/-innen:* Einzelne Aspekte des kooperativen Arbeitens, wie es in den Videostudien beschrieben wird, können zwar im Syntheseframework gefunden werden (z. B. „[...] individu-

Tab. 4 Erweitertes Syntheseframework generischer und naturwissenschaftsdidaktischer Perspektivierungen auf Unterrichtsqualität

Unterrichtsqualität in Perspektivierungen der allgemeinen Unterrichtsforschung: Dimensionen und Subdimensionen	Unterrichtsqualität in Perspektivierungen naturwissenschaftsdidaktischer Forschung: Subdimensionen formuliert als Sammelbegriffe für Qualitätskriterien (Verhältnis von Generik und Fachspezifik) (Anzahl der Kriterien die unter einem Begriff gesammelt wurden)
I. Auswahl und Thematisierung von Inhalten und Fachmethoden	
Auswahl von bedeutungsvollen, dem Lernstand angemessenen Inhalten sowie Fachmethoden	Auswahl und Einbindung von Fachinhalten (S) (10 Kriterien) Auswahl und Einbindung naturwissenschaftlicher Denk- und Arbeitsweisen (S) (23 Kriterien)
Motivierung von Inhalten sowie Fachmethoden	Lernen im Kontext und Verknüpfung zu gesellschaftsrelevanten Problemstellungen (G*S) (19 Kriterien) Motivierende Einbettung der Inhalte (G*S) (5 Kriterien)
Strukturierung der thematisierten Inhalte sowie Fachmethoden	Zielklarheit (G + S) (7 Kriterien) Strukturieren und Fokussieren von Schlüsselaspekten (G*S) (6 Kriterien) Horizontale und vertikale Vernetzung (S) (17 Kriterien) Strukturierter Ablauf und Sequenzierung der Stunde (G*S) (18 Kriterien) Progression innerhalb der Stunde (G*S) (1 Kriterium) Angemessene Repräsentation der Inhalte (G*S) (9 Kriterien)
Akkuratheit und Korrektheit der thematisierten Inhalte sowie Fachmethoden	Nutzung präziser Fachsprache (S) (6 Kriterien) Vermeidung inhaltlicher Fehler (S) (2 Kriterien) Fachlich adäquate Einbindung von Inhalten und Denk- und Arbeitsweisen (S) (5 Kriterien)
<i>Ergänzungen aus fachspezifischen Untersuchungen</i>	Adäquate didaktische Reduktion unter Berücksichtigung zukünftiger fachlicher Lernschritte (theoriebasiert ergänzt aus Fachspezifika)
II. Kognitive Aktivierung	
Auswahl fachlich gehaltvoller und auf das kognitive Niveau der Schüler*innen abgestimmter Aufgaben	Auswahl herausfordernder Lerngelegenheiten (G*S) (18 Kriterien)
Einsatz fachlich gehaltvoller Aufgaben	Problemlösendes Lernen (G*S) (8 Kriterien) Nutzung multipler Repräsentationen und Lösungswege (G + S) (3 Kriterien) Einsatz komplexer und vernetzender Aufgaben (G*S) (9 Kriterien) Kognitiv aktivierender Einsatz naturwissenschaftlicher Denk- und Arbeitsweisen (S) (11 Kriterien)
Unterstützung der kognitiven Aktivität der Schüler*innen	Aktivierung von Vorwissen (G*S) (7 Kriterien) Kognitiv-konstruktive Fehlerkultur (G*S) (6 Kriterien) Explizierung von Denkprozessen (G*S) (9 Kriterien) Konstruktive Einbindung von eigenen Ideen und Schülervorstellungen in den Unterricht (G*S) (22 Kriterien) Unterstützung kognitiv aktivierender Prozesse (G + S) (14 Kriterien)
Unterstützung des metakognitiven Lernens der Schüler*innen anhand kognitiv aktivierender Aufgaben	Unterstützung und Einbindung metakognitiver Prozesse (G*S) (12 Kriterien) Explizierung von naturwissenschaftlichen Denk- und Arbeitsweisen (G*S) (3 Kriterien)
<i>Ergänzungen aus fachspezifischen Untersuchungen</i>	Kooperatives Arbeiten zur zielführenden Aktivierung der Schüler*innen (G*S) (13 Kriterien)

Tab. 4 (Fortsetzung)

Unterrichtsqualität in Perspektivierungen der allgemeinen Unterrichtsforschung: Dimensionen und Subdimensionen	Unterrichtsqualität in Perspektivierungen naturwissenschaftsdidaktischer Forschung: Subdimensionen formuliert als Sammelbegriffe für Qualitätskriterien (Verhältnis von Generik und Fachspezifik) (Anzahl der Kriterien die unter einem Begriff gesammelt wurden)
III. Unterstützung des Übens	
Unterstützung bei der Festigung von Lern- und Anwendungsprozessen	Wiederholende Anwendung von Fachinhalten und Methoden (G*S) (4 Kriterien)
Konstruktiver Umgang mit Fehlern und Schwierigkeiten von Schüler*innen beim Üben	–
IV. Formatives Assessment	
Klare Ausrichtung der Beurteilung auf die zu erlernenden Kompetenzen	Unterrichtsbezogene Rückmeldung (G + S) (1 Kriterium)
Regelmäßige Überprüfung des Verständnisses der Schüler*innen	Regelmäßige Überprüfung des schülerseitigen Verständnisses (G + S) (3 Kriterien)
Qualitativ hochwertiges Feedback an die Schüler*innen	Konstruktives Feedback (G*S) (6 Kriterien)
Nutzung des Feedbacks als Grundlage für die Ausrichtung des weiteren Unterrichts	Konstruktive Nutzung des Feedbacks (G*S) (1 Kriterium)
V. Unterstützung des Lernens aller Schüler*innen	
Bereitstellung eines Lernumfelds, das produktives Verhalten fördert (z. B. Eigenverantwortung, eigenständiges Lernen, Identität, Durchhaltevermögen)	Autonomie der Schüler/-innen (G*S) (8 Kriterien) Selbstwahrnehmung der Schüler/-innen (G) (2 Kriterien)
Differenzierung und Adaptivität	Individualisierung/Differenzierung (G*S) (8 Kriterien) Adaptive Erleichterung (G*S) (2 Kriterien)
Förderung der aktiven Mitwirkung von allen Schüler*innen	Förderung der Schülerbeteiligung (G) (3 Kriterien) Lernbegleitende Unterstützung der Schüler/-innen (G*S) (14 Kriterien)
VI. Sozio-emotionale Unterstützung	
Beziehung zwischen Lehrperson und Schüler*innen	Unterstützende Lehrer-Schüler-Interaktion (G) (9 Kriterien)
Beziehung der Schüler*innen untereinander	–
<i>Ergänzungen aus fachspezifischen Untersuchungen</i>	Beziehung zum Inhalt (Antizipation von Angst, Respekt oder Ekel; theoriebasiert ergänzt aus Fachspezifika)
VII. Klassenführung	
Verhaltensmanagement	Allgegenwärtigkeit (G) (6 Kriterien) Prävention (G) (15 Kriterien) Intervention (G) (2 Kriterien)
Zeitmanagement	Zeitliche Strukturierung der Stunde (G*S) (9 Kriterien) Phasenübergänge (G*S) (2 Kriterien) Pacing (G + S) (8 Kriterien)
<i>Ergänzungen aus fachspezifischen Untersuchungen</i>	Sicherheit (S) (2 Kriterien) Raum- und Materialmanagement (theoriebasiert ergänzt aus Fachspezifika)

ell oder in Gruppen gehaltvolle Aufgaben zu bearbeiten.“ in der Subdimension „Unterstützung der kognitiven Aktivität der Schüler/-innen“). In den Videostudien wird das kooperative Arbeiten jedoch anhand von 13 Kriterien als spezifisches Mittel zur kognitiven Aktivierung beschrieben. Hierbei wird es häufig stark ausdifferenziert betrachtet und teilweise naturwissenschaftsspezifisch auf praktische Arbeitsphasen oder daran anschließende Diskussionen ausgelegt. Im Syntheseframework kann es dagegen nur als eine austauschbare Methode erfasst werden und nimmt somit nicht denselben Stellenwert ein, wie in den Videostudien. Der Fokus des Syntheseframeworks liegt hierbei auf der kognitiven Aktivierung, wie diese erreicht wird steht jedoch nicht im Fokus. Aus der Untersuchung naturwissenschaftsdidaktischer Unterrichtsforschung ging dagegen hervor, dass ein kooperatives Arbeiten, je nach Zielstellung des Unterrichts, der Einzelarbeit gegenüber zu bevorzugen ist, was mit dem Syntheseframework nicht erfasst werden kann. Durch die Fokussierung des kooperativen Arbeitens auf bestimmte Aspekte des Unterrichts, beispielsweise auf die Durchführung von Experimenten, kann es zu einer Überschneidung zu bereits bestehenden Perspektivierungen kommen. Da jedoch die besondere Rolle des kooperativen Arbeitens durch diese „teilweisen Übereinstimmungen“ nicht abgedeckt werden kann und die differenzierte Betrachtung des kooperativen Arbeitens sich auf mehr als eine Subdimension des Syntheseframeworks erstrecken würde, wird es als eine Ergänzung aufgeführt.

VII. Klassenführung

- *Sicherheit*: Der Sicherheitsaspekt stellt im naturwissenschaftlichen Unterricht besonders beim Experimentieren ein zentrales Thema dar. Bei Schulz (2011) wird die Sicherheit beim Experimentieren durch die Kriterien *personale Sicherheit* und *Umgang mit Geräten und Chemikalien* abgedeckt. Diese Aspekte werden bisher nicht im Syntheseframework von Praetorius und Charalambous (2018) erfasst, sollten aber auf Basis der zentralen Stellung des Experiments im naturwissenschaftlichen Unterricht, berücksichtigt werden.

5.2 Einblicke in ausgewählte Spezifika der Ziele, Inhalte und Methoden der naturwissenschaftlichen Unterrichtsfächer

Im Abschn. 4 wurden bereits Bezüge zwischen der Beschreibung von Zielen, Inhalten und Methoden der naturwissenschaftlichen Fächer und der Unterrichtsqualitätsforschung angedeutet. Im Folgenden wird dargestellt, inwiefern sich exemplarische Erweiterungen von Kriterien im Syntheseframework ergeben, was zur Beantwortung der zweiten Forschungsfrage beitragen soll. Dabei wird sich an den Dimensionen des Syntheseframeworks orientiert und auf Ebene der Subdimensionen ergänzt. Die Ergänzungen werden jeweils kursiv dargestellt.

I. Auswahl und Thematisierung von Inhalten und Fachmethoden

- *Adäquate didaktische Reduktion unter Berücksichtigung zukünftiger fachlicher Lernschritte*: Lehrkräfte sind bei der Auswahl, bei der Aufbereitung und bei der

didaktischen Reduktion von Inhalten gefordert, ihre Schülerinnen und Schüler nicht bei zukünftigen Lernschritten zu behindern („Prinzip der Ausbaufähigkeit“, Risch und Peifer 2018). So existieren für den Chemieunterricht Hinweise, die Teilchen des einfachen Teilchenmodells nicht allesamt kugelförmig zu repräsentieren. Eine „fachliche Sackgasse“ für zukünftige Lernschritte ergibt sich dann, wenn der Molekülbegriff erarbeitet wird und dabei deutlich wird, dass nicht alle Teilchen kugelförmig sind, sondern Moleküle über eine spezifische Geometrie verfügen und Atomen die Repräsentation der Kugel zugeordnet wird.

VI. Sozio-emotionale Unterstützung

- *Beziehung zum Inhalt (Motivierung, Förderung von Interesse an Inhalten, Antizipation von Angst, Respekt oder Ekel)*: Gerade vor dem Hintergrund eines abnehmenden Interesses an den naturwissenschaftlichen Fächern und Fachinhalten mit zunehmenden Alter ist die Motivierung und Unterstützung einer Motivations- und Interessensentwicklung nicht nur lernunterstützend, sondern auch als eigenständiges Unterrichtsziel anerkannt. Aber auch das Empfinden von Gefahr, z. B. beim Anzünden eines Brenners im Anfangsunterricht, oder das Empfinden von Ekel, z. B. beim Sezieren von Schweineherzen im Biologieunterricht, sind emotionale Bezüge zum Inhalt, die im Sinne einer fachspezifischen Unterrichtsqualität antizipiert oder aufgegriffen werden sollten.

VII. Klassenführung

- *Raum- und Materialmanagement*: In diesem Kontext ist auch der Raum bedeutsam, der für experimentelle Tätigkeiten genutzt werden muss. Experimentiertische sollten abgeräumt sein, Kittel und Schutzbrillen sollten bereitliegen. Auch ausreichend viele Experimentiermaterialien oder auch funktionierende Gas- oder Wasserhähne oder Mikroskope sind Aspekte, um naturwissenschaftliches Lernen überhaupt zu ermöglichen. Veraltete Lösungen oder unreine Chemikalien, z. B. eingetrübte Calciumhydroxidlösungen, in denen sich schon Kohlenstoffdioxid aus der Luft gelöst hat, verhindern erfolgreiches Arbeiten im Unterricht. Der oben beschriebene Aspekt der Sicherheit wird durch Kriterien aus Videostudien ergänzt. Beim *Raum- und Materialmanagement* geht es nicht lediglich um einen besseren Unterrichtsfluss, vielmehr werden Aspekte beschrieben, die für den naturwissenschaftlichen Unterricht eine notwendige Voraussetzung darstellen.

6 Diskussion

6.1 Vergleichbarkeit von naturwissenschaftsdidaktischen Forschungen mit dem generischen Syntheseframework

Zur Diskussion der Befundlage dieses Artikels werden drei Perspektiven angelegt: erstens, die Perspektive des Syntheseframeworks auf naturwissenschaftsdidaktische Unterrichtsforschungen, zweitens, die Perspektive der naturwissenschaftsdidaktischen Unterrichtsforschungen auf das Syntheseframework, drittens die Perspektive

der Fachspezifika auf die Unterrichtsforschung und das Syntheseframework. Durch dieses Vorgehen sollen die Forschungsfragen zusammengeführt werden, um somit den Vergleich der naturwissenschaftsdidaktischen Unterrichtsforschungen mit dem Syntheseframework im Rahmen dieser Arbeit abzuschließen.

Aus der ersten Perspektive wird deutlich, dass naturwissenschaftsdidaktische Studien zahlreiche, der in Dimensionen und Subdimensionen abgebildeten, Aspekte des Syntheseframeworks in den Blick genommen haben. Auch wenn sich eine deutliche Fokussierung auf die Dimensionen I (Auswahl und Thematisierung von Inhalten und Fachmethoden) und II (Kognitive Aktivierung) nachweisen lässt, erscheint der Stand der videobasierten Unterrichtsforschung in den Naturwissenschaftsdidaktiken als vergleichsweise breit. Vermutlich liegen gerade in den – in Dimension I und II abgebildeten – Facetten von Unterrichtsqualität die größten Verknüpfungen zum Umgang mit Fachinhalten und fachspezifischen Denk- und Arbeitsweisen, so dass hier das Kerngeschäft naturwissenschaftsdidaktischer Forschung abgebildet wird. Damit stellt sich aber auch die Frage, ob nicht auch andere Facetten von Unterrichtsqualität, z. B. im Falle der Unterstützung des Übens, des formativen Assessments, der sozio-emotionalen Unterstützung oder Differenzierung, in den Blick genommen werden können, so dass sich auf der einen Seite vielversprechende Schnittmengen zur generisch orientierten Unterrichtsforschung zeigen sowie, auf der anderen Seite, Fragen nach fachspezifischen Operationalisierungen dieser Felder aufgeworfen werden. Welche Rolle spielt das Üben für „guten“ naturwissenschaftlichen Unterricht? Gibt es naturwissenschaftsspezifische Formen des formativen Assessments und wie und mit welchen Effekten werden sie in den Unterricht eingebunden? In diesem Sinne lässt sich das Syntheseframework als Reflexionsschablone für die Ausrichtung und Schwerpunktsetzung von fachdidaktischer Forschung verstehen.

In der zweiten Perspektive zeigt sich das Syntheseframework als weitgehend tragfähig für einen Vergleich und eine Systematisierung naturwissenschaftsdidaktischer Unterrichtsforschungen. Dies zeigt sich vor allem darin, dass der Anteil von Kriterien, die mit „teilweiser Übereinstimmung“ oder „starker Abweichung“ in das Syntheseframework eingeordnet wurden, mit circa einem Viertel der Kriterien vergleichsweise gering erscheint. Gründe für Abweichungen liegen vor allem in der stärkeren Ausdifferenzierung fachspezifischer Qualitätskriterien, die dafür sorgen, dass die Beschreibungen der Subdimensionen im Syntheseframework nicht ausreichen um diese vollständig abzubilden. Des Weiteren können einige der fachspezifischen Kriterien lediglich durch mehrere Subdimensionen des Syntheseframeworks gleichzeitig erfasst werden, wodurch die betreffenden Kriterien nicht eindeutig einer Subdimension zugeordnet werden können. Ein letzter Abweichungsgrund liegt in der schülerseitigen Auslegung und somit Fokussierung der Angebotsnutzung einiger, in den Videostudien verwendeter, Kriterien, die bisher im Syntheseframework nur in stark reduzierter Form erfasst werden. Eine Erfassung der Nutzung des Lernangebotes ist somit nicht mit demselben Grad der Ausdifferenzierung möglich wie die Erfassung des Lernangebotes selbst. Das verdeutlicht, dass weniger grundlegend-konzeptuelle Abweichungen bestehen, sondern vielmehr unterschiedliche Logiken in der Systematisierung von Kriterien für Unterrichtsqualität angewendet werden. Die nicht stets gegebene Trennschärfe kann bei der Zusammenführung von Studienlagen Probleme aufwerfen, die im konkreten Fall zu diskutieren und zu lösen

sind. Die Befunde dieses Artikels können einen Hinweis darauf geben, bei welchen Kriterien das insbesondere der Fall sein kann.

Der Befund der hohen Übereinstimmung zwischen Syntheseframework und Videostudien (bei gleichzeitiger naturwissenschaftsdidaktischer Konkretisierung) mag darin begründet sein, dass zahlreiche Arbeiten der naturwissenschaftsdidaktischen Unterrichtsforschung auch generisch ausgerichtete Studien zur Unterrichtsqualität im Rahmen ihrer theoretischen Fundierungen wahrnehmen. Gleichzeitig könnten sich darin aber auch Parallelen zeigen zwischen dem Lehren und Lernen von Mathematik, einem Bereich, in dem zahlreiche Studien der Unterrichtsforschung durchgeführt wurden, und dem Lehren und Lernen in den naturwissenschaftlichen Fächern, so dass eine Verortung von Unterrichtsqualitätskriterien leichter möglich sein könnte, als es zwischen anderen Fächern der Fall ist. Schließlich stellt das Syntheseframework selbst eine Synthese verschiedener Ansätze dar und ist – zumindest auf der Ebene der Dimensionen und Subdimensionen – auf einer vergleichsweise allgemeinen Ebene beschrieben. Auf diesen Ebenen hilft die Verallgemeinerung, Verortungen und Verknüpfungen zwischen den naturwissenschaftsdidaktischen und den generischen Ansätzen herstellen zu können. Betrachtet man die Ebene der Subdimensionen zeigt sich, dass fachspezifische Theoretisierungen und Terminologien zunehmen. Auf der Ebene der Items schließlich, auf der die Kodierungen im Review basierten, herrschen zu großen Teilen fach- und teilweise themenspezifische Operationalisierungen vor. Hier bildet sich in hohem Maße die fachspezifische Perspektivierung und Konkretisierung der Qualitätskriterien ab.

Das erweiterte Syntheseframework lässt sich aus dieser Perspektive als Kommunikationstool zwischen den naturwissenschaftsdidaktischen Disziplinen, der fachübergreifend ausgerichteten allgemeinen Unterrichtsforschung und weiteren fachdidaktischen Disziplinen interpretieren. Mit der Möglichkeit, Merkmale aus unterschiedlichen Disziplinen auf Ebene der Subdimensionen des Syntheseframeworks zu verorten, kann disziplinenübergreifende Verständigung unterstützt werden. Im Falle unterschiedlicher Begrifflichkeiten und Operationalisierungen kann Kommunikation ermöglicht werden, indem die Bezüge zur Ebene der Dimensionen hergestellt werden. So lassen sich beispielsweise Parallelen zwischen dem Unterrichtsqualitätsmerkmal der „metakognitiv-epistemologische Ebene“ (Zülsdorf-Kersting, in diesem Heft) und dem, in diesem Beitrag verorteten, Kriterium der „Explizierung von Denkprozessen“ identifizieren, indem eine Verknüpfung über die Ebene der kognitiven Aktivierung hergestellt wird.

Schließlich zeigt die Perspektive der Fachspezifika auf die Unterrichtsforschung und das Syntheseframework Möglichkeiten zur theoriebasierten Erweiterung von Kriterien für Unterrichtsqualität auf. Es deutet sich gerade mit fachspezifischem Fokus an, dass die Reflexion der oben beschriebenen impliziten Bezüge zwischen den Diskursen über Unterrichtsqualität und naturwissenschaftsdidaktischer Forschung fruchtbar für eine weiterführende Theoriebildung sein könnten. Die Ergänzung der Fragen nach einer adäquaten didaktischen Reduktion unter Berücksichtigung zukünftiger fachlicher Lernschritte könnte darauf hindeuten, dass durchaus fachdidaktisch relevante Aspekte existieren, die bisher eher selten in der naturwissenschaftsdidaktischen Unterrichtsforschung berücksichtigt wurden.

6.2 Typisierung von Fachspezifik im Verhältnis zu generischen Unterrichtsqualitätskriterien

Aus der Zuordnung von Kriterien in das Syntheseframework und der Erarbeitung naturwissenschaftsdidaktischer Perspektivierungen wird eine Typisierung des Verhältnisses fachspezifischer und generischer Kriterien deutlich. Diese Typisierung ist hierbei nicht als Ersatz zur Einteilung in generische und fachspezifische Anteile zu verstehen, wie sie in Anlehnung an Praetorius und Charalambous (2018), im Methodenteil beschrieben wurde. Vielmehr ergab sich die Typisierung bei der Identifikation der fachspezifischen Anteile der naturwissenschaftsdidaktischen Perspektivierung und stellt somit eine Ergänzung zur Präzisierung fachspezifischer Aspekte dar. Sie umfasst drei Typen, die insbesondere beschreiben, wie Fachspezifik im Verhältnis zu generischen Kriterien ausgestaltet sein kann:

1. *Fachspezifik durch Fokussierung generisch formulierter Kriterien auf einen spezifischen Ausschnitt des Faches („Fokussierungs-Spezifik“)*: Fachspezifische Kriterien können durch eine Fokussierung von generischen Merkmalen auf Spezifika eines Faches etabliert werden, indem ein spezifischer Ausschnitt des Faches beurteilt wird und andere Bereiche des Unterrichts nicht im Fokus sind. Dabei kann das Verständnis des generischen Merkmals vergleichbar bleiben, jedoch auf ein spezifisches Fachmerkmal gerichtet sein. Beispiele hierfür sind die „kognitive Aktivierung beim Experimentieren“ oder „Klarheit und Strukturiertheit des Experiments“
2. *Fachspezifik durch Konkretisierung und Operationalisierung generisch formulierter Kriterien mittels fachdidaktischer Theorieelemente („Theoretisierungs-Spezifik“)*: Fachspezifische Kriterien können gebildet werden, indem generische Begrifflichkeiten genutzt, aber mittels fachdidaktischer Terminologien und Theorien operationalisiert werden. Ein Beispiel hierfür sind die Kriterien zur „Konstruktive Einbindung von eigenen Ideen und Schülervorstellungen in den Unterricht“ (Dimension II „Kognitive Aktivierung“), bei denen eine fachspezifische Theorie den Fokus vorgibt, die Übertragung auf den Unterricht jedoch mit Hilfe verschiedener generischer Aspekte erfolgt. Schülervorstellungen stellen in diesem Fall ein Beispiel dar, das in der naturwissenschaftsdidaktischen Forschung einen zentralen Platz einnimmt und durch zahlreiche naturwissenschaftsdidaktische Theorien beschrieben wird (für einen Überblick siehe z. B. von Aufschnaiter und Rogge 2015). Wenngleich für eine kognitive Aktivierung der Einbezug des aktuellen Lernstandes von Schülerinnen und Schülern in einer generischen Perspektive für eine Vielzahl von Fächern wichtig ist, ist die Nutzung von Schülervorstellungen ein Spezifikum, das stärker auf die naturwissenschaftlichen Fächer zutrifft, da hier ein Konzept genutzt wird, das zur Beschreibung von fachspezifischer Kognition beim naturwissenschaftlichen Lernen genutzt wird. Fachspezifik wird durch die Einbringung dieses Konzeptes in den Diskurs der Unterrichtsqualität generiert.
3. *Fachspezifik durch Ergänzung von Kriterien, die nicht fächerübergreifend sind („genuine Fachspezifik“)*: Im fachdidaktischen Diskurs werden Theorieelemente benannt, die in generisch orientierten Diskursen oder Frameworks nicht auftauchen. Beispiele hierfür sind die in diesem Beitrag genannten Ergänzungen, wie

die Berücksichtigung zukünftiger fachlicher Lernschritte bei der Auswahl und der Einbindung von Inhalten in den Unterricht.

6.3 Merkmale von Unterrichtsqualität in Abhängigkeit der Zieldimensionen von Unterricht

Aber nicht allein die Ableitung von weiteren Kriterien für Unterrichtsqualität kann von dieser Perspektive profitieren. Vielmehr erscheint die Frage nach Unterschieden in der Bedeutung von Qualitätskriterien für das Erreichen von verschiedenen Zielen, die Unterricht haben kann, als bisher kaum bearbeitet. Eine Systematisierung oder gar Modellierung solcher Abhängigkeiten zwischen Zieldimensionen, wie hier im Artikel beschrieben, und Qualitätsmerkmalen liegt für die naturwissenschaftlichen Fächer kaum vor – obwohl die Definition von Zielen einer Unterrichtsstunde und deren Konsequenzen für die Beurteilung der Qualität des Unterrichts durchaus Gegenstand der Lehramtsausbildung sind und für die unterschiedliche Relevanz drei Basisdimensionen für kognitive oder affektive Ziele beschrieben wurden (Klieme und Rakoczy 2008).

An dieser Stelle sei dieser Aspekt exemplarisch anhand der hier beschriebenen Zieldimensionen und Qualitätsmerkmalen konkretisiert: Für das Erreichen von Zielen aus der Zieldimension „knowing science“ (in der Sprache der Nationalen Bildungsstandards: Kompetenzbereich Fachwissen) kann eine hohe Explizierung von Vorgehensweisen von naturwissenschaftlichen Denk- und Arbeitsweisen als Teil einer kognitiven Aktivierung weniger zielführend sein, als für Ziele in der Zieldimension „doing science“ oder „knowing about science“ (in der Sprache der Nationalen Bildungsstandards: Kompetenzbereich Erkenntnisgewinnung). Wenn z. B. im letzteren Fall die verstehende Anwendung von Strategien zur Umsetzung von Experimenten erlernt werden soll, ist eine Explizierung von Vorgehensweisen insbesondere lernwirksam (Schwchow et al. 2016a), während sie im ersten Fall Unterrichtszeit binden kann, die eher für die Einbindung von Schülervorstellungen über einen Fachinhalt (kognitive Aktivierung) nützlich sein könnte. Ebenso könnte die „Gelungenheit des Experiments“ (Schulz 2011; verortet in der „Auswahl und Thematisierung von Inhalten und Fachmethoden“) für Ziele des „knowing science“ (Kompetenzbereich Fachwissen) insbesondere bedeutsam sein. Hier würde Unterrichtszeit genutzt, um ein Phänomen aus Natur oder Technik in den Unterricht zu integrieren und dann zu theoretisieren, um Fachwissen aufzubauen. Dass ein Experiment dann „klappt“, kann insbesondere für die im Rahmen der Dimension 1 „Auswahl und Thematisierung von Inhalten und Fachmethoden“ beschriebenen Perspektivierung „Strukturierter Ablauf und Sequenzierung der Stunde“ bedeutsam sein. Jedoch für Ziele im Bereich „knowing about science“ (Kompetenzbereich Erkenntnisgewinnung) könnte ein nicht gelungenes Experiment jedoch gerade ein wertvoller und kognitiv aktivierender Anlass sein, über das Wesen von Experimenten zu lernen und herauszuarbeiten, dass Experimente im engeren wissenschaftstheoretischen Sinne nicht „nicht gelingen“ können. So werden konfundierte, „nicht gelungene“ Experimente als Möglichkeit zur Förderung der Variablenkontrollstrategie beschrieben (Schwchow et al. 2016b). Anhand dieses Beispiels mit Bezügen zur Dimension 1

(„Auswahl und Thematisierung von Inhalten und Fachmethoden“) wird exemplarisch deutlich, dass das Syntheseframework einen Mehrwert gegenüber einer reinen Betrachtung von Unterrichtsqualität durch die Basisdimensionen bieten kann. Eine theoriebasierte Ableitung und Systematisierung solcher Zusammenhänge mit dem Ziel empirisch prüfbare Hypothesen abzuleiten, könnte Teil zukünftiger Aktivitäten der empirischen Unterrichtsforschung sein. Mit der Übersicht über Zieldimensionen und Qualitätsmerkmale könnte das vorliegende Paper auch dabei einen Beitrag aus naturwissenschaftsdidaktischer Perspektive leisten.

6.4 Limitationen der Untersuchung

Durch das beschriebene Vorgehen konnten die aufgestellten Forschungsfragen grundsätzlich beantwortet werden, dennoch ergeben sich einige Limitationen in Bezug auf das Ergebnis, die berücksichtigt werden müssen. Zunächst beschränkt sich der Abgleich zwischen dem Syntheseframework und den naturwissenschaftsdidaktischen Untersuchungen auf den deutschsprachigen Raum. Eine Erweiterung auf den internationalen Raum steht somit aus, sollte jedoch berücksichtigen, dass Unterschiede zwischen den deutschsprachigen und den internationalen Konzeptionen von Unterrichtsqualität auftreten können und dies im Forschungsansatz mitberücksichtigen. Weiterhin bleibt offen, ob das Syntheseframework den besten Ansatz zur Abbildung naturwissenschaftsdidaktischer Qualitätskriterien darstellt, auch wenn es sich für einen fächerübergreifenden Vergleich als gute Herangehensweise herausgestellt hat. An dieser Stelle würde auch die Frage anschließen, inwiefern das Syntheseframework den Ergänzungsbedarf der naturwissenschaftlichen Unterrichtsfächer, der innerhalb der drei Basisdimensionen besteht, vollständig decken kann. Zusätzlich sollte hierbei beachtet werden, dass die Naturwissenschaftsdidaktik in diesem Fall mehrere Fächer umfasst, die mitunter eine eigene Strukturierung von Unterrichtsqualität bevorzugen könnten. Hier steht die Unterrichtsqualitätsforschung insbesondere vor bisher kaum gelösten Herausforderungen, da auch innerhalb der Fächer unterschiedliche Inhaltsbereiche ausgemacht werden können (z. B. Fachgebiete der Chemie wie anorganische, organische, analytische oder physikalische Chemie, die sich im Schulcurriculum abbilden), in denen prinzipiell andere Dimensionen bedeutsam sein könnten. Abschließend bleibt anzumerken, dass der Vergleich der unterschiedlichen Systematisierungen von Unterrichtsqualität aufgezeigt hat, dass die Hierarchie bei der Darstellung von Unterrichtsqualität je nach Argumentation unterschiedlich ausgelegt werden kann. Das bedeutet für die Anwendung eines Frameworks zur Systematisierung von Unterrichtsqualität, dass die gewählten Ebenen und die Zuordnung von Aspekten zu diesen Ebenen durchaus diskutiert werden können und ein Ansatz zur strukturierten Darstellung von Unterrichtsqualität je nach Anwendungszweck mehr oder weniger hilfreich sein kann. Zielführend für die weitere Forschung wird die Offenlegung der konkreten Ziele der Verwendung eines Frameworks im Verhältnis zu seinen Zwecken sein.

Danksagung Die Autoren bedanken sich bei *Prof. Dr. Susanne Weßnigk* und *Dr. Sarah Dannemann* für ihre Anregungen zu den physik- und biologiedidaktischen Anteilen des Reviews sowie bei *Jasmin Meyer* für die Unterstützung während der Ratings und der Diskussion der Kriterien.

Funding Open Access funding provided by Projekt DEAL.

Open Access Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Artikel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.

Weitere Details zur Lizenz entnehmen Sie bitte der Lizenzinformation auf <http://creativecommons.org/licenses/by/4.0/deed.de>.

Literatur

- von Aufschnaiter, C., & Rogge, C. (2015). Conceptual change in learning. In R. Gunnstone (Hrsg.), *Encyclopedia of science education* (S. 209–218). Dordrecht, Heidelberg, New York, London: Springer.
- Barke, H.-D. (2006). *Chemiedidaktik. Diagnose und Korrektur von Schülervorstellungen*. Berlin: Springer.
- Beerenwinkel, A., & von Arx, M. (2016). Constructivism in practice: an exploratory study of teaching patterns and student motivation in physics classrooms in Finland, Germany and Switzerland. *Research in Science Education*. <https://doi.org/10.1007/s11165-015-9497-3>.
- Berliner, D. C. (2005). The near impossibility of testing for teacher quality. *Journal of Teacher Education*, 36(3), 205–213. <https://doi.org/10.1177/0022487105275904>.
- Bernholt, S., Parchmann, I., & Commons, M. L. (2009). Kompetenzmodellierung zwischen Forschung und Unterrichtspraxis. *Zeitschrift Für Didaktik Der Naturwissenschaften*, 15, 219–245.
- Börlin, J. (2012). Das Experiment als Lerngelegenheit: Vom interkulturellen Vergleich des Physikunterrichts zu Merkmalen seiner Qualität. In H. Niedderer, H. Fischler & E. Sumfleth (Hrsg.), *Studien zum Physik- und Chemielernen* Bd. 132. Berlin: Logos.
- Börlin, J., & Labudde, P. (2014). Practical work in physics instruction: an opportunity to learn? In H. E. Fischer, P. Labudde, K. Neumann & J. Viiri (Hrsg.), *Quality of instruction in physics. Comparing Finland, Switzerland and Germany* (S. 111–127). Münster: Waxmann.
- Brovelli, D. (2018). *Wirksamer Physikunterricht*. Baltmannsweiler: Schneider Verlag Hohengehren.
- Brunner, E. (2018). Qualität von Mathematikunterricht: Eine Frage der Perspektive. *Journal Für Mathematik-Didaktik*, 39(2), 257–284. <https://doi.org/10.1007/s13138-017-0122-z>.
- Bybee, R. W. (1997). *Achieving scientific literacy: from purposes to practices*. Portsmouth: Heinemann.
- Cauet, E. (2015). *Testen wir relevantes Wissen? Zusammenhänge zwischen dem Professionswissen von Physiklehrkräften und gutem und erfolgreichem Unterrichten*. Dissertation: Universität Duisburg-Essen.
- Chinn, C. A., & Malhotra, B. (2002). Epistemologically authentic inquiry in schools: a theoretical framework for evaluating inquiry tasks. *Science Education*, 86(2), 175–218. <https://doi.org/10.1002/sc.10001>.
- Dorfner, T., Förtsch, C., & Neuhaus, B. J. (2017). Die methodische und inhaltliche Ausrichtung quantitativer Videostudien zur Unterrichtsqualität im mathematisch-naturwissenschaftlichen Unterricht. *Zeitschrift für Didaktik der Naturwissenschaften*, 23(1), 261–285. <https://doi.org/10.1007/s40573-017-0058-3>.
- Duit, R., Gropengießer, H., Kattmann, U., Komorek, M., & Parchmann, I. (2012). The model of educational reconstruction—a framework for improving teaching and learning science. In D. Jorde & J. Dillon (Hrsg.), *Science education research and practice in Europe* (S. 13–37). Rotterdam: SensePublishers. https://doi.org/10.1007/978-94-6091-900-8_2.
- Fauth, B., Decristan, J., Rieser, S., Klieme, E., & Büttner, G. (2014). Student ratings of teaching quality in primary school: dimensions and prediction of student outcomes. *Learning and Instruction*, 29, 1–9.

- Förtsch, C., Heidenfelder, K., Spangler, M., & Neuhaus, B.J. (2018). How does the use of core ideas in biology lessons influence students' knowledge development? *Zeitschrift Für Didaktik Der Naturwissenschaften*, 24(1), 35–50. <https://doi.org/10.1007/s40573-018-0071-1>.
- Givvin, K.B., Hiebert, J., Jacobs, J.K., Hollingsworth, H., & Gallimore, R. (2005). Are there national patterns of teaching? Evidence from the TIMSS 1999 video study. *Comparative Education Review*, 49(3), 311–343. <https://doi.org/10.1086/430260>.
- Glemnitz, I.B. (2007). Vertikale Vernetzung im Chemieunterricht. Ein Vergleich von traditionellem Unterricht mit Unterricht nach Chemie im Kontext. In H. Niedderer, H. Fischler & E. Sumfleth (Hrsg.), *Studien zum Physik- und Chemielernen* Bd. 62. Berlin: Logos.
- Gräber, W. (1992). Interesse am Unterrichtsfach Chemie, an Inhalten und Tätigkeiten. *Chemie in Der Schule*, 39, 354–358.
- Gräber, W., Nentwig, P., Koballa, T., & Evans, R. (Hrsg.). (2002). *Scientific Literacy: Der Beitrag der Naturwissenschaften zur Allgemeinen Bildung*. Opladen: Leske + Budrich.
- Gropengießer, H. (2009). Beobachten. In H. Gropengießer, U. Harms & U. Kattmann (Hrsg.), *Fachdidaktik Biologie* (S. 273–277). Hallbergmoos: Aulis.
- Gropengießer, H., & Kattmann, U. (2013). Didaktische Rekonstruktion. In H. Gropengießer, U. Kattmann & U. Harms (Hrsg.), *Fachdidaktik Biologie* (S. 16–23). Hallbergmoos: Aulis.
- Gyllenpalm, J., & Wickman, P.-O. (2011). “Experiments” and the inquiry emphasis conflation in science teacher education. *Science Education*, 95(5), 908–926. <https://doi.org/10.1002/sce.20446>.
- Hammann, M. (2019). Organisationsebenen biologischer Systeme unterscheiden und vernetzen: Empirische Befunde und Empfehlungen für die Praxis. In J. Groß, M. Hammann, P. Schmiemann & J. Zabel (Hrsg.), *Biologiedidaktische Forschung: Erträge für die Praxis* (S. 77–91). Berlin, Heidelberg: Springer.
- Herweg, C. (2008). *Zielorientierung im deutschen und schweizerischen Physikunterricht – eine Videostudie*. Dissertation: Christian Albrechts Universität zu Kiel.
- Hodson, D. (2014). Learning science, learning about science, doing science: different goals demand different learning methods. *International Journal of Science Education*, 36(15), 2534–2553. <https://doi.org/10.1080/09500693.2014.899722>.
- Höft, L., Bernholt, S., Blankenburg, J.S., & Winberg, M. (2019). Knowing more about things you care less about: cross-sectional analysis of the opposing trend and interplay between conceptual understanding and interest in secondary school chemistry. *Journal of Research in Science Teaching*, 56(2), 184–210. <https://doi.org/10.1002/tea.21475>.
- Höttecke, D., & Rieß, F. (2015). Naturwissenschaftliches Experimentieren im Lichte der jüngeren Wissenschaftsforschung – Auf der Suche nach einem authentischen Experimentbegriff der Fachdidaktik. *Zeitschrift Für Didaktik Der Naturwissenschaften*. <https://doi.org/10.1007/s40573-015-0030-z>.
- Johnstone, A.H. (1991). Thinking about thinking. *International Newsletter on Chemical Education*, 36, 7–10.
- Kattmann, U. (2015). *Schüler besser verstehen. Alltagsvorstellungen im Biologieunterricht*. Hallbergmoos: Aulis.
- Kauertz, A., Fischer, H., Mayer, J., Sumfleth, E., & Walpuski, M. (2010). Standardbezogene Kompetenzmodellierung in den Naturwissenschaften der Sekundarstufe I. *Zeitschrift Für Didaktik Der Naturwissenschaften*, 16, 135–154.
- Kind, P.M. (2013). Conceptualizing the science curriculum: 40 years of developing assessment frameworks in three large-scale assessments. *Science Education*, 97(5), 671–694. <https://doi.org/10.1002/sce.21070>.
- Klein, P., Kuhn, J., & Müller, A. (2018). Förderung von Repräsentationskompetenz und Experimentbezug in den vorlesungsbegleitenden Übungen zur Experimentalphysik. *Zeitschrift Für Didaktik Der Naturwissenschaften*. <https://doi.org/10.1007/s40573-018-0070-2>.
- Klieme, E., & Rakoczy, K. (2008). Empirische Unterrichtsforschung und Fachdidaktik. Outcome-orientierte Messung und Prozessqualität des Unterrichts. *Zeitschrift für Pädagogik*, 54, 222–237.
- KMK (2005a). *Bildungsstandards im Fach Biologie für den Mittleren Schulabschluss*. München: Luchterhand.
- KMK (2005b). *Bildungsstandards im Fach Chemie für den Mittleren Schulabschluss*. München: Luchterhand.
- KMK (2005c). *Bildungsstandards im Fach Physik für den Mittleren Schulabschluss*. München: Luchterhand.
- Kunter, M., & Ewald, S. (2016). Bedingungen und Effekte von Unterricht: Aktuelle Forschungsperspektiven aus der pädagogischen Psychologie. In N. McElvany, W. Bos, H.G. Holtappels, M.M. Gebauer & F. Schwabe (Hrsg.), *Bedingungen und Effekte guten Unterrichts* (S. 9–31). Münster: Waxmann.

- Lau, A. (2011). *Passung und vertikale Vernetzung im Chemie- und Physikunterricht*. Berlin: Waxmann.
- Mikelskis-Seifert, S., & Fischler, H. (2003). Die Bedeutung des Denkens in Modellen bei der Entwicklung von Teilchenvorstellungen – Stand der Forschung und Entwurf einer Unterrichtskonzeption. *Zeitschrift Für Didaktik Der Naturwissenschaften*, 9, 75–88.
- Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and Meta-analyses: the PRISMA statement. *Annals of Internal Medicine*, 151(4), 264–270.
- Muckenfuss, H. (2001). Retten uns die Phänomene? Anmerkungen zum Verhältnis von Wahrnehmung und Theorie. *Naturwissenschaften im Unterricht. Physik*, 12(63/64), 74–77.
- Nehring, A., Stiller, J., Nowak, K. H., Upmeyer zu Belzen, A., & Tiemann, R. (2016). Naturwissenschaftliche Denk- und Arbeitsweisen im Chemieunterricht – eine modellbasierte Videostudie zu Lerngelegenheiten für den Kompetenzbereich der Erkenntnisgewinnung. *Zeitschrift Für Didaktik Der Naturwissenschaften*. <https://doi.org/10.1007/s40573-016-0043-2>.
- Osborne, J. (2007). Science education for the twenty first century. *Eurasia Journal of Mathematics, Science and Technology Education*, 3(3), 173–184. <https://doi.org/10.1080/13586840903194748>.
- Podschuweil, S., Bernholt, S., & Brückmann, M. (2016). Classroom learning and achievement: how the complexity of classroom interaction impacts students' learning. *Research in Science & Technological Education*, 34(2), 142–163. <https://doi.org/10.1080/02635143.2015.1092955>.
- Polte, S., & Wilde, M. (2018). Wirkt Ekel vor lebenden Tieren bei Schülerinnen und Schülern als Prädiktor für ihr Flow-Erleben? *Zeitschrift Für Didaktik Der Naturwissenschaften*, 24(1), 287–292. <https://doi.org/10.1007/s40573-018-0075-x>.
- 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., Pauli, C., Reusser, K., Rakoczy, K., & Klieme, E. (2014). One lesson is all you need? Stability of instructional quality across lessons. *Learning and Instruction*, 31, 2–12. <https://doi.org/10.1016/j.learninstruc.2013.12.002>.
- Praetorius, A., Herrmann, C., Gerlach, E., Zülsdorf-Kersting, M., Heinitz, B., Nehring, A. (eingereicht). Unterrichtsqualität in den Fachdidaktiken – zwischen generischen Dimensionen und fachspezifischer Ausdifferenzierung.
- Rehm, M. (2018). *Wirksamer Chemieunterricht*. Unterrichtsqualität: Perspektiven von Expertinnen und Experten, Bd. 2. Baltmannsweiler: Schneider Verlag Hohengehren.
- (2016). Eine Studie zwischen Kontinuität und Innovation. In K. Reiss, C. Sälzer, A. Schiepe-Tiska, E. Klieme & O. Köller (Hrsg.), *PISA 2015. Eine Studie zwischen Kontinuität und Innovation*. Münster: Waxmann.
- Risch, B., & Peifer, P. (2018). Didaktische Reduktion – Elementarisierung. In K. Sommer, J. Wambach-Laicher & P. Peifer (Hrsg.), *Konkrete Fachdidaktik Chemie* (S. 45–63). Seelze: Aulis.
- Schecker, H., Wilhelm, T., Hopf, M., & Duit, R. (2018). *Schülervorstellungen und Physikunterricht. Ein Lehrbuch für Studium, Referendariat und Unterrichtspraxis*. Berlin, Heidelberg: Springer Spektrum. <https://doi.org/10.1007/978-3-662-57270-2>.
- Schlesinger, L., & Jentsch, A. (2016). Theoretical and methodological challenges in measuring instructional quality in mathematics education using classroom observations. *ZDM – Mathematics Education*, 48(1/2), 29–40. <https://doi.org/10.1007/s11858-016-0765-0>.
- Schulz, A. (2011). Experimentierspezifische Qualitätsmerkmale im Chemieunterricht: Eine Videostudie. In H. Niedderer, H. Fischler & E. Sumfleth (Hrsg.), *Studien zum Physik- und Chemielernen* Bd. 113. Berlin: Logos.
- Schwichow, M., Croker, S., Zimmerman, C., Höffler, T., & Härtig, H. (2016a). Teaching the control-of-variables strategy: a meta-analysis. *Developmental Review*, 39, 37–63. <https://doi.org/10.1016/j.dr.2015.12.001>.
- Schwichow, M., Zimmerman, C., Croker, S., & Härtig, H. (2016b). What students learn from hands-on activities. *Journal of Research in Science Teaching*, 53(7), 980–1002.
- Seidel, T., Prenzel, M., Rimmle, R., Dalehefte, I. M., Herweg, C., Kobarg, M., & Schwindt, K. (2006a). Blicke auf den Physikunterricht. Ergebnisse der IPN Videostudie. *Zeitschrift für Pädagogik*, 52(6), 799–821.
- Sjöström, J., & Talanquer, V. (2014). Humanizing chemistry education: from simple contextualization to multifaceted problematization. *Journal of Chemical Education*, 91(8), 1125–1131. <https://doi.org/10.1021/ed5000718>.
- Sommer, K., & Pfeifer, P. (2018). Experiment und Erkenntnis. In K. Sommer, J. Wambach-Laicher & P. Pfeifer (Hrsg.), *Konkrete Fachdidaktik Chemie. Grundlagen für das Lernen und Lehren im Chemieunterricht* (S. 70–88). Hannover: Friedrich.

- Stigler, J. W., Gonzales, P., Kawanaka, T., Knoll, S., & Serrano, A. (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 DC: U.S. Government Printing Office.
- Sumfleth, E., & Fischer, H. E. (Hrsg.). (2013). *nwu-essen 10 Jahre Forschung zum naturwissenschaftlichen Unterricht*. Berlin: Logos.
- Szogs, M., Krüger, M., & Korneck, F. (2017). Erhebung von Unterrichtsqualität mittels hoch-inferenter Videorating – das Ratingmanual der Φ actio-Studie. In C. Maurer (Hrsg.), *Implementation fachdidaktischer Innovation im Spiegel von Forschung und Praxis. Gesellschaft für Didaktik der Chemie und Physik Jahrestagung in Zürich 2016* (S. 256–259).
- Talanquer, V. (2011). Macro, Submicro, and symbolic: the many faces of the chemistry “triplet.” *International Journal of Science Education*, 33(2), 179–195. <https://doi.org/10.1080/09500690903386435>.
- Talanquer, V. (2015). Central ideas in chemistry: an alternative perspective. *Journal of Chemical Education*, 93(1), 3–8. <https://doi.org/10.1021/acs.jchemed.5b00434>.
- Treagust, D., Chittleborough, G., & Mamiala, T. (2003). The role of submicroscopic and symbolic representations in chemical explanations. *International Journal of Science Education*, 25(11), 1353–1368. <https://doi.org/10.1080/0950069032000070306>.
- Unfallkasse NRW (2018). *Gemeinsames Lernen im Chemieunterricht der Sekundarstufe I*
- Vorholzer, A., von Aufschnaiter, C., & Boone, W. J. (2018). Fostering upper secondary students’ ability to engage in practices of scientific investigation: a comparative analysis of an explicit and an implicit instructional approach. *Research in Science Education*. <https://doi.org/10.1007/s11165-018-9691-1>.
- Wagenschein, M. (1976). *Die Pädagogische Dimension der Physik*. Braunschweig: Westermann.
- Wellnitz, N., & Mayer, J. (2013). Erkenntnismethoden in der Biologie – Entwicklung und Evaluation eines Kompetenzmodells. *Zeitschrift für Didaktik der Naturwissenschaften*, 19, 315–346.
- Widodo, A., & Duit, R. (2004). Konstruktivistische Sichtweisen vom Lehren und Lernen und die Praxis des Physikunterrichts. *Zeitschrift für Didaktik der Naturwissenschaften*, 10, 233–255.
- Wilhelm, M. (2018). *Wirksamer Biologieunterricht*. Baltmannsweiler: Schneider Verlag Hohengehren.
- Wirtz, M., & Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität: Methoden zur Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Categoriesystemen und Ratingskalen*. Göttingen: Hogrefe.
- Wüsten, S. (2010). *Allgemeine und fachspezifische Merkmale der Unterrichtsqualität im Fach Biologie. Eine Video- und Interventionsstudie*. Berlin: Logos.
- Ziegler, A., Antes, G., & König, I. (2011). Bevorzugte Report Items für systematische Übersichten und Meta-Analysen: Das PRISMA-Statement. *DMW – Deutsche Medizinische Wochenschrift*, 136(08):e9–e15.