

Unterrichtsqualität aus mathematikdidaktischer Sicht: Grundlegung, exemplarische Konkretisierung und empirische Überprüfung

Habilitationsschrift

Dr. phil. Esther Brunner

TUM School of Education
Technische Universität München

Januar 2020

Inhaltsverzeichnis

1	Einleitung	1
2	Mehrwert einer fachdidaktischen Sicht	6
2.1	Empirische Bildungsforschung und Primat der Erziehungswissenschaften	6
2.2	Aufsatz 1: Fachdidaktische Forschung als notwendige Perspektive	7
2.3	Aufsatz 2: Interdisziplinärer Beitrag zur Theorieentwicklung von Unterrichtsqualität	10
2.4	Diskussion	11
3	Unterrichtsqualität und der fachdidaktische Blick	14
3.1	Unterrichtsqualität: generisch oder fachspezifisch?	14
3.2	Studie 1: Qualitätssteigerung dank fachspezifischem Unterrichtscoaching	16
3.3	Studie 2: Qualität durch unterschiedliche Unterrichtsmethoden?	19
3.4	Studie 3: Hierarchisches Modell fachlicher Unterrichtsqualität	20
3.5	Diskussion	22
4	Qualitätsfragen am Beispiel mathematischen Begründens, Argumentierens und Beweisens	24
4.1	Studie 4: Beweistypen und Unterrichtsgespräch	25
4.2	Studie 5: Beweistypen und Merkmale von Lehrpersonen	26
4.3	Studie 6: Konzeptualisierung frühen mathematischen Argumentierens	27
4.4	Studie 7: Förderung frühen mathematischen Argumentierens	29
4.5	Diskussion	31
5	Qualitätsfragen am Beispiel des Mathematikunterrichts in unterschiedlichen Kontexten	35
5.1	Studie 8: Gestaltungselemente von Mathematikunterricht in Mehrjahrgangsklassen	37
5.2	Studie 9: Gestaltungsmuster von Mathematikunterricht im Kantonsvergleich	38
5.3	Studie 10: Überzeugungen von Lehrpersonen zu jahrgangsgemischtem Mathematikunterricht im Ländervergleich	39
5.4	Diskussion	42

6	Übergreifende Gesamtdiskussion	43
6.1	Kurze Zusammenfassung zentraler Ergebnisse	43
6.2	Epistemisches Potenzial eines interdisziplinären Ansatz	48
6.3	Implikationen für die Praxis	49
6.4	Weiterer Forschungsbedarf	50
	Literatur	52
	Anhang	

1 Einleitung

Was genau ist unter „gutem Mathematikunterricht“ zu verstehen und wie lässt sich dies bestimmen? Ist „guter Mathematikunterricht“ einfach „guter Unterricht“ oder gibt es eine fachspezifische oder möglicherweise gar eine fach- und inhaltspezifische Unterrichtsqualität? Und falls ja, welche Konsequenzen hätte dies für die Messung von Unterrichtsqualität? Für eine Klärung dieser Fragen liegen bislang keine fundierten theoretischen und empirischen Grundlagen vor. Dies erstaunt insbesondere vor dem Hintergrund der hohen Relevanz für die Praxis und die Forschung, die sich beide gleichermaßen intensiv mit der Frage der Unterrichtsqualität, ihrer Bestimmung, ihres Zustandekommens oder ihrer Auswirkungen auf das Lernen der Schülerinnen und Schüler bzw. auf die berufliche Zufriedenheit von Lehrpersonen befassen. Was genau unter „gutem“ Unterricht verstanden wird, wie diese Qualitätsbestimmung theoretisch konzeptualisiert, operationalisiert und empirisch erfasst wird und welche Zielsetzung der Qualitätsbestimmung insgesamt zugrunde liegt, wird zum einen längst nicht immer explizit ausgeführt und zeigt sich zum anderen in verschiedenen Arbeiten teils sehr different. Diese Unschärfe führt dazu, dass das Konstrukt „Unterrichtsqualität“ unterschiedlich aufgefasst wird und sich diesbezügliche Diskussionen in Praxis und Forschung nicht zwingend auf dieselben theoretischen Grundlagen, Operationalisierungen und anschließende Bestimmung und Erfassung beziehen, und zwar selbst dann nicht, wenn dieselben Begrifflichkeiten verwendet werden. Dies manifestiert sich beispielsweise in Konstrukten wie „Klarheit“. Dieses Konstrukt wird zum einen im Sinne von Klarheit der Handlungsanweisungen (Ophardt & Thiel, 2017) und damit als Aspekt der Klassenführung und zum anderen als „Strukturelle Klarheit“ (Drollinger-Vetter, 2011; Rakoczy, Klieme, Lipowsky & Drollinger-Vetter, 2010) und damit als Aspekt inhaltlich-fachlicher Kohärenz interpretiert. Solche unterschiedlichen Definitionen und anschließenden Operationalisierungen von Konstrukten lassen vermuten, dass die Bestimmung von Unterrichtsqualität von der zugrunde gelegten Analyseperspektive abhängt und dass deshalb die Einschätzung desselben Unterrichts durchaus unterschiedlich ausfallen und dieser in der Folge auch different bewertet werden kann. Obwohl diese Überlegung weitreichende Konsequenzen nach sich ziehen würde und sowohl für die Forschung als auch für die Praxis von hoher Bedeutung ist, wurde diesbezüglich noch kaum eine breite Diskussion lanciert.

Um Antworten auf mit Unterrichtsqualität zusammenhängende Fragen finden zu können, dürfte sich der Einbezug einer fachspezifischen Sicht als unabdingbar erweisen, denn Unterricht ist in der Regel immer auch Fachunterricht, der sich auf einen bestimmten Lerngegenstand bezieht, wie dies auch im didaktischen Dreieck ersichtlich ist, das schulische Vermittlungsprozesse als wechselseitige Interaktionen von Lehrperson, Lernenden und (fach)inhaltlichem Gegenstand (Ecken des Dreiecks) konzipiert (Abbildung 1) und drei unterschiedliche Bereiche (Seiten des Dreiecks) fokussiert: den Bereich der Lehrstoff- und Aufgabekultur, den Bereich der Lernprozess- und Interaktionskultur sowie den Bereich der Lernhilfe- und Unterstützungskultur. Reusser (2006, S. 161) bezeichnet diese drei Seiten bzw. diese drei Teilkulturen als „Basisdimensionen der Gestaltung konstruktivistischer Lehr-Lernarrangements“, wobei der Konstruktionsbegriff bei ihm im strukturgenetischen Ansatz von Piaget (2003) begründet ist und sich im Gegensatz zum „fertigen Stoff“ auf die Konstruktion der fachlichen Wissensinhalte bezieht. Der Fokus auf die „werdende Mathematik“ in Ergänzung zur „fertigen Mathematik“ wurde bereits von Polya (1949)

aufgegriffen und hat in der Mathematikdidaktik Tradition, die sich bei Freudenthal (1977) und seinem Ansatz des lokalen Ordners fortschrieb. Unterrichtsqualität entsteht innerhalb dieser wechselseitigen Beziehungen des didaktischen Dreiecks. Unterrichtsqualität im Fach Mathematik bedeutet demnach, diese wechselseitigen Beziehungen des didaktischen Dreiecks im Sinne der „werdenden Mathematik“ möglichst optimal zu nutzen, damit im Unterricht mathematisches Verstehen und dadurch ein Schritt in Richtung „fertiger Mathematik“ erreicht werden kann.

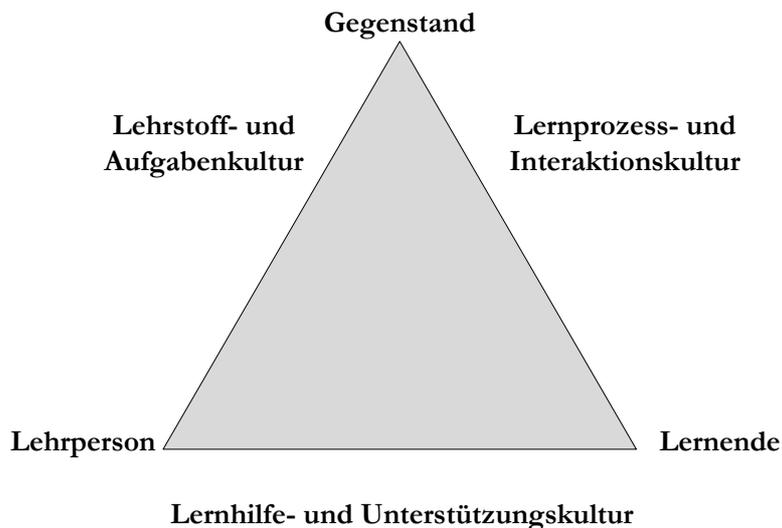


Abbildung 1: Didaktisches Dreieck (Reusser, 2006, S. 162)

Das Erfassen von Unterrichtsqualität in einem bestimmten Fach wie beispielsweise der Mathematik besteht somit darin, diese drei Bereiche oder Teilkulturen zu beleuchten und bezüglich ihrer Qualität einzuschätzen. Den drei Bereichen liegen allerdings unterschiedliche theoretische Ansätze zugrunde (Abbildung 2). Der Lehrstoff- und Aufgabenkultur, die sich auf die Interaktion zwischen Lehrperson und Gegenstand bezieht, liegt die Sicht der Disziplin Mathematik zugrunde, die sich mit dem Inhalt und dem Gegenstand befasst und in einer stoffdidaktischen Auseinandersetzung zentrale Wissens Elemente, ihre Voraussetzungen und ihre inhärenten Zusammenhänge offenlegt. Die Untersuchung der Lernprozess- und Interaktionskultur und damit einhergehend der Beziehung zwischen Lernenden und Gegenstand erfordert hingegen verstärkt nicht nur eine fachspezifische, sondern auch eine lerntheoretische Perspektive, indem der Blick auf das Verstehen des Gegenstands gerichtet wird, und zwar im Hinblick auf eine bestimmte zu erwerbende mathematische Kompetenz und deren Entwicklung. Die dritte Teilkultur, der Fokus auf die Lernhilfe- und Unterstützungskultur, verlangt sodann eine stärker didaktische Sichtweise und bezieht sich auf die fachspezifische Gestaltung des Lehr- und Lern-Prozesses.

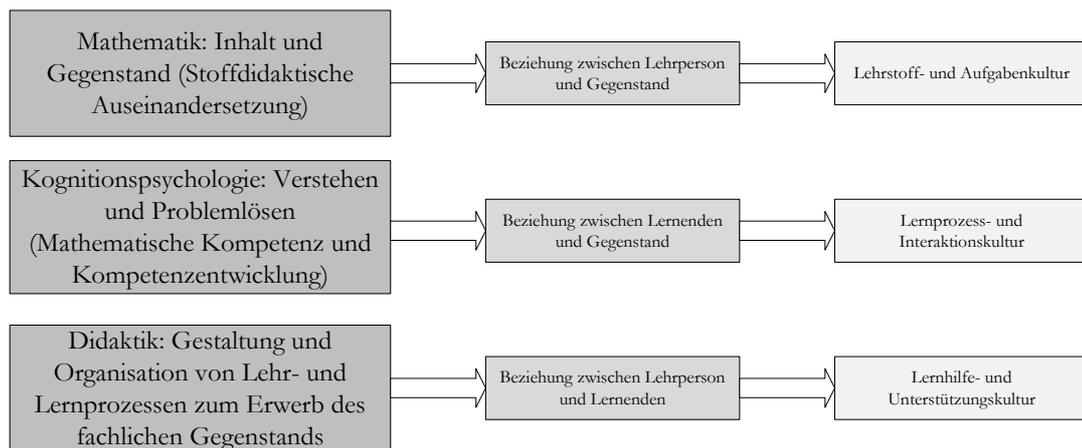


Abbildung 2: Von der Perspektive zur Teilkultur von Unterricht (in Anlehnung an Brunner, 2013, S. 23)

In diesen drei Perspektiven lassen sich auch die Konzeptionen verschiedener Bildungsstandards (Common Core State Standards Initiative, 2012; D-EDK, 2016; KMK, 2005) wiederfinden: Die erste Sichtweise bezieht sich auf den Inhalt und den Inhaltsbereich bzw. die inhaltlichen Leitideen, während sich die zweite Perspektive in erster Linie mit der Kompetenz bzw. dem Handlungsaspekt innerhalb dieser Inhalte befasst. Die dritte Perspektive schließlich, die sich der Gestaltung dieser Handlungen in Verbindung mit einem bestimmten Inhalt befasst, zielt auf die Anforderungen, die Bildungsziele und deren Modellierung im Unterricht ab.

Eine solche fachspezifische Sichtweise fehlt in der Bestimmung von Unterrichtsqualität im deutschsprachigen Kontext bislang weitgehend, während sie sich im amerikanischen Sprachraum zwar nicht in theoretischer Hinsicht, aber doch teilweise in verschiedenen Instrumenten zur Erfassung von Unterrichtsqualität etabliert hat. Dies zeigt sich auch in der großen Vielfalt zurzeit vorliegender Analyseinstrumente (zusammenfassend Schlesinger & Jentsch, 2016) zur Erfassung von Unterrichtsqualität, die von Charalambous und Paetorius (2018, S. 356) in „generische“, „fachspezifische“ und „hybride“ Instrumente eingeteilt werden.¹ Generische Instrumente erfassen Unterrichtsqualität aus einer allgemeinen, fachunabhängigen Sicht; fachspezifische Instrumente hingegen sind mehrheitlich auf das Unterrichtsfach bezogen und hybride Instrumente nehmen eine Verbindung der anderen beiden Kategorien von Instrumenten vor. Ungeklärt sind allerdings das Verhältnis dieser unterschiedlichen Kategorien von Instrumenten, ihr Einsatz und die folgenreiche Frage, ob die verschiedenen Instrumente zu derselben Einschätzung von Unterrichtsqualität führen oder ob mit abweichenden Einschätzungen zu rechnen ist. Im Kern geht es dabei um die Bedeutung der Domäne, innerhalb derer Unterrichtsqualität untersucht wird. Es stellt sich somit die grundsätzliche Frage nach Domänenspezifität von Unterrichtsqualität: Inwiefern gibt es eine „allgemeine“ Unterrichtsqualität und inwiefern ist dieses Konstrukt auf alle Fä-

¹ Als Beispiel eines generischen Ansatzes zur Bestimmung von Unterrichtsqualität nennen Charalambous und Paetorius (2018) das Instrument von Muijs et al. (2014) und das Instrument CLASS (Pianta & Hamre, 2009). Auch die im deutschen Sprachraum verbreiteten Basisdimensionen (Klieme, Lipowsky, Rakoczy & Ratzka, 2006) gehören hierzu. Als Beispiel eines klar fachspezifischen Ansatzes nennen Charalambous und Paetorius das Instrument MQI (Learning Mathematics for Teaching Project, 2011) und das Instrument MScan (Walkowiak, Berry, Meyer, Rimm-Kaufman & Ottmar, 2014), während das Instrument TRU (Schoenfeld, 2013; Schoenfeld, Floden & The Algebra Teaching and Mathematics Assessment Project, 2014) eine Verbindung von generischen Ansätzen mit fachspezifischen vornimmt und deshalb als „hybrides“ Instrument bezeichnet werden kann.

cher übertragbar? Was ist unter „hybrider“ Unterrichtsqualität zu verstehen, wie müsste ein entsprechendes Analyseinstrument genau beschaffen sein und welchen Mehrwert bietet ein solches Instrument beispielsweise gegenüber einem generischen Instrument?

Aber nicht nur eine mögliche Domänenspezifität von Unterrichtsqualität wird vergleichsweise selten diskutiert. Noch seltener wird die Frage nach einer möglichen Inhaltsspezifität von Unterrichtsqualität aufgeworfen. Ist von einer umfassenden, allgemeinen – das heißt in diesem Fall fachspezifischen, aber inhaltsunspezifischen – Unterrichtsqualität beispielsweise für Mathematik auszugehen? Oder ist diese entlang verschiedener zu bearbeitender Inhalte unterschiedlich zu konzipieren? Weist die Qualität von Beweis- und Begründungssequenzen andere Merkmale auf als die Qualität einer Übungslektion zu einem bestimmten mathematischen Verfahren? Auch diese Fragen wurden bislang weder in der Literatur noch in empirischen Studien berücksichtigt, obwohl sie bereits im didaktischen Dreieck angelegt wären, in dem die Lehrstoff- und Aufgabekultur eine konstitutive Funktion einnimmt. Da Lehrstoff nicht gleich Lehrstoff ist und die Aufgabekultur innerhalb desselben Fachs oder sogar Themenbereichs deutlich differieren kann, liegt der Schluss nahe, dass sich Unterrichtsqualität auch in Abhängigkeit vom bearbeiteten Inhalt unterschiedlich oder zumindest unterschiedlich akzentuiert bestimmen lassen müsste und somit eine inhaltspezifische Komponente aufweisen dürfte.

Unterricht kann aber nicht nur aus eher fachspezifischer oder aus generischer Sicht analysiert werden, sondern grundsätzlich auch durch unterschiedliche „Linsen“ (vgl. Givvin et al., 2005) betrachtet werden, beispielsweise mit Blick auf das Ganze aus einer Weitwinkelperspektive oder mit einem speziellen Fokus auf Details durch eine Art „Makroobjektiv“. Je nach gewählter Linse rücken andere Qualitätsmerkmale ins Zentrum. Dies wird deutlich, wenn ein weiteres terminologisches Gegensatzpaar der Erziehungswissenschaften herangezogen wird, nämlich dasjenige der Sicht- und der Tiefenstruktur (Reusser, 2005) von Unterricht. Eine Untersuchung von Tiefenstrukturmerkmalen zielt auf das (inhaltliche) Verstehen ab, während sich eine Analyse der Sichtstruktur mit dem Möglichkeitsraum (Kunter & Ewald, 2016) befasst, innerhalb dessen sich das Lernen der Schülerinnen und Schüler vollzieht. Die Sichtstruktur, die der Beobachtung unmittelbar zugänglich ist, gibt dabei den Rahmen vor, innerhalb dessen sich das der Beobachtung nicht unmittelbar zugängliche (fachliche) Lernen abspielt. Auch bezüglich dieses Gegensatzpaares dürfte davon auszugehen sein, dass derselbe Unterricht mit Blick auf Tiefenstrukturmerkmale nicht zwingend gleich eingeschätzt wird, wie wenn der Fokus auf den Sichtstrukturmerkmalen liegt.

Eine weitere offene und bislang kaum thematisierte Frage betrifft eine mögliche Hierarchisierung verschiedener Qualitätsaspekte. In den meisten Konzeptionen und den darauf aufbauenden Analyseinstrumenten zur Erfassung von Unterrichtsqualität werden die Qualitätsmerkmale auf derselben Hierarchieebene angesiedelt, obwohl sich spätestens seit der Ausdifferenzierung von *Basisdimensionen* von Unterrichtsqualität (Klieme et al., 2006) sowohl aus Praxissicht als auch aus der Sicht der Forschung zum einen die Frage nach einer möglichen Hierarchisierung von Qualitätsmerkmalen und zum anderen die Frage nach Zusammenhängen zwischen einzelnen Qualitätsmerkmalen oder Qualitätsmerkmalen verschiedener Ebenen bzw. ihrer gegenseitigen Bedingtheit stellen sollten. Berücksichtigt werden muss zudem, dass Unterricht und damit auch Unterrichtsqualität innerhalb eines bestimmten kulturell geprägten normativen Rahmens realisiert wird. Soziomathematische Normen (Yackel & Cobb, 1996) prägen nicht nur den Unterricht, sondern ebenso sehr die Bestimmung von Unterrichtsqualität. Die Einschätzung von Unter-

richtsqualität ist deshalb stets auch davon abhängig, was eine bestimmte Kultur als „guten“ Unterricht betrachtet (z. B. Givvin et al., 2005; Kaur, Anthony & Ohtani, 2013; Reusser & Pauli, 2003; Shimizu, Kaur, Huang & Clarke, 2010). Vor diesem Hintergrund ist davon auszugehen, dass die Qualität von Mathematikunterricht weder für unterschiedliche kulturelle Kontexte noch für unterschiedliche Schulstufen und möglicherweise auch nicht für unterschiedliche Organisationsstrukturen einheitlich definiert und beschrieben werden kann.

Zusammenfassend kann festgehalten werden, dass das Konstrukt „Unterrichtsqualität“ bezüglich verschiedener zentraler Fragen zurzeit weder zufriedenstellend theoretisch konzipiert noch empirisch operationalisiert vorliegt, sondern verschiedene – lokale – Ansätze und Instrumente größtenteils unverbunden nebeneinander bestehen und es mehr oder weniger zufällig erscheint, welches Instrument wofür eingesetzt wird. Dies führt zwangsläufig zu unterschiedlichen Ergebnissen der Qualitätseinschätzung eines bestimmten Unterrichts und verunmöglicht deshalb eine systematische Untersuchung von Unterrichtsqualität. Die gegenwärtig zentralen offenen Fragen beziehen sich auf folgende Dimensionen: 1) Fachspezifität vs. generische Qualität, 2) Fachspezifität vs. Inhaltsspezifität, 3) Tiefen- vs. Sichtstruktur von Unterricht, 4) Hierarchisierung von Qualitätsmerkmalen, 5) Wechselwirkungen von verschiedenen Qualitätsmerkmalen untereinander und 6) Einfluss des kulturellen Hintergrundes und der systemisch bedingten Rahmenbedingungen auf die Unterrichtsqualität. Diese sechs Dimensionen bilden den inhaltlichen Rahmen der vorliegenden Habilitationsschrift.

Ausgehend von zwei Aufsätzen soll zunächst die epistemische Basisfrage geklärt werden, welcher Mehrwert durch den Einbezug einer fachdidaktischen Perspektive grundsätzlich zu erwarten ist (vgl. Kapitel 2). Danach geht es mit einem engeren thematischen Fokus um den Beitrag der Fachdidaktik zur Bestimmung der Qualität von Fachunterricht (vgl. Kapitel 3). In diesem Zusammenhang werden drei unterschiedliche Studien vorgestellt, die sich alle mit der grundsätzlichen Frage der Konzeptualisierung fachlicher Qualitätsmerkmale und deren empirischer Bestimmung befassen. Anhand von zwei thematischen Schwerpunkten werden sodann verschiedene Qualitätsaspekte von Mathematikunterricht dargestellt, zum einen mit Blick auf eine prozessbezogene Dimension (vgl. Kapitel 4) im Sinne der „werdenden Mathematik“ und zum anderen bezüglich kontextbezogener Dimensionen (vgl. Kapitel 5). Als prozessbezogene Dimension wird das für die Disziplin (Heintz, 2000) und den Unterricht unverzichtbare mathematische Argumentieren, Begründen und Beweisen (Common Core State Standards Initiative, 2012; D-EDK, 2016; KMK, 2005; NCTM, 2000) für verschiedene Alters- und Schulstufen gewählt. Im Fokus der ersten beiden Studien steht die Sekundarstufe I (vgl. Kapitel 4.1–4.2.) während sich die nächsten beiden Studien (vgl. Kapitel 4.3–4.4) mit dem frühen mathematischen Argumentieren im Kindergarten befassen. Die ebenfalls thematisierte kontextbezogene Dimension betrifft Qualitätsfragen von Mathematikunterricht unter Berücksichtigung einer bestimmten Schul- bzw. Organisationsstruktur (vgl. Kapitel 5.1–5.3). Das letzte Kapitel (vgl. Kapitel 0) dient dem Zweck einer übergreifenden Diskussion und zieht ein Fazit zu den sechs oben aufgeführten Dimensionen, in denen zurzeit offene Fragen bestehen. Des Weiteren werden Implikationen für die Praxis abgeleitet und weitere offene Forschungsfragen benannt.

2 Mehrwert einer fachdidaktischen Sicht

2.1 Empirische Bildungsforschung und Primat der Erziehungswissenschaften

Braucht es eine eigenständige, forschende Fachdidaktik oder ist wissenschaftsbasierte, forschende Fachdidaktik ein Teil empirischer Lehr-Lern-Forschung, wie dies verschiedene Autorinnen und Autoren nahelegen? Oder sind empirische Bildungsforschung und empirisch forschende Fachdidaktik genauso Gegensatzpaare und „fremde Schwestern“, wie dies Terhart (2002, S. 77) für die Allgemeine Didaktik als normativ geprägte, historisch aus der Praxis heraus entstandene Wissenschaft einerseits und die empirisch arbeitende Lehr-Lern-Forschung bzw. Bildungsforschung andererseits beschreibt? In diesem Grundkonflikt einer zunehmenden Professionalisierung der Fachdidaktiken (Heitzmann & Pauli, 2015), die sich vermehrt auch im Feld der empirischen Bildungsforschung und nicht nur in der Praxis der Lehrerinnen- und Lehrerbildung etablieren, gerät das bisher vorherrschende Primat der Erziehungswissenschaften zunehmend unter Druck, wenn es um die Erforschung von Lehr-Lern-Prozessen geht. Fachdidaktik und damit fachdidaktische Forschung nehmen für sich in Anspruch, die fachlichen Lernprozesse, um deren Erforschung es im Zusammenhang mit Unterrichtsforschung in der Regel geht, fachspezifisch modellieren, operationalisieren und analysieren zu können und damit über allgemeine und generische Fragestellungen hinaus einen Beitrag zur Erforschung fachlichen Lernens leisten zu können. Reiss und Ufer (2009, S. 199) sprechen am Beispiel der mathematikdidaktischen Forschung präzisierend von „mathematikdidaktischer Bildungsforschung“ und machen dadurch zum einen deutlich, dass das Konzept einer empirisch arbeitenden Fachdidaktik zu kurz greife, weil es *die* Fachdidaktik nicht gebe, sondern immer nur die jeweilige Fachdidaktik eines bestimmten Schulfaches wie beispielsweise die Mathematikdidaktik, die sich mit der Erforschung von Mathematikunterricht befasst. Zum anderen wird im Begriff der Autorin und des Autors die Nähe zur empirischen Bildungsforschung im Hinblick auf Arbeitsweisen und Methoden erkennbar. Mathematikdidaktische Forschung arbeitet somit mit etablierten Methoden der Bildungsforschung, adaptiert diese aber für fachspezifische Fragestellungen und Analysen.

Das Problemlösen beispielsweise wurde traditionell während vieler Jahre aus der Perspektive der Erziehungswissenschaften und der Psychologie erforscht (z. B. Dörner, 1974; Duncker, 1935; Funke, 2003; Reusser, 1984) und oftmals an Beispielen aus der Mathematik konkretisiert (Aebli, 1981; Wertheimer, 1964), weshalb es lange Zeit an einer spezifisch fachlichen Sicht auf die zugrunde liegenden mathematischen Probleme im Sinne mathematischen Problemlösens fehlte. Ein solcher fokussierter Zugang zur Beschreibung und Erforschung mathematischen Problemlösens (z. B. Polya, 1995; Schoenfeld, 1985, 1994) wie auch die vertiefte mathematikdidaktische Erforschung von Problemlösen im Zusammenhang mit Modellierungsaufgaben (z. B. Schukajlow et al., 2012; Schukajlow & Leiss, 2011) oder entlang von heuristischen Strategien (z. B. Rott, 2014; Zöttl, Ufer & Reiss, 2010) zeugen mittlerweile von einer zunehmenden Professionalisierung der Fachdidaktiken, die sich als empirisch forschende und eigenständige Disziplinen begreifen (Reusser, 1991). Durch ihren spezifischen fachlichen Blick stellen diese Arbeiten eine Erweiterung der Erkenntnisse der breiten Forschung zum Problemlösen dar und ergänzen die psychologisch orientierte Forschung zu Textaufgaben (z. B. Reusser, 1989; Verschaffel et al., 1999; Verschaffel, Dooren, Greer & Mukhopadhyay, 2010).

Das Beispiel des Problemlösens zeigt, dass es nicht um eine Vorherrschaft oder um ein Entweder-oder geht, sondern vielmehr um eine gewinnbringende, erkenntniserweiternde Ergänzung empirischer Bildungsforschung durch eine fachspezifische Perspektive. Denn die in Kapitel 1 aufgeführten sechs Dimensionen, in denen sich im Zusammenhang mit der Bestimmung und der Erforschung von Unterrichtsqualität zentrale offene Fragen stellen, lassen sich kaum aus einer Perspektive heraus umfassend erforschen. Die Klärung der Frage nach Fachspezifität vs. generischer Qualität (1) verlangt den Einbezug allgemein ausgerichteter Bildungsforschung wie auch fachdidaktischer Forschung. Gleiches gilt für die Auseinandersetzung mit der Frage zur Tiefen- vs. Sichtstruktur von Unterricht (3) oder zur Hierarchisierung von Qualitätsmerkmalen (4). Die Befassung mit der Frage zu Fachspezifität vs. Inhaltsspezifität (2) hingegen dürfte eher in den Bereich der fachdidaktischen Forschung fallen, während Wechselwirkungen von verschiedenen Qualitätsmerkmalen untereinander (5) vermutlich vor allem aus der Perspektive der empirischen Bildungsforschung unter Einbezug der entsprechenden Fachdidaktik zu untersuchen wären.

2.2 Aufsatz 1: Fachdidaktische Forschung als notwendige Perspektive

Brunner, E. (2015). Mathematikdidaktische Forschung. Eine notwendige vertiefende Perspektive. *Beiträge zur Lehrerinnen- und Lehrerbildung BzL*, 33(2), 235-245.

Für die Schweiz stellen die Entwicklung und die Etablierung fachdidaktischer Forschung eine besondere Herausforderung dar, weil Fachdidaktiken größtenteils nicht an universitären Lehrstühlen angesiedelt sind, sondern mehrheitlich oder beinahe ausschließlich in der Lehrerinnen- und Lehrerbildung an den Pädagogischen Hochschulen verortet sind. Da die Pädagogischen Hochschulen kein Promotionsrecht haben, besteht zum einen ein großes Nachwuchsproblem hinsichtlich qualifizierter Personen im Bereich der Fachdidaktik. Zum anderen werden Fachdidaktiken infolge ebendieses Mangels an qualifizierten Mitarbeitenden nicht an allen Pädagogischen Hochschulen als forschende Disziplinen verankert. Diese spezielle Situation wird im Aufsatz dargestellt. An einem Beispiel aus einer Pädagogischen Hochschule, die mathematikdidaktische Forschung etablieren wollte, wird aufgezeigt, worin die Aufgaben fachdidaktischer Forschung bestehen, welche Herausforderungen sich zeigen und welcher Mehrwert von diesem Zugang für die Forschung und die Praxis zu erwarten ist. Konkretisiert wird dies an einem zum Zeitpunkt der Publikation laufenden Projekt zur Erforschung von Mathematikunterricht in altersdurchmischten Klassen der Primarschule (vgl. Kapitel 5.1).

Der Mehrwert fachdidaktischer Forschung wird im Aufsatz zum einen anhand ihrer unterschiedlichen Aufgabenfelder dargestellt und zum anderen durch den Verweis auf pädagogisch relevantes Handlungswissen, das auf der Basis der generierten wissenschaftlichen Erkenntnisse entwickelt wird. Diskutiert wird, inwiefern bzw. weshalb mathematikdidaktische Forschung nicht nur eine notwendige vertiefende Perspektive darstellt, sondern auch eine Sonderrolle einnimmt. Diese Sonderrolle wird mit der Nähe zum schulischen Fachunterricht und dem Bezug zur fachlichen Domäne und zur wissenschaftlichen Disziplin begründet. Im Zentrum fachdidaktischer Forschung stehen der fachliche Untersuchungsgegenstand und seine Vermittlung sowie der fachliche Lernprozess. Dies erfordert von Forschenden profunde Kenntnisse des jeweiligen Unterrichtsfachs und der Inhalte.

In Anlehnung an Reiss und Ufer (2009) werden im Aufsatz die vier Aufgabenfelder von Fachdidaktik und fachdidaktischer Forschung umrissen, konkretisiert und durch ein weiteres, fünftes Feld ergänzt: 1) Aufgabenfeld „Schulfach und seine Inhalte und Ziele“, 2) Aufgabenfeld „Fachliche Lehr-Lern-Prozesse“, 3) Aufgabenfeld „Unterrichtspraxis mit Lehrmaterialien und Lernumgebungen“, 4) Aufgabenfeld „Gesellschaftliche Anforderungen“, das beispielsweise Standards und Tests umfasst, und 5) Aufgabenfeld „Inter- und Transdisziplinarität“ (Defila, Di Giulio & Scheuermann, 2006). Zu diesen fünf Aufgabenfeldern fachdidaktischer Forschung werden im Aufsatz jeweils Beispiele aus der mathematikdidaktischen Forschung gegeben.

Der Aufsatz fragt sodann nach der Qualität von fachdidaktischer Forschung und verlangt mit Reiss und Ufer (2009, S. 210), dass fachdidaktische Forschung anschlussfähig an Methoden und Ergebnisse aus den jeweiligen Bezugsdisziplinen zu sein habe, wobei unter Anschlussfähigkeit nicht ein unhinterfragtes, blindes Übernehmen von Methoden und theoretischen Grundlagen einer Bezugsdisziplin verstanden wird. Vielmehr soll in Kenntnis dieser Grundlagen und Methoden geprüft werden, inwiefern sich diese gewinnbringend für die eigene fachdidaktisch angelegte Forschung nutzen, ergänzen oder spezifizieren lassen. Dies wird am Beispiel mathematischen Argumentierens konkretisiert, indem verdeutlicht wird, dass bestimmte methodische Zugänge, beispielsweise Vorgehensweisen und theoretische Grundlagen aus der Argumentationstheorie (Van Eemeren & Grootendorst, 2004), eine auf etablierten Termini beruhende Beschreibung von Argumentstrukturen (Toulmin, 1996) oder Ansätze aus der Gesprächsforschung (Henne & Rehbock, 2001), notwendigerweise beizuziehen sind, weil sich mathematisches Argumentieren in einem Diskurs ereignet.

Diskussion

Mit Blick auf Unterrichtsqualität stellt der Beitrag zur postulierten Notwendigkeit der Perspektive fachdidaktischer Forschung eine Grundlage für weiterführende Überlegungen dar. Denn die im Aufsatz thematisierten unterschiedlichen Aufgabenfelder fachdidaktischer Forschung können unter anderem auch im Kontext von Fragen der Unterrichtsqualität im Hinblick auf ihren Mehrwert diskutiert werden. Insbesondere bei der Bestimmung und der Analyse der Qualität von Fachunterricht ist der Einbezug einer fachspezifischen Perspektive unabdingbar und stellt einen Mehrwert gegenüber ausschließlich erziehungswissenschaftlicher Bildungsforschung dar. Dies soll nachfolgend anhand der fünf Aufgabenfelder kurz aufgezeigt werden.

1) *Aufgabenfeld „Schulfach und seine Inhalte und Ziele“*: Im Zusammenhang mit fachlicher Unterrichtsqualität ist es bedeutsam, zu fragen, inwiefern der für den Unterricht ausgewählte inhaltliche Gegenstand von Relevanz für das weitere fachliche Lernen der Schülerinnen und Schüler und für deren Alltagsbewältigung ist und inwiefern dem Gegenstand darüber hinaus eine exemplarische Bedeutung beigemessen werden kann (Klafki, 1991). Dies zu klären, erfordert gemäß der im Aufsatz entwickelten Argumentation eine fachspezifische Sicht und eine fundierte fachdidaktische Perspektive.

2) *Aufgabenfeld „Fachliche Lehr-Lern-Prozesse“*: Wenn Unterrichtsqualität im Sinne der Output-Variablen „Leistung“ bzw. „Lernzuwachs“ definiert wird, stellt sich die Frage, wie sich fachliches Verstehen eines bestimmten fachlichen Inhalts bzw. Konzepts entwickelt oder welche zentralen, fachlich bedingten Verstehenshürden dabei auftreten können. Auch diese Frage kann nur aus einer auf einer vertieften Kenntnis des Fachs beruhenden inhaltlichen und fachspezifischen Sicht heraus fundiert beantwortet werden.

3) *Aufgabenfeld „Unterrichtspraxis mit Lehrmaterialien und Lernumgebungen“*: Der Rahmen für das Entstehen von Unterrichtsqualität konstituiert sich auf der Ebene der Sichtstruktur von Unterricht, welcher den Möglichkeitsraum für die Tiefenstruktur (Charalambous & Praetorius, 2018) und damit für das fachliche Lernen darstellt. Es ist deshalb aus fachlicher Sicht zentral, im Zusammenhang mit Unterrichtsqualität auch die fachliche Qualität der eingesetzten Lehrmaterialien und der Lernumgebung einzuschätzen und diese Elemente mit den zu erwerbenden fachlichen Kompetenzen und Bildungszielen des Unterrichts in Verbindung zu bringen. Wenn die Lehrmaterialien oder die Lernumgebung eine zu geringe Qualität aufweisen, dann dürfte davon auszugehen sein, dass sich dies negativ auf das fachliche Lernen der Schülerinnen und Schüler auswirkt, unabhängig davon, ob der Unterricht bezüglich generischer Qualitätsmerkmale als gut eingeschätzt wird.

4) *Aufgabenfeld „Gesellschaftliche Anforderungen“* (beispielsweise Standards und Tests): Auch das vierte Aufgabenfeld ist im Zusammenhang mit dem Einschätzen der Qualität von Fachunterricht bedeutsam und erfordert ebenfalls den Einbezug einer fachspezifischen Sichtweise. Gesellschaftliche Anforderungen einerseits und normativen Grundlagen der Disziplin andererseits müssen in Einklang gebracht werden. Im Hinblick auf Unterrichtsqualität bedeutet dies, zu fragen, inwiefern der Unterricht dazu geeignet ist, Ansprüche des Fachs mit denjenigen der Gesellschaft zu verbinden, beispielsweise durch eine Orientierung an Anwendbarkeit von erworbenem Wissen oder durch Realitätsbezüge. Gleichzeitig ist es aber auch notwendig, aus Fachsicht heraus nicht alles der gesellschaftlichen Forderung nach Anwendbarkeit, wie dies das Literacy-Konzept (OECD, 2013) nahelegt, unterzuordnen. Gerade aus einem vertieften Fachverständnis heraus zeigt sich unter anderem auch die Notwendigkeit des Verstehens und Arbeitens mit reiner Mathematik, das heißt der Beschäftigung mit dem innermathematischen Bereich. Gerade dies ist unabdingbar, wenn es beispielsweise um – für die Disziplin konstitutive – Bereiche wie mathematisches Argumentieren und Beweisen (vgl. Kapitel 4) geht. Denn das Einnehmen einer ausschließlich erziehungswissenschaftlichen Perspektive könnte zur Annahme führen, dass es nicht notwendig sei, im schulischen Mathematikunterricht beweisen zu lernen, was aus der Sicht der Disziplin keineswegs zutrifft.

5) *Aufgabenfeld „Inter- und Transdisziplinarität“*: Auch dieses Aufgabenfeld wird bei der Erforschung von fachlicher Unterrichtsqualität notwendigerweise angesprochen. Wenn generische und fachspezifische Qualitätsmerkmale untersucht werden sollen und ihr Zusammenhang untereinander geprüft werden soll (vgl. Kapitel 3.2), scheint inter- und transdisziplinäres Zusammenarbeiten unabdingbar zu sein. Transdisziplinarität ist immer dann gefragt, wenn das Erkenntnisinteresse einer bestimmten Disziplin oder eine Fragestellung mehrere andere Disziplinen tangiert und das Ziel verfolgt wird, eine gemeinsame, breit abgestützte Antwort zu finden. Dies ist beispielsweise beim Thema der Textaufgaben oder des Problemlösens der Fall (vgl. Kapitel 2.1), während die Frage nach der Qualität fachsprachlicher Formulierungen zum besseren Verstehensaufbau bestimmter mathematischer Konzepte eine interdisziplinäre Zusammenarbeit mit linguistischer Forschung als sinnvoll erscheinen lässt.

2.3 Aufsatz 2: Interdisziplinärer Beitrag zur Theorieentwicklung von Unterrichtsqualität

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

Als eine Weiterführung der bisherigen Arbeiten rund um die Fragen von Unterrichtsqualität kann der vorliegende Aufsatz interpretiert werden. Das interdisziplinär zusammengesetzte Team von Autorinnen und Autoren aus den Bereichen Erziehungswissenschaften, Mathematikdidaktik, Schulpädagogik und Schulpraxis bzw. Bildungsdirektion versucht darin, am Beispiel des Konzepts der drei Basisdimensionen² von Unterrichtsqualität (Klieme et al., 2006) aufzuzeigen, inwiefern eine allgemeine Theorie zu Unterrichtsqualität fehlt und welchen Aspekten eine umfassende Theorie guten Lehrens/Unterrichtens genügen müsste. Dazu wird auf philosophische und wissenschaftstheoretische Grundlagen zurückgegriffen, die sodann auf das Beispiel der Basisdimensionen von Unterrichtsqualität angewandt werden.

Im Aufsatz wird zunächst die Frage erörtert, ob Forschende in der empirischen Unterrichtsforschung sowie Lehrende in der Praxis der Lehrerinnen- und Lehrerbildung eine umfassende Unterrichtstheorie im Sinne einer „Theory of teaching“ diskutieren und ihren Arbeiten eine solche Theorie zugrunde legen, was bereits Terhart (2014) infrage stellte. Während die Lehrerinnen- und Lehrerbildung auf eine lange Tradition (allgemein)didaktischer Erörterungen (z. B. Baer, 2006; Reusser, 2006; Terhart, 2002) im Hinblick auf die Schulpraxis zurückgreift, fehlt eine vertiefte theoretisch-analytische Perspektive im Sinne einer Unterrichtstheorie oder einer Theorie des Unterrichtens jenseits von praktischen Erörterungen und Nutzbarkeit beinahe vollständig. Dieser Umstand führt dazu, dass Lehrerinnen- und Lehrerbildung stark auf die professionelle Praxis abzielt und entsprechendes Handeln reflektiert, was auch zu den zentralen Aufgaben der Lehrerinnen- und Lehrerbildung gehört. Was aber nicht selten ausbleibt, ist die empirische Validierung der diskutierten didaktischen Konzepte oder es wird eng und einseitig auf das Curriculum fokussiert. Seidel (2014) macht in ihrer Analyse deutlich, dass im Zusammenhang mit Unterrichtstheorie bzw. Theorie des Unterrichtens zwei unterschiedliche Paradigmen unterschieden werden können. Das erste Paradigma baut auf psychologischen Theorien auf (z. B. kognitionspsychologischen, motivationalen oder Theorien zu Metakognition) und identifiziert Merkmale im Unterricht, welche entsprechende Aspekte von Lernen unterstützen. Das zweite Paradigma hingegen versucht, Aspekte von Unterrichten zu systematisieren und diese mit der Leistung bzw. der Leistungsentwicklung der Schülerinnen und Schüler zu verbinden: Guter Unterricht, eine hohe Unterrichtsqualität und erfolgreiches Unterrichten manifestieren sich gemäß diesem Paradigma in einem hohen Lernerfolg der Schülerinnen und Schüler. Das im Aufsatz diskutierte und analysier-

² Als die drei Basisdimensionen von Unterrichtsqualität werden von Klieme et al. (2006) Klassenführung, konstruktive Unterstützung und kognitive Aktivierung beschrieben und empirisch bezüglich ihrer Wirksamkeit im Hinblick auf die Leistungsentwicklung der Schülerinnen und Schüler belegt.

te Modell der drei Basisdimensionen von Unterrichtsqualität (Klieme et al., 2006) gehört zum zweiten Paradigma.

Das Modell der drei Basisdimensionen von Unterrichtsqualität wird im Aufsatz ausführlich vorgestellt und anschließend entlang von zentralen wissenschaftstheoretischen Kriterien anhand der Fragestellung, inwiefern dabei von einer Unterrichtstheorie bzw. einer Theorie des Unterrichtens gesprochen werden kann, analysiert. Die zentralen wissenschaftstheoretischen Fragen, die herangezogen werden, beziehen sich u. a. darauf, inwiefern die getroffenen Aussage logisch und theoretisch konsistent aufeinander bezogen werden, inwiefern sie hierarchisch strukturiert, sparsam und gleichzeitig verständlich und nachvollziehbar sind und inwiefern sie präzise genug formuliert sind, um für eine empirische Überprüfung operationalisiert werden zu können bzw. ob sie bereits empirisch geprüft worden sind. Der Aufsatz kommt in seiner Analyse zum Schluss, dass das im deutschen Sprachraum im Zusammenhang mit Unterrichtsqualitätsforschung sehr bekannte und oft eingesetzte und zitierte Modell der drei Basisdimensionen den wissenschaftstheoretischen Ansprüchen, die an eine Theorie gestellt werden, nicht genügt und somit nicht als eine Unterrichtstheorie oder Theorie des Unterrichtens interpretiert werden kann.

Der Aufsatz mündet in eine Konklusion, in der zentrale zu bearbeitende Forschungsfragen mit Blick auf die Entwicklung einer umfassenden Unterrichtstheorie benannt werden. Dabei werden insbesondere fünf relevante Themen für zukünftige Forschung aufgeführt. Es sind dies die Notwendigkeit, 1) die den empirischen Untersuchungen zugrunde gelegten theoretischen Annahmen, Theorieteile und Modelle explizit darzulegen und nicht nur implizit darauf aufzubauen, 2) vermehrt auch theoretische Erörterungen und Forschungssynthesen zum Thema zu erarbeiten, 3) expliziter zu werden bezüglich der formulierten Hypothesen, angenommenen Werte und der durchgeführten und geplanten Forschungsprozesse, 4) die verwendeten Modelle und Frameworks explizit aufeinander zu beziehen bzw. sich davon abzugrenzen und schließlich 5) die im Beitrag ausgeführten Kriterien einer guten Theorie nicht nur selektiv, sondern in ihrer Gesamtheit zu berücksichtigen und die Qualität der eigenen Forschung daran zu prüfen.

2.4 Diskussion

Obwohl der vorgestellte zweite Aufsatz die Frage des Theoriebezugs und des Vorliegens einer umfassenden Theorie des Unterrichtens anhand eines in den Erziehungswissenschaften und der empirischen Bildungsforschung etablierten Modells von Unterrichtsqualitätsmerkmalen untersucht, lassen sich in der Fachdidaktik und der fachdidaktischen Forschung einerseits zahlreiche Parallelen beschreiben. Andererseits müssen sich Fachdidaktik und fachdidaktische Forschung – am Beispiel der Mathematikdidaktik – der Frage stellen, inwiefern ihre (didaktischen) Überlegungen, Modelle und Theorien zum einen Theorien im eigentlichen Sinne sind, das heißt den im Aufsatz dargelegten Kriterien entsprechen, und zum anderen als eigenständige und fachspezifische Theorien und Modelle interpretiert werden können.

Im Hinblick auf die erste Frage wäre zu klären, welche Theorien und Modelle die Mathematikdidaktik genau verwendet und inwiefern diese tatsächlich den Ansprüchen genügen, die an wissenschaftliche Theorien und Modelle gestellt werden. Die Frage nach der Eigenständigkeit kann an einem Beispiel aus der Mathematikdidaktik konkretisiert werden: Es liegen zahlreiche Unterrichtsplanungsvorschläge und teils elaborierte Gestaltungselemente zur Realisierung „guten Unterrichts“ aus allgemeindidaktischer Sicht vor (z. B. Ball & Forzani, 2009; Grell & Grell, 2010; Kron, 2008; Meyer, 1987, 2009; Wellenreuther, 2005), die in jüngerer Zeit in der fachdidaktischen

Literatur aufgegriffen und weitergeführt wurden (z. B. Barzel, Holzäpfel, Leuders & Streit, 2017; Bruder, Leuders & Büchter, 2016; Leuders, 2003; Leuders & Prediger, 2016). Es lässt sich allerdings konstatieren, dass diese in der Fachdidaktik vorgelegten Konzepte und Modelle kaum fachspezifisch und damit genuin fachdidaktisch konzipiert sind. Unterrichtsplanung und Unterrichtsreflexion erfolgen zwar anhand einer Mathematikstunde (Barzel et al., 2017), bleiben aber allgemeindidaktisch und fachunspezifisch konzipiert. Es geht um Zielklarheit und um systematische Unterrichtsplanung in einem bestimmten Inhaltsbereich, der aus einer übergeordneten Perspektive allerdings als austauschbar erscheint und erst in der Umsetzung der Planung für einen bestimmten Inhalt an Bedeutung gewinnt. Von den üblichen Standardsituationen³ zur Planung von Mathematikunterricht (Barzel et al., 2017) ist die Mehrheit allgemeindidaktisch ausgerichtet. Lediglich das Erarbeiten eines Verfahrens entlang von (heuristischen) Lösungsbeispielen, das in der empirischen Forschung beispielsweise von Zöttl et al. (2010) aufgegriffen wird, kann als fachspezifisch betrachtet werden, wobei auch hier ein Anteil von allgemeindidaktischen Unterrichtsprinzipien enthalten sein dürfte. Gleiches zeigt sich auch in aktuellen Debatten wie beispielsweise derjenigen zur Inklusion. In der Sonderpädagogik sind zwar normative Setzungen etabliert und es liegen didaktische Konzepte vor (z. B. Feuser, 2013). Diese erweisen sich bei genauerer fachdidaktischer Betrachtung für die Unterrichtsgestaltung eines bestimmten fachlichen Inhalts jedoch als wenig präzise und teilweise auch als nicht praktikabel (z. B. Brunner, 2015) und erfordern eine fachspezifische und inhaltsbezogene Konkretisierung (z. B. Pöhler, Prediger & Wessel, 2016; Prediger, 2016), die empirisch überprüft werden kann (z. B. Freeseemann, 2014; Pool Maag & Moser Opitz, 2014; Schindler, Moser Opitz, Cadonau-Bieler & Ritterfeld, 2019; Wessel, 2015).

Das von Reusser (1991) bereits früh thematisierte zu klärende Verhältnis zwischen Allgemeiner Didaktik und Fachdidaktik erweist sich hier erneut als ein noch bzw. permanent zu bearbeitendes Aufgabenfeld. Die Metapher der „fremden Schwestern“, die Terhart (2002, S. 77) zur Charakterisierung des Verhältnisses von empirischer Lehr-Lern-Forschung und Allgemeiner Didaktik verwendet, wurde in Kapitel 2.1 bereits im Zusammenhang mit dem Primat der Erziehungswissenschaften beigezogen und kann auch für das Verhältnis von Allgemeiner Didaktik und Fachdidaktik adaptiert werden, wobei zurzeit auch hier noch ein Primat der Allgemeinen Didaktik gegenüber der Fachdidaktik besteht.

Vor diesem Hintergrund soll der vorgestellte Aufsatz des interdisziplinären Autorinnen- und Autorenteam (vgl. Kapitel 2.3) mit seinen fünf Schlussfolgerungen und Forderungen, die es bei weiterer Forschung zu berücksichtigen gelte, dazu anregen, die bereits vorliegende empirische Forschung zu Unterrichtsqualität und zu Unterrichtsgestaltung weiter voranzutreiben und diese gleichzeitig auf ein theoretisch solider abgestütztes Fundament zu stellen. Das Ziel besteht darin, einen Beitrag zur Entwicklung einer Theorie zu leisten, die nicht nur im Feld der Erziehungswissenschaften, sondern ebenso sehr auch in den Fachdidaktiken und in der fachdidaktischen Forschung ihre Verwendung findet und sowohl aus erziehungswissenschaftlicher und allgemeindidaktischer Perspektive als auch aus fachdidaktischer Sicht im disziplinübergreifenden Diskurs gemeinsam weiterentwickelt wird. In diesem Sinne spiegelt der Aufsatz das Bemühen, Fachdidaktik als ergänzende Perspektive mit erziehungswissenschaftlichen Grundlagen fruchtbar zusam-

³ Barzel et al. (2017, S. 27 ff.) unterschieden fünf Standardsituationen für die Unterrichtsplanung im Fach Mathematik: 1) fragend-entwickelnde Erarbeitung, 2) Verfahren erarbeiten an Lösungsbeispielen, 3) forschend-entdeckendes Lernen, 4) Sammeln – Sichern – Systematisieren, 5) differenzierendes Üben.

menzuführen und aus einem fundierten Verständnis des Fachs und seines Inhalts heraus einen Beitrag dazu zu leisten, dass sich fachdidaktische Forschung und Theoriebildung etablieren können und sich sowohl disziplinär als auch interdisziplinär mit spezifischen Fragen von Unterrichtsgestaltung und Unterrichtsqualität befassen. Aufsatz 2 stellt somit zum einen eine inhaltliche Weiterführung von Aufsatz 1 (vgl. Kapitel 2.2) dar, die durch die Kooperation mit weiteren Autorinnen und Autoren aus anderen Disziplinen erreicht werden konnte, und bietet zum anderen eine Grundlage für die empirischen Arbeiten, die in den nachfolgenden Kapiteln vorgestellt werden.

3 Unterrichtsqualität und der fachdidaktische Blick

3.1 Unterrichtsqualität: generisch oder fachspezifisch?

Unterrichtsqualität zu beschreiben, zu erfassen und zu bewerten ist in verschiedener Hinsicht von Relevanz und zudem von aktueller Brisanz. Im US-amerikanischen Kontext wird die Messung von Unterrichtsqualität zur Beurteilung der Arbeit und zur Qualifikation von Lehrpersonen herangezogen, was jedoch aus mehreren – theoretischen, methodischen und unterrichtspraktischen – Gründen problematisch ist: 1) *Messgrundlage und Anzahl eingeschätzter Unterrichtsstunden*: Eine Qualifikation der Arbeit einer Lehrperson auf der Basis einer beurteilten Unterrichtssequenz vorzunehmen ist unzureichend, weil die Messgrundlage dafür zu schmal ist. So konnten beispielsweise Praetorius, Pauli, Reusser, Rakoczy und Klieme (2014) nachweisen, dass die Einschätzung von generischen Qualitätsmerkmalen wie der Klassenführung anhand einer einzigen Unterrichtsstunde relativ zuverlässig eingeschätzt werden kann, während Einschätzungen eher fachspezifischer Merkmale wie beispielsweise kognitiver Aktivierung deutlich mehr Variabilität aufweisen, weshalb sehr viel mehr Messungen bzw. mehrere Unterrichtsstunden einbezogen werden müssten. 2) *Messinstrument*: Eine Messung von Unterrichtsqualität – generischer oder fachspezifischer Art – setzt voraus, dass ein valides und reliables Instrument eingesetzt wird. Diesbezüglich zeigt der Blick auf die verschiedenen vorhandenen Instrumente aber große Unterschiede (Übersicht in Schlesinger & Jentsch, 2016) und es stellt sich die Frage, wie so viele unterschiedliche Instrumente zu einem verlässlichen Urteil zur Qualität der geleisteten Arbeit einer Lehrperson kommen sollen und ob sie dasselbe Messergebnis erzeugen würden. 3) *Theoretisches Paradigma der Unterrichtsforschung*: Wird die festgestellte Unterrichtsqualität einzig auf die Person der Lehrperson zurückgeführt, wird Unterrichtsqualität als Produkt des Handelns von Lehrpersonen definiert, wobei wesentliche Faktoren aufseiten der Schülerinnen und Schüler, Variablen der Rahmenbedingungen und des Kontexts sowie grundsätzlich die Nutzung des schulischen Angebots durch die Schülerinnen und Schüler vernachlässigt werden. Die empirische Bildungsforschung hat deshalb im Zusammenhang mit Unterrichtsforschung und Unterrichtsqualität einen Paradigmenwechsel vollzogen – weg vom lehrpersonenzentrierten Ansatz hin zum variablenzentrierten (Helmke, 2006; Helmke & Weinert, 1997; Weinert, Schrader & Helmke, 1990). Diese Entwicklung hat dazu geführt, dass weniger (Charakter-)Eigenschaften und Personenmerkmale von Lehrpersonen als ursächlich für das Lernen der Schülerinnen und Schüler und die Qualität des Unterrichts angesehen werden, als vielmehr das Wechselspiel verschiedener Variablen, die im Angebot-Nutzungs-Modell (Pauli, Reusser & Grob, 2010) von Unterricht ihren Niederschlag gefunden haben.

Unterrichtsqualität wird seither als mehrdimensionales Konstrukt verstanden, das zum einen theoretisch im Kontext der Allgemeinen Didaktik (z. B. Meyer, 2009) diskutiert wird und zum anderen empirisch insbesondere in der empirischen Bildungsforschung untersucht wird (z. B. Praetorius, 2014). In der deutschsprachigen Unterrichtsforschung gelten Angebot-Nutzungs-Modelle von Unterricht, die es mittlerweile in einer Vielzahl gibt, als Standard für eine solide, theoretische Grundlage, auf der sich verschiedene angenommene und bestätigte Wirkungszusammenhänge modellieren lassen. Wie sich diesbezüglich feststellen lässt, fehlen eine solche Auseinandersetzung mit Angebot-Nutzungs-Modellen im amerikanischen Raum beinahe vollständig. Unterschieden wird stattdessen etwa nach Berliner (2005) zwischen „good teaching“, „effective teaching“ und „quality teaching“. Ersteres – übersetzt als „guter Unterricht“ (Kunter & Ewald, 2016) – bezieht sich auf das Einhalten normativer Prinzipien und Standards des Feldes und

schließt damit eng an die Sicht der Allgemeinen Didaktik an. Im Zentrum stehen normative Überlegungen zur Bestimmung von gutem Unterricht und letztlich von angemessener Bildung. Unter „effective teaching“ hingegen wird die Effektivität des Unterrichts verstanden. Gegenstand des Interesses ist hier das Erreichen bestimmter Ziele. Diese können ihrerseits wiederum sehr vielfältig sein (Kunter, 2005). Im Zentrum stehen somit Outcome-Variablen von Unterricht, insbesondere die Leistung der Schülerinnen und Schüler. Auf dieser Basis wird retrospektiv auf eine hohe Unterrichtsqualität geschlossen. „Quality teaching“ schließlich subsumiert die beiden Bereiche „good teaching“ und „effective teaching“. Eine hohe Unterrichtsqualität ist demnach definiert durch normative Anteile einerseits und durch die empirisch nachweisbare Erreichung entsprechender (Bildungs-)Ziele andererseits.

Unberücksichtigt bleibt in diesem Verständnis von Berliner (2005), wie bereits erwähnt, die Interaktion zwischen dem schulischen Angebot und seiner Nutzung durch die Lernenden. Ebenfalls nicht berücksichtigt werden Kontextmerkmale, die einen erheblichen Einfluss auf das Angebot wie auch dessen Nutzung haben können, wie dies verschiedene Studien beispielsweise für die Klassenzusammensetzung (z. B. Biedermann, Weber, Herzog-Punzenberger & Nagel, 2016; Carman & Zhang, 2012; Ditton, 2013) nachweisen konnten. Ausgeblendet bleibt in diesem Verständnis von Unterrichtsqualität aber auch die Domänenspezifität von Unterricht. Unterrichtsqualität wird hier generisch gefasst und untersteht dem Primat der Erziehungswissenschaften, obwohl sich sowohl normative Aspekte als auch empirisch erhärtete Merkmale auf der Basis des Outputs von Unterricht auch fachspezifisch interpretieren lassen. Ein normatives Postulat der Mathematikdidaktik ist beispielsweise in der Forderung von Radatz et al. (1996) begründet, dass das Mathematiklernen in Sinnzusammenhängen oder auf individuellen Wegen erfolgen solle. Ähnliche Postulate lassen sich auch im internationalen Kontext der Mathematikdidaktik finden (z. B. Clements et al., 2004; Heuvel-Panhuizen, 2001; NCTM, 2000). Ein Schwachpunkt besteht darin, dass diese normativen Postulate, so plausibel sie auch erscheinen mögen, nicht immer empirisch überprüft wurden. Im Gegensatz dazu liegen empirische Befunde zu generischen Unterrichtsmerkmalen vor, die nachweisbar mit einem hohen Lernerfolg der Schülerinnen und Schüler einhergehen, beispielsweise eine hohe Strukturiertheit und Klarheit, Motivierungsqualität, Klassenführung oder die Nutzung der effektiven Lernzeit (z. B. Brophy & Good, 1986; Helmke, 2004; Lipowsky, 2015; Praetorius, 2014).

In jüngerer Zeit wurden in der empirischen Unterrichtsforschung vermehrt auch fachnähere fachdidaktische Aspekte von Unterrichtsqualität untersucht (z. B. Klieme et al., 2009), welche die generischen Qualitätsmerkmale ergänzen. Im Zentrum standen dabei Merkmale wie kognitive Aktivierung im Fachunterricht (z. B. Pauli et al., 2010; Rakoczy et al., 2010) oder die Aufgabenqualität (z. B. Klieme et al., 2001; Neubrand, 2002). Eine vergleichsweise neue Entwicklung stellen Arbeiten dar, die Methoden der empirischen Bildungsforschung auf das Untersuchen von fachdidaktischen Konstrukten im engeren Sinne anwenden (z. B. Brunner, 2013; Drollinger-Vetter, 2011). Dennoch lässt sich im Feld der Unterrichtsforschung nach wie vor das Primat der Erziehungswissenschaften und der empirischen Bildungsforschung konstatieren. Interdisziplinäre Forschungsteams oder der Einbezug ausgewiesener Expertinnen und Experten aus dem Bereich der Fachdidaktik in Studien zur Bestimmung von Unterrichtsqualität sind nach wie vor eher selten. Das Primat der Erziehungswissenschaften ist aber nur so lange sinnvoll, als es sich um die Erforschung fachunspezifischer Qualitätsaspekte handelt, und stößt überall dort an Grenzen, wo einschlägige inhaltliche Expertise notwendig ist. Es ist daher unabdingbar, die Qualität von Mathematikunterricht auch aus dieser fachspezifischen und inhaltspezifischen Sicht heraus zu erfor-

schen, um die generische Perspektive zu ergänzen, wie dies die im Rahmen dieses Kapitels dargestellten Studien beabsichtigen. Diese Diskussion wurde in den letzten Jahren auch in anderen Studien aufgegriffen. Auch Lipowsky, Drollinger-Vetter, Klieme Pauli, und Reusser (2018) kamen in ihrer Analyse der Pythagoras-Daten (Klieme et al., 2009) zum Schluss, dass es sich bei den generischen und den fachspezifischen Qualitätsdimensionen nicht um zwei untrennbare Seiten einer Medaille handle, sondern um unterscheidbare Säulen von Unterrichtsqualität, wie dies die nachfolgend dargestellte Studie (vgl. Kapitel 3.2) schon früher zeigen konnte. Dies deutet ebenfalls darauf hin, dass fachspezifische Qualitätsmerkmale eigenständig und als von den generischen unabhängig zu konzipieren sind und dass ihr Einbezug zentral ist, wenn es um die Einschätzung der Qualität von Fachunterricht geht.

Zur weiteren Bearbeitung der offenen Fragen ist zum einen die Entwicklung geeigneter fach- und inhaltspezifischer Analyseinstrumente notwendig und zum anderen ist auf Theorieebene eine integrierende Sichtweise von generischen und fachspezifischen Qualitätsmerkmalen zu leisten, die in einem Modell zur Qualität von Fachunterricht zusammengeführt werden können. Von diesen Versuchen berichten die nachfolgenden Kapitel 3.2–3.5.

3.2 Studie 1: Qualitätssteigerung dank fachspezifischem Unterrichtscoaching

Brunner, E., Kreis, A., Staub, F. C., Schoy-Lutz, M. & Kosorok Labhart, C. (2014). Qualitätssteigerung von Mathematikunterricht dank Fachspezifischem Unterrichtscoaching. In J. Roth & J. Ames (Hrsg.), *Beiträge zur Mathematikdidaktik 2014* (S. 273-276). Koblenz: Universität Koblenz-Landau.

Im Rahmen einer Interventionsstudie der Pädagogischen Hochschule Thurgau (Kreis, 2012; Kreis & Staub, 2011) wurden Praxislehrpersonen (Mentorinnen und Mentoren) der Interventionsgruppe in zweierlei Hinsicht gezielt weiterqualifiziert: Mittels eines innovativen Coachingansatzes wurden sie zu fachspezifischen Unterrichtscoachs (West & Staub, 2003) ausgebildet. Parallel dazu erhielten sie eine mathematikdidaktische Fortbildung zu verschiedenen relevanten und aktuellen Fragen aus der Mathematikdidaktik. Die Fortbildung umfasste ca. 50 Kursstunden und 70 Stunden Selbststudium während eines Schuljahres und war entlang der sieben Merkmale erfolgreicher Fortbildungen von Lipowsky (2004) konzipiert. Im Anschluss an diese Fortbildung der Praxislehrpersonen folgte ein siebenwöchiges Praktikum von Lehramtsstudierenden, die eine Ausbildung zur Primarschullehrperson absolvierten, entweder bei den ausgebildeten Coachs (Interventionsgruppe, $N = 16$ Dyaden) oder bei Praxislehrpersonen, die keine solche Fortbildung besucht hatten (Kontrollgruppe, $N = 16$ Dyaden).

Ziel der Studie war es, zu prüfen, ob es mit dem Ansatz des Fachspezifischen Unterrichtscoachings gelingt, die Qualität des Mathematikunterrichts von Studierenden, das heißt der Coachees, zu steigern. In einem ersten Schritt sollte untersucht werden, ob und inwiefern sich Merkmale von Unterrichtsqualität von videografierten Mathematikktionen der Lehramtsstudierenden in der Interventions- und der Kontrollgruppe unterscheiden. Dazu wurde je eine Sequenz von 45 Minuten Unterricht der Studierenden videografiert und in einem hochinferenten Rating bezüglich ihrer Qualität eingeschätzt. Ein breit angelegtes Ratinginstrument (Ditton & Merz, 2000), das ausschließlich auf generischen Merkmalen beruhte, war zu diesem Zweck in der For-

schungsgruppe durch verschiedene Einzelitems zu fachspezifischen Qualitätsmerkmalen ergänzt worden. Aus der ursprünglichen Fragebogenskala von Ditton (2000; Ditton & Merz, 2000) wurden 16 Items zu Merkmalen von (generischer) Unterrichtsqualität einbezogen. Dazu gehörten „Klarheit“, „Verständlichkeit“, „Transparenz der Erwartungen“, „Differenzierungsqualität“, „diagnostische Kompetenz der Lehrperson“, „Interessantheit des Unterrichts“, „motivierende Unterstützung und Hilfestellung“, „Einübung des gelernten Unterrichtsstoffs“, „Zeitnutzung“, „Klassenführung“, „positives Verhältnis der Schülerinnen und Schüler zur Lehrperson“, „Schwierigkeit des Unterrichts“, „Leistungserwartungen“ (an die Lernenden), „inhaltliche Strukturiertheit“, „formal-kognitive Strukturiertheit“. Diese Merkmale wurden durch weitere Items ergänzt, die zusätzlich Aspekte fachdidaktischer Qualität fokussierten: Es waren dies „Offenheit der Aufgabenstellungen“, „Fehlerkultur“, „Verwendung fachspezifischer Begriffe“, „Alltagsbezug“, „fachliche Richtigkeit“, „Visualisierung und Verständnishilfen“ sowie „kognitive Aktivierung“. Die Anleitung zum Vorgehen beim Rating und die Operationalisierung der Items wurden orientiert an einem Instrument von Rakoczy und Pauli (2006) ausgearbeitet. Die Qualitätsmerkmale wurden analog zu den Merkmalen des generischen Instruments auf einer Skala von 1 bis 10 hochinferent eingeschätzt (Tabelle 1). Die Interraterreliabilität, die bei zehn zufällig bestimmten Mathematikstunden berechnet wurde, zeigte für die einzelnen Qualitätsaspekte akzeptable bis hohe Werte (Spearman's rho: von .496 bis .757, $p < .05$).

Tabelle 1: Beispiel Qualitätsmerkmal im Ratinginstrument nach Ditton (2000) und Rakoczy und Pauli (2006)

Bitte in jeder Zeile nur eine Ziffer ankreuzen.	1	2	3	4	5	6	7	8	9	10	Nicht beobachtbar
Offene Aufgabenstellungen:		○	○	○	○	○	○	○	○	○	○
Aufgabenstellungen lassen verschiedene Lösungsmöglichkeiten zu											

Die Ergebnisse der Studie lassen sich wie folgt zusammenfassen: Die Qualität des Unterrichts fiel in der Interventionsgruppe bezüglich einiger fachdidaktischer Merkmale, die in der Intervention fokussiert worden waren, signifikant höher aus ($t = 2.67$, $df = 30$, $p = .012$), während sich hinsichtlich generischer Merkmale kaum Unterschiede zeigten. Unterschiede ließen sich bezüglich fachlicher Verständlichkeit, der Interessantheit des Unterrichts, inhaltlicher Strukturiertheit (z. B. fachlich kohärenter und logischer Aufbau) sowie formal-kognitiver Strukturiertheit finden, jeweils zugunsten der Interventionsgruppe. In der Studie konnte somit gezeigt werden, dass eine intensive Fortbildung zu Fachspezifischem Unterrichtscoaching und Mathematikdidaktik für Praxislehrpersonen Wirkungen in der Unterrichtsqualität der Studierenden zeigt. Studierende, die ein Praktikum bei einer Praxislehrperson aus der Interventionsgruppe absolviert und somit in der Vorbereitung des gemeinsam verantworteten Unterrichts Fachspezifisches Unterrichtscoaching erhalten hatten, erreichten in verschiedenen relevanten Merkmalen eine höhere Unterrichtsqualität als ihre Mitstudierenden. Die höheren Qualitätswerte in den vier fachdidaktischen Merkmalen lassen vermuten, dass Praxislehrpersonen, die in der Fortbildung fokussierten mathematischen Grundfragen in der Coachingsituation intensiver berücksichtigt hatten.

In einer unveröffentlichten Weiterführung der Studie wurde anhand von Faktorenanalysen vertiefend geprüft, inwiefern sich zwei unterschiedliche Skalen von Unterrichtsqualität beschrei-

ben lassen, und zwar eine eher generische und eine eher fachspezifische. Theoretisch lassen sich die insgesamt 22 Unterrichtsqualitätsmerkmale einem der beiden Bereiche zuzuordnen. Merkmale wie „Verständlichkeit“, „fachliche Richtigkeit“, „Visualisierung und Verständnishilfen“, „offene Aufgabenstellungen“, „diagnostische Kompetenz“, „Interessantheit des Unterrichts“, „Fehlerkultur“, „Schwierigkeiten des Unterrichts“, „Leistungserwartungen“, „Verwendung von Fachbegriffen“, „Alltagsbezug“, „inhaltliche Strukturiertheit“, „formal-kognitive Strukturiertheit“ und „kognitive Aktivierung“ lassen sich nicht unabhängig vom fachspezifischen Inhalt einschätzen und können daher auch als fachspezifische Merkmale betrachtet werden. Aspekte wie „Klarheit“ (im Sinne der Klarheit des Handlungsprogramms), „Transparenz der Erwartungen“, „motivierende und unterstützende Hilfeleistungen“, „Differenzierung“, „Einübung des gelernten Lernstoffs“, „Zeitnutzung“, „Klassenführung“ und „positives Verhältnis zur Lehrperson“ hingegen können weitgehend unabhängig vom Fachinhalt eingeschätzt werden und beziehen sich stärker auf generische Aspekte von Unterrichtsqualität. Auf der Basis der 22 Einzelitems ließen sich gemäß dieser theoretisch begründeten Zuordnung zwei Skalen bilden: Die Skala zur „generischen Unterrichtsqualität“ ($\alpha = .70$) umfasste sieben Items, jene zur „fachspezifischen Unterrichtsqualität“ ($\alpha = .84$) elf Items. Für diese Skalen sowie für eine aus allen 22 Items gebildete Gesamtskala ($\alpha = .90$) wurden t -Tests für unabhängige Stichproben und das Effektstärkemaß Cohens d (Cohen, 1988) für die beiden Gruppen – Interventions- und Kontrollgruppe – berechnet. Für die Skala der „generischen Merkmale“ zeigte sich kein statistisch signifikanter Unterschied in der Qualität der Interventions- und der Kontrollgruppe ($M_{IG} = 8.77$, $SD_{IG} = 0.78$; $M_{KG} = 8.20$, $SD_{KG} = 1.40$). Ein solcher Unterschied ließ sich hingegen für die Skala „Fachspezifische Merkmale“ nachweisen ($t_{(30)} = 1.38$; $p < .05$), und zwar zugunsten der Interventionsgruppe ($M_{IG} = 8.00$, $SD_{IG} = 1.13$) im Vergleich mit der Kontrollgruppe ($M_{KG} = 6.52$, $SD_{KG} = 1.46$).

Auch wenn diese erste Studie sowie die nicht publizierte Weiterführung mit einigen methodischen Mängeln⁴ behaftet ist, so wird doch deutlich, dass Unterrichtsqualität unterschiedlich definiert werden kann und sich die Ergebnisse bei deren Einschätzung je nach Fokus durchaus unterscheiden können. Dies zeigen auch die beiden unterschiedlichen Skalen von Qualitätsmerkmalen, die sich nachweisen ließen. Deutlich wird ferner, dass sich generische und fachspezifische Merkmale nicht zwingend gleich stark ausgeprägt zeigen und dass eine gezielte Intervention mit fachspezifischen Inhalten möglicherweise einen Effekt auf die fachspezifische Betreuung von Lehramtsstudierenden hat und einen Mehrwert gegenüber einer ausschließlichen Fokussierung auf generische Merkmale aufweist. Damit stellt diese erste Studie eine Ausgangslage im Sinne eines ersten Schritts für die nachfolgende Auseinandersetzung mit dem Thema der Qualität von Mathematikunterricht aus fachdidaktischer Sicht dar.

⁴ Es fehlten beispielsweise Qualitätseinschätzungen des Unterrichts der Praxislehrpersonen vor der Intervention oder Qualitätseinschätzungen der Studierenden aus einem früheren Praktikum. Zudem konnten die Praxislehrpersonen wählen, ob sie der Interventions- oder der Kontrollgruppe angehören möchten. Die Stichprobe der Praxislehrpersonen ist somit nicht randomisiert. Ersteres wiegt schwerer, weil bei Letzterem angenommen werden kann, dass aufgrund der zufälligen Einteilung der Studierenden nicht von einer systematischen Verzerrung der Stichprobe ausgegangen werden muss, während es durchaus sein kann, dass bei Lehrpersonen, welche sich an der Langzeitfortbildung beteiligt hatten, aufgrund höheren Interesses auch eine höhere Unterrichtsqualität und ggf. auch eine höhere Betreuungsqualität der Lehramtsstudierenden vorlagen. Ein weiteres Problem der Studie besteht darin, dass nicht gesagt werden kann, auf welche Fortbildungs-Komponente (Coaching- oder Fachdidaktikinputs) die Effekte zurückgeführt werden können.

3.3 Studie 2: Qualität durch unterschiedliche Unterrichtsmethoden?

Bieg, M., Goetz, T., Sticca, F., Brunner, E., Becker, E. S., Morger, V. & Hubbard, K. (2017). Teaching methods and their impact on students' emotions in mathematics: an experience-sampling approach. *ZDM Mathematics Education*, 49(3), 411-422.

Der Fokus einer weiteren Studie zur Qualität von Mathematikunterricht lag auf Sichtstrukturmerkmalen von Mathematikunterricht. Anhand von verschiedenen Unterrichtsmethoden sollte die Frage geklärt werden, inwiefern diese einen Einfluss auf das emotionale Erleben von Schülerinnen und Schülern haben. Diesbezüglich liegen erst wenige Befunde vor und die verfügbaren Studie beziehen sich meist auf Selbstberichte von Lehrpersonen bzw. von Lernenden (z. B. Goetz, Lohrmann, Ganser & Haag, 2005) oder auf Videodaten (Givvin et al., 2005; Hiebert et al., 2003), weshalb sie nicht das unmittelbare emotionale Erleben während einer Mathematikstunde erfassen, sondern auf retrospektiven Einschätzungen beruhen. Die nachfolgend dargestellte Studie verfolgte deshalb das Ziel, einen Beitrag zur Schließung dieser Forschungslücke zu leisten. Leitende Fragestellung war, inwiefern sich Unterschiede im emotionalen Erleben von Schülerinnen und Schüler der Klassenstufe 9 im Mathematikunterricht in Phasen direkter Instruktion und Phasen von Gruppenarbeiten im Vergleich nachweisen und beschreiben lassen. Dazu wurde eine Stichprobe von $N = 141$ Schülerinnen und Schülern (54.5 % weiblich, $M_{\text{Alter}} = 15.64$ Jahre, $SD = 0.62$) aus 43 Klassen untersucht. Dieses Sample war Teil einer größeren Studie und wurde gebildet, indem aus jeder Klasse zufällig je 2 bis 4 Schülerinnen und Schüler gezogen wurden, die an der Experience-Sampling-Phase der Studie teilnehmen sollten. Erfasst wurden die Emotionen im Mathematikunterricht während zweier Wochen. Dies führte zu insgesamt 807 eingeschätzten Mathematikstunden. Am häufigsten kamen Phasen direkter Instruktion vor (42.6 %), gefolgt von individueller Einzelarbeit (24.5 %) und Gruppen- oder Tandemarbeiten (14.1 %). Die Mehrebenenanalysen zeigten, dass Phasen direkter Instruktion im Vergleich zu den anderen beiden Unterrichtsmethoden mit einem niedrigeren Level positiver Emotionen und einem höheren Level von Langeweile einhergegangen waren. Für die Emotionen Ärger und Angst zeigten sich demgegenüber keine signifikanten Zusammenhänge mit der Unterrichtsmethode auf der Sichtstrukturebene.

Die Ergebnisse der Studie deuten darauf hin, dass Sichtstrukturmerkmale bei den Schülerinnen und Schülern einen Einfluss auf das emotionale Erleben von Mathematikunterricht haben können. Dies gilt insbesondere für Formen direkter Instruktion, während denen mehr Langeweile berichtet wurde. Dass keine entsprechenden Befunde für die beiden anderen Unterrichtsmethoden – die Phase der selbstständigen individuellen Arbeit und die Gruppen- bzw. Tandemarbeitsphase – gefunden wurden, legt die Vermutung nahe, dass in Phasen der selbstständigen individuellen und gruppenbezogenen Arbeit weniger die Sichtstruktur das emotionale Erleben prägt als vielmehr der zu bearbeitende mathematische Inhalt. Das dabei festgestellte eigene Kompetenzerleben dürfte diesbezüglich ein bedeutsamer Faktor sein. Dies verdeutlicht im Hinblick auf das Thema der Unterrichtsqualität aus fachlicher Sicht, dass ein generisches Konzept von Unterrichtsqualität oder Sichtstrukturmerkmale wie Unterrichtsmethoden möglicherweise zu wenig präzise Voraussetzungen für die Beschreibung von emotionalem Erleben im Mathematikunter-

richt bilden und dass in Ergänzung dazu inhalts- und fachspezifische Qualitätsaspekte einbezogen werden müssten.

3.4 Studie 3: Hierarchisches Modell fachlicher Unterrichtsqualität

Brunner, E. (2018). Qualität von Mathematikunterricht: Eine Frage der Perspektive. *Journal für Mathematik-Didaktik JMD*, 39(2), 257-284.

In der vorliegenden explorativen Einzelfallstudie wurde am Beispiel einer gefilmten Mathematikstunde aus dem 6. Schuljahr zunächst untersucht, inwiefern die Qualität derselben Stunde ähnlich bzw. divergierend eingeschätzt wird, wenn sie anhand von drei unterschiedlichen hochinformativen Analyseinstrumenten bewertet wird. Dieser Fragestellung lagen die Vorarbeiten aus den früheren Studien (vgl. Kapitel 2.2, 3.2, 3.3) zugrunde. In einem zweiten Schritt wurde auf der Basis der gewonnenen Erkenntnisse ein hierarchisch gegliedertes Modell zur Qualität von Fachunterricht entwickelt und zur Diskussion gestellt.

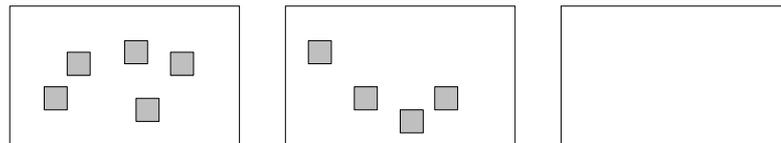
Die ausgewählte einzuschätzende Mathematikstunde wurde von einer Lehramtsstudentin am Ende eines Langzeitpraktikums zum Thema „Direkte Proportionalität“ bzw. „Maßstabgetreues Umrechnen von Größenangaben“ gehalten. Während der Durchführung der Stunde unterliefen der Studentin verschiedene gravierende fachliche Fehler, die in der Folge in erhöhte Anforderungen bezüglich der Klassenführung mündeten, die von der Studentin jedoch problemlos gemeistert wurden. Die gefilmte Stunde wurde ausgewählt, weil sie zum einen fachliche Fehler enthält, die von der Studentin selbst nicht wahrgenommen wurden, die sich aber in der Interaktion mit den Schülerinnen und Schülern auswirkten, und weil sie zum anderen durch gute Klassenführung der Studentin überzeugte. Der erste Eindruck ließ die Hypothese zu, dass die Stunde aus fachdidaktischer Sicht deutlich schlechter eingeschätzt werden dürfte als aus allgemeiner Sicht im Hinblick auf generische Qualitätsmerkmale.

Die drei für die Analyse ausgewählten Instrumente (Abbildung 3) unterschieden sich sowohl bezüglich ihrer zugrunde liegenden theoretischen Perspektive und ihres Fokus auf bestimmte Qualitätsaspekte als auch hinsichtlich ihrer Berücksichtigung und Gewichtung von fachspezifischen Qualitätsmerkmalen. Dadurch repräsentieren sie je einen anderen Prototyp von Analyseinstrumenten: Die Konzeption des ersten Instruments (Brunner, Kreis, Staub, Schoy-Lutz & Kosorok Labhart, 2014; Ditton & Merz, 2000) war additiv angelegt, da es eine Anzahl generischer Qualitätsmerkmale mit weiteren fachspezifischen ergänzte (vgl. Kapitel 3.2). Das zweite Instrument (Rakoczy & Pauli, 2006) kam in der Pythagoras-Studie (Klieme et al., 2009) zum Einsatz und wird als integratives Instrument bezeichnet, weil innerhalb der drei verwendeten Skalen jeweils einzelne Items fachspezifisch operationalisiert wurden. Als Beispiel für ein inklusives Instrument wurde das dritte Analyseinstrument (Schoenfeld, Floden & The Algebra Teaching and Mathematics Assessment Project, 2014) einbezogen, das auf der Grundlage von Mathematik und einem vertieften Fachverständnis weitere – u. a. auch generische – Dimensionen fachspezifisch operationalisiert. Von Charalambous und Praetorius (2018) wurde dieses Instrument als „hybrid“ kategorisiert (vgl. Kapitel 1). Diese Unterscheidung von Analyseinstrumenten in additive, integrierende und inklusive Instrumente, die der Beitrag vornimmt, ist neuartig und ergänzt aus ma-

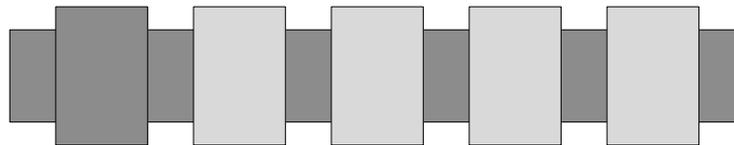
thematikdidaktischer Perspektive die Kategorisierung von Charalambous und Praetorius (2018), die beinahe zeitgleich publiziert wird.



Instrument 1: Ditton (2000); Ditton und Merz (2000); ergänzt durch Brunner et al. (2014)



Instrument 2: Rakoczy und Pauli (2006)



Instrument 3: Schoenfeld et al. (2014)

Abbildung 3: Übersicht über die drei Instrumente (dunkelgrau: fachspezifische Qualitätsaspekte; weiß: fachunspezifische, allgemeine Qualitätsaspekte; hellgrau: fachspezifische Ausgestaltung allgemeiner Aspekte) (Abbildung aus: Brunner, 2018, S. 266)

Die Ergebnisse der Studie zeigen deutlich, dass die Qualität derselben Mathematikstunde je nach Analyseinstrument unterschiedlich eingeschätzt wird. Insbesondere die mangelhafte fachliche Korrektheit des Unterrichts wurde in den einzelnen Instrumenten höchst unterschiedlich berücksichtigt. Mit dem additiven Instrument wurden die generischen Qualitätsmerkmale deutlich positiver eingeschätzt als die fachspezifischen. Mit dem zweiten, dem integrierenden, Instrument wurde fachliche Korrektheit nicht erfasst und fachspezifische Operationalisierungen insbesondere im Bereich des Wissensaufbaus wurden kritischer eingeschätzt als die generischen Merkmale. Das inklusive dritte Instrument schließlich wies die fachlichen Fehler mit niedrigen Werten im Bereich „Fach“ aus, in den anderen vier Qualitätsdimensionen wurde die Stunde in einem mittleren Bereich eingeschätzt.

Die Reflexion der Ergebnisse mündete in erste Überlegungen für einen Entwurf eines hierarchisch gegliederten Modells von fachlicher Unterrichtsqualität (Abbildung 4). Dieses Modell beruht auf der Annahme, dass wirkungsvoller Mathematikunterricht auf der Voraussetzung angemessener Klassenführung basiert und aus einer soliden fachlichen Fundierung heraus erfolgt. Auf den Ebenen 3 und 4 folgen sodann weitere Qualitätsmerkmale, die teils generisch, teils fachspezifisch sind. Auf Ebene 3 finden sich inhaltsunspezifische generische und fachspezifische Merkmale, auf Ebene 4 schließlich inhaltspezifische generische und fachspezifische Merkmale. Das Modell postuliert, dass die zu beobachtenden Qualitätsaspekte auf der obersten Ebene, das heißt Ebene 4, in Abhängigkeit vom zu bearbeitenden Inhalt ausgewählt werden müssen. Beweis-

sequenzen beispielsweise erfordern den Einbezug anderer inhaltspezifischer Qualitätsmerkmale als eine Übungssequenz zur Festigung des kleinen Einmaleins.

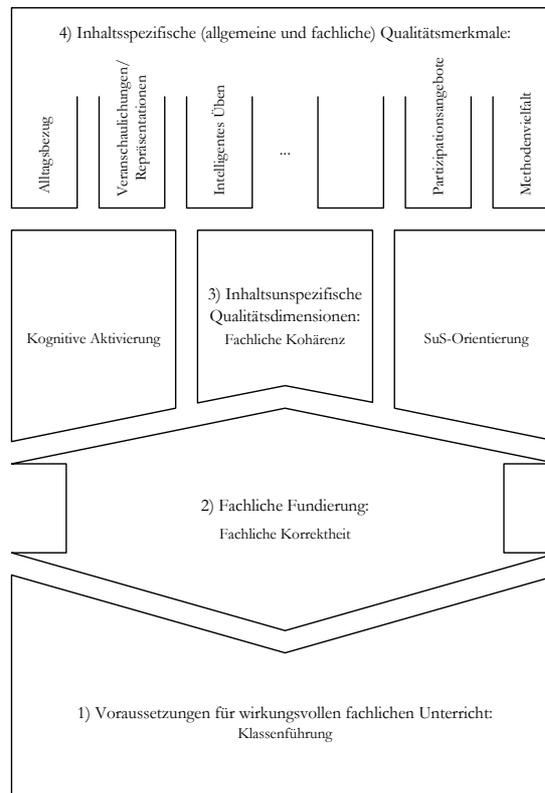


Abbildung 4: Hierarchisches Modell der Qualität von Fachunterricht (SuS: Schülerinnen und Schüler) (Abbildung aus: Brunner, 2018, S. 278)

Unterrichtsqualität als ein hierarchisch gegliedertes Konstrukt zu konzipieren, das aus inhaltsunspezifischen generischen und fachspezifischen (Ebene 3) und inhaltspezifischen Merkmalen (Ebene 4) besteht und auf den Voraussetzungen der Klassenführung (Ebene 1) und fachlicher Fundierung (Ebene 2) beruht, stellt ein Novum in der Forschungslandschaft dar.

3.5 Diskussion

In diesem Kapitel wurden drei Studien zur Beschreibung, Messung und Konzipierung fachlicher Unterrichtsqualität vorgestellt. Die erste Studie (vgl. Kapitel 3.2) unternahm als eine erste Auseinandersetzung mit dem Thema den Versuch, ein etabliertes hochinferentes generisches Ratinginstrument durch fachspezifische Qualitätsaspekte zu ergänzen und beleuchtete damit die Frage der Bedeutung der Domänenspezifität von Qualitätseinschätzungen. In der zweiten Studie (vgl. Kapitel 3.3) wurde nach dem Einfluss von Sichtstrukturmerkmalen auf das emotionale Erleben der Schülerinnen und Schüler, das im weiteren Sinne auch als ein Qualitätsmerkmal betrachtet werden kann (vgl. z. B. Skala „Motivationsunterstützung“ bei Rakoczy & Pauli, 2006), gefragt und festgestellt, dass die unmittelbar erkennbare Sichtstruktur zwar durchaus einen Einfluss auf das Erleben von Langeweile hat, dies aber nur dann, wenn der Unterricht als instruktionale Sequenz erfolgt. Sichtstrukturmerkmale scheinen somit zu wenig aussagekräftig zu sein für die Bestimmung von Unterrichtsqualität allgemein und zur Erforschung emotionalen Erlebens der

Schülerinnen und Schüler im Besonderen. Die dritte Studie (vgl. Kapitel 3.4) schließlich bietet Evidenz für die These, dass die Ergebnisse der Messung von Unterrichtsqualität vom gewählten Analyseinstrument abhängen und dass gerade für die Qualitätseinschätzung von Mathematikunterricht vorliegende Instrumente ungenügend sind, weil den meisten davon die zentrale Dimension fachlicher Korrektheit als Fundierung des Unterrichts fehlt bzw. weil fachliche Korrektheit in den Instrumenten – im Gegensatz zu angemessener Klassenführung – bereits vorausgesetzt wird. Diese dritte Studie legte sodann mit dem hierarchisch gegliederten Modell der Qualität von Mathematikunterricht eine Diskussionsgrundlage vor, die verschiedene der in Kapitel 1 aufgeführten Dimensionen, in denen sich offene Fragen stellen, anspricht: 1) Fachspezifität vs. generische Qualität, 2) Fachspezifität vs. Inhaltsspezifität, 3) Tiefen- vs. Sichtstruktur von Unterricht und 4) Hierarchisierung von Qualitätsmerkmalen. Noch unbearbeitet bleibt in den drei Studien die fünfte Dimension, die Frage nach Wechselwirkungen von verschiedenen Qualitätsmerkmalen untereinander umfasst. Die Frage der Fach- vs. Inhaltsspezifität wird im nächsten Kapitel vertiefend bearbeitet.

4 Qualitätsfragen am Beispiel mathematischen Begründens, Argumentierens und Beweisens

Unterrichtsqualität – fachlich oder generisch interpretiert – bezieht sich auf die gemeinsame, ko-konstruktive (z. B. Greeno, 2006; Pauli & Reusser, 2015; Resnick, 1992; Reusser, 2006) Bearbeitung eines Lerngegenstands (vgl. didaktisches Dreieck, Abbildung 1) und damit auf einen Inhalt. Wenn es sich bei diesem Lerngegenstand um einen fachlichen Inhalt bzw. eine fachliche Kompetenz handelt, sind die Standards und Normen des Fachs entscheidend (z. B. Common Core State Standards Initiative, 2012; D-EDK, 2016; KMK, 2005). Diese beschreiben die von den Schülerinnen und Schülern zu einem bestimmten Zeitpunkt zu erreichenden Anforderungen inhaltlich und strukturieren darüber hinaus die fachliche Bildung, die die Lernenden zu einem bestimmten Zeitpunkt von der Schule erwarten können.⁵ Als übergreifende Zielsetzung von Mathematikunterricht im Sinne einer umfassenden mathematischen Bildung gilt seit vielen Jahren das Verstehen von Mathematik als zentrale Leitlinie (Freudenthal, 1977; Hiebert et al., 1997; Winter, 1975). Dies wird verbunden mit der Notwendigkeit, erworbenes und aufgebautes Wissen in neuen Kontexten anwenden und somit mathematisch kompetent handeln zu können. Der Anspruch des Verstehens und des kompetenten fachlichen Handelns gilt aber nicht nur für anwendungsbezogene Kontexte, sondern ebenso sehr für innermathematische Bereiche. In diesen das Fach strukturierenden Inhaltsbereichen sollen ebenfalls zentrale Handlungskompetenzen erworben und aktiv angewandt werden.

Mathematisches Argumentieren gehört zu den zentralen inhaltlichen Handlungskompetenzen, die aufzubauen sind – und zwar vom Kindergarten an bis zur Hochschule im Sinne eines konsequenten kumulativen Wissensaufbaus (Heinze & Grüßing, 2009) und einer Enkulturation in den Fachbereich (Bishop, 1997). Mathematisches Argumentieren, Begründen und Beweisen gehört zu den herausfordernden Kompetenzen, und dies sowohl für die Lernenden als auch für die Lehrenden (zusammenfassend Brunner, 2014), wie dies verschiedene empirische Studien zeigen. Mathematisches Argumentieren und Beweisen setzt einerseits einen anspruchsvollen kognitiven Prozess voraus und ist andererseits in einem sozialen Kontext situiert, weil mathematische Argumente im Dialog vorgebracht und von der Community validiert werden müssen (Brunner, 2014). Zentral beim zugrunde liegenden Denkprozess sind das Erkennen mathematischer Strukturen und ihrer sie konstituierenden Merkmale, das Erkennen und Aufdecken von inhärenten Beziehungen, das Formulieren möglicher Hypothesen und Argumente zum Zustandekommen einer bestimmten Struktur sowie die Verallgemeinerung auf der Basis logischen Schließens (Jahnke & Ufer, 2015). Auf der Seite der (Lern-)Gemeinschaft sind Aspekte des Überprüfens und Validierens vorgebrachter Argumente entscheidend.

Wird aus fachspezifischer Sicht die Qualität von Beweisequenzen oder Phasen mathematischen Argumentierens fokussiert, bedeutet dies, entlang des Beweisprozesses essenzielle fachlich-

⁵ Die schweizerischen Bildungsstandards (D-EDK, 2014) definieren Mindeststandards, die von (beinahe) allen (95 %) Schülerinnen und Schülern jeweils am Ende eines Zyklus von vier bzw. drei Jahren zu erreichen sind. Die deutschen Bildungsstandards (KMK, 2005) hingegen definieren die Anforderungen als Regelstandard, der sich an die Mehrheit der Lernenden richtet und nicht von allen erreicht werden muss. Dieser Umstand ist für die Festlegung von Bildungszielen und überprüfbaren Leistungen erheblich.

inhaltliche Qualitätsaspekte zu definieren, die im Weiteren in ihrer Umsetzung im Unterricht analysiert werden. Essenzielle Qualitätsaspekte von Beweisequenzen betreffen die Gestaltung des Gesprächs (Studie 4, vgl. Kapitel 4.1), die Repräsentation bestimmter Beweise (Studie 5, vgl. Kapitel 4.2) und deren Bezug zu den Lernvoraussetzungen der Schülerinnen und Schüler im Sinne adaptiven Unterrichtens (vgl. Kapitel 4.2). Des Weiteren muss die Qualität von Beweisequenzen bzw. von Phasen mathematischen Argumentierens im Kontext der Voraussetzungen der Lernenden einerseits und der Schulstufe andererseits beleuchtet werden. So gelten beispielsweise zur Einschätzung von erfolgreichen Phasen mathematischen Argumentierens im Kindergarten (Studie 7, vgl. Kapitel 4.4) andere Merkmale als im Hochschulkontext oder in der Sekundarstufe I (z. B. Brunner, 2013), weil diese Qualitätsmerkmale in Abhängigkeit von der Konzeptualisierung des mathematischen Prozesses und seiner Entwicklung (Studie 6, vgl. Kapitel 4.3) zu berücksichtigen sind. Mit diesen Fragen befassen sich die nachfolgend dargestellten Studien dieses Kapitels.

4.1 Studie 4: Beweistypen und Unterrichtsgespräch

Brunner, E. (2014). Verschiedene Beweistypen und ihre Umsetzung im Unterrichtsgespräch. *Journal für Mathematik-Didaktik*, 35(2), 229-249.

Im Rahmen dieser Studie wurden Kommunikationsmuster als ein Merkmal generischer Unterrichtsqualität auf die zur Lösung einer bestimmten Aufgabenstellung eingesetzten, inhaltlich bearbeiteten Beweistypen bezogen. Die vier Kommunikationsmuster wurden entlang der Dimensionen „Steuerung durch die Lehrperson“ und „Aktive Partizipation der Lernenden“ konzipiert und lassen sich zusammenfassend als 1) Modelling (Collins, Brown & Newman, 1989), 2) Lehr-Lern-Gespräch entlang des I-R-E-Musters (Mehan, 1979), 3) Scaffolding (Collins et al., 1989) und 4) ko-konstruktiver Dialog beschreiben. Dabei ist die Steuerung durch die Lehrperson von Muster 1 hin zu Muster 4 abnehmend, während die aktive Partizipationsmöglichkeit für die Lernenden zugleich zunimmt. Die Studie untersuchte jedoch nicht nur die Qualität des Gesprächs, wie dies zahlreiche andere Arbeiten zuvor getan hatten (z. B. Pauli & Lipowsky, 2007; Pauli & Reusser, 2015; Sfard, 2007; van de Pol et al., 2017, 2010), sondern versteht die unterschiedlichen Gesprächsmuster als grundsätzlich unterschiedliche Ansätze, die es ermöglichen, Beweisprozesse in adaptivem Sinne den Möglichkeiten der Schülerinnen und Schüler anzupassen. Ein sehr anspruchsvoller Beweis beispielsweise könnte mit Lernenden, die über geringe mathematische Voraussetzungen und algebraische Fähigkeiten verfügen, durch ein Modelling bearbeitet werden, bei dem die Lernenden kaum mit aktiven Beiträgen partizipieren müssen. Die drei Beweistypen von Wittmann und Müller (1988) – experimentelle „Beweise“, operative oder inhaltlich-anschauliche Beweise und formal-deduktive Beweise (zusammenfassend Brunner, 2014) – bieten deshalb ein unterschiedliches didaktisches Potenzial für den Brückenschlag zwischen inhaltlicher Anforderung einerseits und Voraussetzungen der Lernenden andererseits.

In der vorliegenden Arbeit wurde am Datensatz der Pythagoras-Studie (Klieme et al., 2009) in 32 Klassen der Sekundarstufe I (8./9. Schuljahr) untersucht, mit welchen Beweistypen und in welchen Kommunikationsmustern dieselbe Beweisaufgabe aus der elementaren Zahlentheorie bearbeitet wird. Die Ergebnisse verdeutlichen zum einen, dass bei der Bearbeitung derselben Aufgabe eine große Vielfalt besteht und alle drei Beweistypen auftreten. Zum anderen zeigte sich, dass anspruchsvolle, abstrakte formal-deduktive Beweise mehrheitlich in einem eng geführten

Unterrichtsgespräch nach dem I-R-E-Muster erarbeitet wurden, während operative oder inhaltlich-anschauliche Beweise im I-R-E-Muster erarbeitet wurden, aber etwa gleich häufig auch in einem Scaffolding mit einer höheren Partizipation der Lernenden umgesetzt wurden. Für die (wenigen beobachteten) experimentellen Beweise zeigte sich keine klare Präferenz. Sie wurden aber erwartungswidrig häufig im I-R-E-Muster durchgeführt, obwohl sie Potenzial für eine hohe Partizipation der Lernenden bieten würden.

Die Studie weist darauf hin, dass nicht nur Sichtstrukturmerkmale von Unterricht allgemein den Möglichkeitsraum für das Lernen auf der Ebene der Tiefenstruktur darstellen, sondern dass der Möglichkeitsraum auch durch die inhaltliche Anforderung definiert wird. Unterrichtsqualität in Beweisssequenzen zu bestimmen, bedeutet zum einen, die Qualität des Gesprächs zu analysieren, und zum anderen, dies in Abhängigkeit von einem bestimmten inhaltlichen Aspekt zu tun. Beim mathematischen Beweisen ist somit weniger von einer „absoluten“ Unterrichtsqualität auszugehen als vielmehr von einer „relationalen“: Das Herstellen einer Brücke zwischen den Voraussetzungen der Lernenden einerseits und den inhaltlichen Anforderungen andererseits kann als spezifische Qualität im Hinblick auf adaptives Unterrichten interpretiert werden.

4.2 Studie 5: Beweistypen und Merkmale von Lehrpersonen

Brunner, E. & Reusser, K. (2019). Type of Mathematical Proof: Personal Preference or Adaptive Teaching Behavior? *ZDM Mathematics Education*, 51(5), 747-758.

Im Rahmen einer vertiefenden Analyse der Daten aus der Pythagoras-Studie (Klieme et al., 2009) wurde in 32 Klassen des 8. und 9. Schuljahres aus der Schweiz und Deutschland der Frage nachgegangen, wie Lehrpersonen dieselbe Beweisaufgabe aus der elementaren Zahlentheorie mit ihren Klassen bearbeiteten. Im Zentrum stand dabei insbesondere die Frage, inwiefern Gestaltungselemente wie Beweistypen, die auf der (fachspezifischen) Sichtstrukturebene angesiedelt sind, aber eine bestimmte Art des Denkens repräsentieren, mit Merkmalen der Lehrperson einerseits und mit mittleren Voraussetzungen der Klasse andererseits zusammenhängen. Die ausgewählte Beweisaufgabe kann unterschiedlich gelöst werden. Sowohl die Repräsentationsformen des Denkens als auch die eingesetzten heuristischen Strategien und Hilfsmittel können sich unterscheiden. Diese feinkörnige Analyse ermöglicht es, Merkmale auf der Ebene der Sichtstruktur mit Merkmalen auf der Ebene der Tiefenstruktur zu verbinden. So lassen beispielsweise die verwendeten Repräsentationsformen einen Rückschluss auf den Denkprozess zu, während heuristische Hilfsmittel und Strategien anzeigen, welche Lösungsweisen und Lösungsideen beim Bearbeitungsprozess verfolgt werden. Nicht alle Lösungs- und Vorgehensweisen sind jedoch gleich einfach umsetzbar. Je nach mathematischen, insbesondere algebraischen Voraussetzungen sind gewisse Lösungsweisen zweckmäßiger als andere. Deshalb ergibt es Sinn, im Hinblick auf eine adaptive Unterrichtsgestaltung die mittleren Voraussetzungen der Klasse zu berücksichtigen.

Die explorative Studie fragte danach, inwiefern der Einsatz eines bestimmten Beweistyps als eine Manifestation adaptiven Lehrens interpretiert werden kann. Um diese Frage untersuchen zu können, wurden mögliche korrelative Zusammenhänge zwischen den drei Beweistypen – experimentelle „Beweise“, operative oder inhaltlich-anschauliche Beweise und formal-deduktive Beweise (Wittmann & Müller, 1988; zusammenfassend Brunner, 2014) – einerseits und den Überzeugungen und weiteren persönlichen Merkmalen der Lehrpersonen andererseits sowie den mathe-

matischen und algebraischen Voraussetzungen der Lernenden gesucht. Dabei zeigte sich, dass die drei Beweistypen in den Klassen unterschiedlich häufig eingesetzt wurden. Feststellbar war eine Dominanz desjenigen Beweistyps, der die höchsten Anforderungen an Formalisierung und Abstraktion stellt, das heißt des formal-deduktiven Beweises. Aber auch inhaltlich-anschauliche oder operative Beweise kamen oft vor. Insgesamt machen die Ergebnisse deutlich, dass besonders der formal-deduktive Beweistyp als eine persönliche Präferenz einer ganz bestimmten Gruppe von Lehrpersonen interpretiert werden kann und dass damit hohe mathematische Voraussetzungen der Schülerinnen und Schüler einhergehen. Der Einsatz dieses anspruchsvollen Beweistyps kann somit als Ausdruck adaptiven Unterrichtens betrachtet werden.

Die Ergebnisse werfen die Frage auf, ob ein bestimmtes fachdidaktisches Merkmal auf der Ebene der Sichtstruktur – auf der auch der Beweistyp zu verorten ist – als Indikator für adaptives Unterrichten verstanden werden kann. Da diese Frage anhand des gewählten Samples nicht beantwortet werden kann, besteht hier weiterer Forschungsbedarf. Zu klären ist, inwiefern Lehrpersonen mithilfe von bestimmten Beweistypen tatsächlich adaptiv unterrichten, das heißt inwiefern sie bei ihrer Unterrichtsplanung auf die mittleren Voraussetzungen der Klassen Bezug nehmen und sich für einen passenden Beweistyp entscheiden. Im Hinblick auf Merkmale von Unterrichtsqualität verdeutlicht die Studie, dass sowohl Sichtstrukturmerkmale als auch Tiefenstrukturmerkmale fach- und inhaltspezifisch erhoben bzw. analysiert werden können und müssen, um vertiefte Erkenntnisse zu fachlichem Lernen und adaptivem Unterrichten von Mathematik gewinnen zu können.

Inwiefern (zukünftige) Lehrpersonen der Sekundarstufe I die drei unterschiedlichen Beweistypen tatsächlich adaptiv im Hinblick auf zentrale Klassenmerkmale wie die mathematischen Vorkenntnisse einsetzen würden, ist Gegenstand einer laufenden weiterführenden Studie mit einem Sample von $N = 183$ Lehramtsstudierenden Mathematik (Sommerhoff, Brunner & Ufer, 2019). Diese Studie geht der Frage nach, wovon die Wahl eines bestimmten Beweistyps abhängt und prüft empirisch mögliche Faktoren, die diese Wahl beeinflussen könnten wie beispielsweise Aufgaben- und Beweismerkmale, Klassenmerkmale, Merkmale der (zukünftigen) Lehrpersonen wie ihre eigene Beweiskompetenz sowie Merkmale der Lehr-Lern-Situation. Die Ergebnisse werden derzeit in einem englischsprachigen Manuskript zusammengefasst und zur Publikation vorbereitet. Erkenntnisse zu möglichen Einflussfaktoren auf die Wahl eines bestimmten Beweistyps könnten sodann im Hinblick auf die Förderung der Qualität von Beweisequenzen gezielt fokussiert werden.

4.3 Studie 6: Konzeptualisierung frühen mathematischen Argumentierens

Lindmeier, A., Brunner, E. & Grüßing, M. (2018). Early mathematical reasoning – theoretical foundations and possible assessment. In E. Bergqvist, M. Österholm, C. Granberg & L. Sumpter (Hrsg.), *Proceedings of the 42th Conference of the International Group for the Psychology of Mathematics Education* (Bd. 3, S. 315–322). Umea: PME.

Mathematisches Argumentieren ist ein komplexer und anspruchsvoller Prozess. Deshalb ergibt es im Hinblick auf einen kontinuierlichen und kumulativen Wissensaufbau (Heinze & Grüßing, 2009) Sinn, mathematisches Argumentieren über alle Bildungsstufen hinweg zu fördern. Damit

rückt auch die frühe Bildungsstufe in den Fokus. Mit dieser befasste sich die vorliegende Studie. Zunächst wurde dargelegt, dass im Unterschied zu fundierten und breit abgestützten Forschungsbefunden zum Prozess von mathematischem Argumentieren und Beweisen auf der Sekundarstufe I nur wenig theoretisches und empirisches Wissen zum mathematischen Argumentieren junger Kinder im Allgemeinen vorliegt. Insbesondere eine fundierte theoretische Grundlage im Sinne einer Modellierung des Prozesses, die zum einen die Möglichkeiten und Voraussetzungen junger Kinder angemessen berücksichtigt und zum anderen mit dem Prozess des späteren mathematischen Beweisens kompatibel ist, ist kaum verfügbar. Im Beitrag wurde frühes mathematisches Argumentieren sodann als ein Prozess von fünf zentralen Schritten beschrieben, die sich auf zwei unterschiedliche Facetten frühen mathematischen Argumentierens beziehen. Das kognitive Modell, das seit dem Erscheinen des Beitrags kontinuierlich weiterentwickelt, aber noch nicht veröffentlicht wurde (Abbildung 5; englischsprachige Publikation nach erfolgtem Peer-Review-Verfahren in Überarbeitung), beschreibt zwei unterschiedliche Levels des Denkprozesses. Das erste Level bezieht sich auf das Erkennen und Beschreiben relevanter mathematischer Strukturen und Beziehungen, während das zweite Level das Begründen und Verallgemeinern ebendieser mathematischen Strukturen und Beziehungen fokussiert. Mit dem ersten Level werden Prozesse wie das Erkennen, Beschreiben und Repräsentieren mathematischer Strukturen und Beziehungen angesprochen. Beim zweiten Level stehen Prozesse des Suchens und Findens einer Begründung, des Konstruierens einer Argumentation, aber auch des Überprüfens einer gegebenen Begründung und das Reproduzieren einer mathematischen Argumentation im Zentrum.

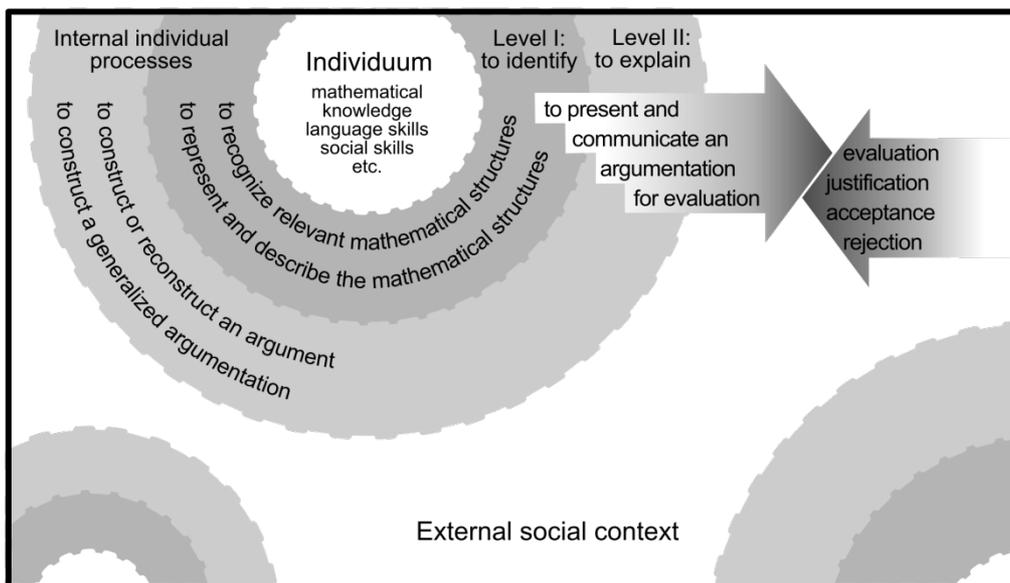


Abbildung 5: Modell frühen mathematischen Argumentierens (Brunner, Heinze, Grüßing & Lindmeier, in Rev.)

In der vorliegenden publizierten Studie wurde auf der Basis der ausgeführten theoretischen Grundlagen am Beispiel von zwei Aufgabenstellungen exemplarisch gezeigt, was sich ein Kind im Kindergartenalter beim Bearbeitungs- bzw. beim Begründungsprozess überlegen muss und in welcher Weise die inhaltliche Anforderung der Aufgabe den Begründungsprozess beeinflusst und welche inhaltlichen Lösungsideen in dieser Bildungsphase überhaupt möglich sind.

Die Analyse der beiden Beispielaufgaben zum mathematischen Argumentieren junger Kinder zeichnet zum einen idealtypisch den Denkprozess nach und verdeutlicht zum anderen, dass die inhaltliche Begründung eines erkannten mathematischen Zusammenhangs unterschiedlichen Lösungsideen folgen kann. Damit können die von Stylianides (2016) beschriebenen drei Komponenten einer Argumentation⁶ durch eine vierte ergänzt werden, nämlich durch diejenige der inhaltlichen Lösungsidee. Alle vier Komponenten lassen sich auch zur Einschätzung der Qualität einer Begründungsleistung nutzen, wie dies andere derzeit laufende Arbeiten tun (Brunner, im Druck).

Auch wenn der Beitrag selbst dies so nicht ausführte, stellt dieser inhaltliche Einbezug der Lösungsidee eine Weiterentwicklung bestehender Arbeiten dar, da er theoretische Grundlagen aus dem Bereich mathematischen Argumentierens mit einer inhaltlichen und damit aufgabenbezogenen Lösungsidee verbindet und fachdidaktische Überlegungen zum Prozess mathematischen Argumentierens mit einer fachwissenschaftlichen Perspektive zu möglichen inhaltlichen Lösungsideen verknüpft.

4.4 Studie 7: Förderung frühen mathematischen Argumentierens

Brunner, E. (2019). Förderung mathematischen Argumentierens im Kindergarten: Erste Erkenntnisse aus einer Pilotstudie. *Journal für Mathematik-Didaktik JMD*, 40(2), 323–356.

Mit der Förderung frühen Argumentierens auf der Basis der in Kapitel 4.3 dargelegten theoretischen Grundlagen befasst sich Studie 7, die als Pilotstudie angelegt war und das Ziel verfolgte, im Rahmen einer Intervention mit neun Kindergartenlehrpersonen während eines Schuljahres Lernumgebungen zum mathematischen Argumentieren im Kindergarten zu entwickeln, zu erproben und zu evaluieren. Dazu absolvierten die Kindergartenlehrpersonen eine mehrteilige Fortbildung, die entlang der Merkmale erfolgreicher Fortbildungen (Lipowsky, 2004) konzipiert worden war. In dieser Fortbildung lernten die Lehrpersonen theoretische Grundlagen frühen Argumentierens sowie curriculare Vorgaben zum frühen mathematischen Argumentieren im Kindergarten (Amt für Volksschule des Kantons Thurgau, 2016) kennen. Anhand von vier eigens für die Fortbildung entwickelten prototypischen didaktischen Situationen zum frühen mathematischen Argumentieren (Brunner, 2016, 2018a) erhielten die beteiligten Lehrpersonen fachspezifische Situationen vorgestellt, die zum einen reichhaltiges mathematisches Argumentieren fokussieren und dieses zum anderen didaktisch mit bekannter Kindergartenpraxis und -tradition verbinden, in der dem Spiel (Gasteiger, 2015; Van Oers, 2010) und dem Lernen in natürlichen Situationen (Gasteiger, 2012) besondere Bedeutung beigemessen wird. Mithilfe von Puppenspielen und dem Einsatz von Bilderbüchern (van den Heuvel-Panhuizen & van den Boogaard, 2008) wurde gezeigt, wie im Kindergarten mit fünf- und sechsjährigen Kindern altersgemäß, aber dennoch fachspezifisch und inhaltsnah argumentiert werden kann. Diese prototypischen Situationen wurden während der

⁶ Stylianides (2016, S. 14) benennt die folgenden drei Komponenten einer Argumentation: 1) das Set akzeptierter Statements, 2) die Art der Argumentation und 3) die Art der Repräsentation der Begründung.

Studie von den Kindergartenlehrpersonen aufgegriffen, für ihre eigene Praxis adaptiert und mit einem zu bearbeitenden mathematischen Inhalt verbunden.

Die erstellten fachlichen Unterrichtsplanungen zum mathematischen Argumentieren bilden einen Teil des umfangreichen und multiperspektivischen Datenkorpus. Zwei der vier durchgeführten Einheiten der Kindergartenlehrpersonen wurden überdies gefilmt und im Hinblick auf die beobachtbare Praxis frühen mathematischen Argumentierens inhaltsanalytisch (Mayring, 2002) entlang verschiedener Aspekte kategorisiert und ausgewertet. Ergänzend wurden zu Beginn und zum Ende des Schuljahres im Rahmen einer schriftlichen Befragung die professionellen Kompetenzen der Kindergartenlehrpersonen, ihre Praxis mathematischen Argumentierens und ihre erlebten bzw. vermuteten Schwierigkeiten in Form von Selbstberichten erfasst. Auf der Ebene der teilnehmenden Kinder wurden zu Beginn und zum Ende des Schuljahres pro Klasse die Leistungen von je 4 bis 5 Kindergartenkindern erfasst, die von der Lehrperson als repräsentativ für das ganze Leistungsspektrum eingeschätzt worden waren.

In der hier berichteten Studie stand die professionelle Entwicklung der Kindergartenlehrpersonen und ihrer Argumentationspraxis im Verlauf eines Schuljahres im Zentrum. Der Beitrag konzentrierte sich auf die Entwicklung und das fachdidaktische Handeln der Kindergartenlehrpersonen bezüglich der Förderung mathematischen Argumentierens während eines Schuljahres, dies allerdings – was für eine Pilotstudie akzeptabel ist, aber keine Wirksamkeitsuntersuchung zulässt – ohne Bezug zu einer Kontrollgruppe. Im Fokus standen die Selbsteinschätzungen der spezifischen Praxis mathematischen Argumentierens und der eigenen fachdidaktischen Kompetenzen zum mathematischen Argumentieren sowie die Argumentationspraxis im Kindergarten. Geklärt werden sollte die Frage, inwiefern es den Kindergartenlehrpersonen gelingt, mathematisches Argumentieren mit verschiedenen Teilkompetenzen (vgl. Kapitel 4.3) umzusetzen. Aus der Sicht partizipativer Lehr-Lern-Dialoge (z. B. Gillies, 2015; Greeno, 2006, 2015; Pauli & Reusser, 2015; Reusser & Pauli, 2011) interessierte ferner, wer ein Argument bzw. einen Teil davon produziert. Ebenfalls von Interesse war, in welcher Repräsentationsform (Bruner, 1971; Stylianides, 2015, 2016) ein Argument vorgebracht wird und inwiefern es sich dabei um ein mathematisches Argument bzw. um ein Argument handelt, das nur im Alltag zugelassen ist und kein im fachlichen Sinne mathematisches ist (Fischer & Malle, 2004; Reiss, 2002).

Erste Ergebnisse zeichnen ein positives Bild der selbst eingeschätzten professionellen Entwicklung der beteiligten Lehrpersonen sowie ihrer Praxis mathematischen Argumentierens und machen deutlich, dass es gelingen kann, im Kindergarten eine vielfältige, reichhaltige Praxis mathematischen Argumentierens zu etablieren. Diese war geprägt von einer altersgerechten und dennoch fachlich fokussierten Praxis mathematischen Argumentierens, die spezifische Argumentationstätigkeiten (z. B. Formulieren oder Überprüfen eines Arguments) einbezog und die jungen Kinder in hohem Masse aktiv am fachlichen Lehr-Lern-Dialog partizipieren ließ. Dass die meisten hervorgebrachten Argumente bzw. Teile davon in narrativer Form repräsentiert wurden, war in Anbetracht von Befunden aus der Primarschule, wonach in dieser Schulstufe narrative Repräsentationen vorherrschend sind (Krummheuer, 2008), erstens erwartbar. Zweitens dürfte dieses Überwiegen narrativer Repräsentation auch auf die didaktische Anlage der vier prototypischen Situationen zurückzuführen sein, die jeweils ausgehend von Bilderbüchern oder Puppenspielen selbst bereits die narrative Form favorisierten.

Obwohl auf der Basis der kleinen Stichprobe und der Anlage der Pilotstudie keine Verallgemeinerungen getroffen oder Wirksamkeitsüberlegungen angestellt werden können, leistet die In-

terventionsstudie doch einen Einblick in aktuelle Kindergartenpraxis, der es gelingen kann, fachlich gehalt- und anspruchsvolle Zielsetzungen wie mathematisches Argumentieren differenziert und altersgerecht umzusetzen. Gleichzeitig macht die Studie deutlich, dass das Vorhandensein von konkreten fachspezifischen Unterrichtsvorschlägen im Sinne von Modellen, die eine Orientierungsfunktion einnehmen, sinnvoll ist für eine gute Praxis mathematischen Handelns und Argumentierens im Kindergarten. Dies ist insbesondere dann bedeutsam, wenn – wie auf der Kindergartenstufe – kaum fachnahe Ansätze zur gezielten Förderung einzelner mathematischer Kompetenzen vorliegen.

4.5 Diskussion

Die vier vorgestellten Studien zu Qualitätsfragen am Beispiel mathematischen Beweisens und Argumentierens in unterschiedlichen Schulstufen beleuchten insbesondere vier zentrale Dimensionen: Unterrichtsqualität ist nicht nur fachspezifisch, sondern darüber hinaus auch (1) kompetenzspezifisch und (2) inhaltspezifisch zu konzeptualisieren. Die Qualität von Beweis- bzw. mathematischen Argumentationssequenzen manifestiert sich – etwas vereinfacht gesagt – darin, dass die die Kompetenz des mathematischen Beweisens und Argumentierens konstituierenden Prozesse im Unterricht gezielt fokussiert und bearbeitet werden. Ein Unterricht, der sich mit Beweisen und mathematischem Argumentieren befasst, muss spezifische Tätigkeiten des Kompetenzerwerbs von Beweisen und Argumentieren beinhalten und entsprechendes Lernen ermöglichen, wie dies im Rahmen der Interventionsstudie im Kindergarten ausgeführt wurde (vgl. Kapitel 4.4). Der Blick auf die Kompetenz und deren Aufbau durch die Adressierung spezifischer Tätigkeiten und Prozesse erfordert eine inhaltspezifische Ergänzung, durch welche die Kreativität des Denkprozesses in den Mittelpunkt rückt. Diese inhaltspezifische Ergänzung ist aufgaben- und themenabhängig, wie dies anhand der Beispiele in Studie 6 (vgl. Kapitel 4.3) gezeigt werden konnte. Es gibt verschiedene inhaltliche Ansätze, die als Lösungs- bzw. Begründungsidee herangezogen werden können. Die Lösungsideen beschreiben und charakterisieren das Denken in inhaltlicher Hinsicht und eröffnen, sofern im Sinne „multipler Lösungswege“ (Neubrand, 2006, S. 162 ff.) unterschiedliche Lösungsideen zugelassen werden, neue Perspektiven auf und Einsichten in andere Denkweisen.

Die Notwendigkeit, im Hinblick auf Qualitätsfragen von Beweis- und Begründungssequenzen im Unterricht nicht nur die Bedeutung des sozialen Kontexts generell einzubeziehen, sondern darüber hinaus auch (3) die Art des Gesprächs und der eingesetzten Muster didaktischer Kommunikation in Abhängigkeit vom gewählten Vorgehen zu betrachten, verdeutlicht Studie 4 (vgl. Kapitel 4.1). Der in dieser Studie dargelegten Argumentation zufolge weisen verschiedene Beweistypen im Sinne eines prototypischen Vorgehens beim Beweisen und Begründen unterschiedliches Potenzial für die gemeinsame Er- und Bearbeitung im Gespräch sowie die Partizipation der Lernenden auf. Eine Beschreibung der Qualität von Beweis- und Begründungsprozessen im Mathematikunterricht hat somit eine weitere Dimension zu berücksichtigen: diejenige (4) des Diskurses. Diesbezüglich wäre neben dem Erfassen von didaktischen Kommunikationsmustern eine Beschreibung des Diskurses entlang der Merkmale des „Accountable Talk“ (Greeno, 2006, 2015) sinnvoll, die sich an den Leitfragen, inwiefern Verantwortlichkeit für den Inhalt, die Strenge des Denkens sowie die Lerngemeinschaft (geteilt) übernommen wird, orientiert.

Als vierte im Zusammenhang mit der Erfassung der Qualität mathematischen Argumentierens und Beweisens zu berücksichtigende Dimension wird in Studie 7 (vgl. Kapitel 4.4) deutlich,

dass (4) die didaktischen Konzepte zum mathematischen Argumentieren zum einen altersangemessen und zum anderen anschlussfähig an die entsprechende Unterrichtstradition sein sollten. Am Beispiel des Kindergartens wurde aufgezeigt, wie die Arbeit mit Bilderbüchern und Puppenspielen eine Brücke schlagen kann zwischen bekannter Kindergartenpraxis einerseits und den Anforderungen des mathematischen Prozesses andererseits. Dass dies durchaus fruchtbar erfolgen kann, zeigen die Ergebnisse der vorgestellten Studie.

Studie 5 (vgl. Kapitel 4.2) schließlich weist darauf hin, wie eng diese Qualitätsdimensionen in ihrer Realisierung, beispielsweise anhand von unterschiedlichen prototypischen Vorgehensweisen, möglicherweise mit dem fachlichen Profil der Unterrichtenden und ihrer Beziehung zum Unterrichtsfach Mathematik zusammenhängen. Obwohl die in der Studie erfragten Überzeugungen kaum zu deutlich akzentuierten Ergebnissen geführt haben, kann vermutet werden, dass ein sehr viel inhaltsnäheres Erfassen von spezifischen Beliefs zu Lehren und Lernen von Beweisen und mathematischem Argumentieren mehr Aufschluss geben dürfte über mögliche Präferenzen von Lehrpersonen und ihre Entscheidungen in Bezug auf das Vorgehen, das sie für die Arbeit mit ihren Klassen wählen bzw. ausschließen.

Die vier im Rahmen dieses Kapitels vorgestellten Studien erweitern und vertiefen am Beispiel der Qualität von Beweis- und Begründungsphasen die einschlägigen Qualitätsdimensionen, wie sie aus der empirischen Bildungsforschung (vgl. Kapitel 2.1) und der Allgemeinen Didaktik (vgl. Kapitel 2.3) bekannt sind, durch mindestens vier wesentliche Aspekte: 1) Kompetenzspezifität, 2) Inhaltsspezifität, 3) Kommunikations- bzw. Diskurspezifität und 4) Alters- und Kontextspezifität. Diese vier zusätzlichen Dimensionen, die bei der Bestimmung von Unterrichtsqualität gemäß den berichteten Befunden auch einbezogen werden sollten, wurden bislang jedoch nicht diskutiert. Sie nehmen den in Studie 3 (vgl. Kapitel 3.4) vorgestellten ersten Entwurf eines hierarchischen Modells von Unterrichtsqualität auf und präzisieren und konkretisieren diesen. Dabei wird insbesondere die vierte Ebene, diejenige der inhaltspezifischen Qualitätsmerkmale, fokussiert (Abbildung 4, vgl. Kapitel 3.4). Ergänzt werden muss der Entwurf eines hierarchischen Modells der Qualität von Fachunterricht um die Dimensionen der Kompetenzspezifität (1) und die Dimension der Alters- und Kontextspezifität (4), während die Dimension der Inhaltsspezifität (2) und die Dimension der Kommunikations- bzw. Diskurspezifität (3) auf Ebene 4 bereits enthalten sind. Als zusammenfassende Schlussfolgerung kann postuliert werden, dass der Inhalt der Mathematikstunde und die darin fokussierte Kompetenz den Rahmen für die weiteren sinnvoll einzubeziehenden und zu untersuchenden Qualitätsdimensionen und -aspekte vorgeben.

Eine erste Konkretisierung der inhalts- und kompetenzfokussierten Erfassung von Qualität von Beweis- und Begründungssequenzen könnte entlang des TRU-Frameworks – „Teaching for Robust Understanding“ (Schoenfeld et al., 2014) – erfolgen und in Form eines drei- bzw. vierstufigen Rubrics entlang der zentralen sechs „Robust Criteria of Understanding“, den RCs⁷, operationalisiert werden. In diesen RCs werden die zentralen Aktivitäten während des Begründungs-, Argumentations- und/oder Beweisprozesses beschrieben. Dabei ist der Prozess des Begründens, Argumentierens und Beweisens – auch als „cognitive unit“ (Garuti, Boero & Lemut, 1998) bezeichnet – ein „seeking why and explaining why“ (Mason, Burton & Stacey, 1982, p. 78). Es geht

⁷ Analog zu den für das Unterrichten von Algebra ausgearbeiteten „Robust Criteria of Understanding“ (RCs) (Wernet & Lepak, 2014).

um ein „trying to see why the conjecture might be justified, that is from seeking WHAT to seeking WHY. ... The answer to WHY is a structure which links what you KNOW to what you have conjectured“ (Mason et al., 1982, p. 82). Somit stellen die sechs RCs einen Versuch dar, die Kompetenzspezifität von Unterrichtsqualität weiter zu präzisieren. Diese sechs Bearbeitungs- bzw. Prozessschritte, die sich auf den Denk- und Handlungsprozess beziehen, wurden für drei Qualitätsstufen operationalisiert, wobei zusätzlich eine Stufe 0 eingeführt wurde (Tabelle 2). Die drei Qualitätsstufen orientieren sich an der Überzeugungsfunktion von Beweisen (Hersh, 1993) und beziehen diese auf drei unterschiedliche Instanzen eines Gegenübers, das die vorgebrachte Begründung hinsichtlich ihrer Überzeugungskraft und ihrer (Allgemein-)Gültigkeit evaluiert. Diese drei Instanzen haben Mason et al. (1982, S. 87) wie folgt beschrieben: 1) Convince yourself. 2) Convince a friend. 3) Convince a sceptic or an enemy.

Tabelle 2: Erster Entwurf zentraler Kriterien („Robust Criteria“) für die Qualität von Beweis- und Begründungssequenzen

	RC 1	RC 2	RC 3	RC 4	RC 5	RC 6
	Spezifischen kognitiven Konflikt einer Unsicherheit erleben und dadurch Notwendigkeit für das Finden einer Begründung erkennen	Forschen und erkunden, um Sicherheit bezüglich der geprüften Beispiele oder eines speziellen Falls zu erlangen (verifizieren oder falsifizieren)	Erkennen von inhärenten Beziehungen in der besonderen Situation bzw. im besonderen Beispiel und erkennen ihres Potenzials im Hinblick auf eine Verallgemeinerung	Verallgemeinerung vornehmen und gültige Schlussfolgerungen ziehen	Zentrale Argumente als Argumentationskette ordnen und formulieren und Denken repräsentieren	Begründung, Argumentation oder Beweis angemessen und überzeugend kommunizieren
0: Niemanden überzeugen	Es wird keine Notwendigkeit, etwas zu begründen oder zu beweisen, erkannt. Die Behauptung oder Vermutung wird ohne Prüfung akzeptiert.	Es ist kein Erforschen oder Erkunden eines möglichen mathematischen Zusammenhangs an Beispielen oder an einem speziellen Fall erkennbar oder es werden nur ein bis zwei Beispiele geprüft oder diese passen nicht zur Behauptung oder zur Vermutung bzw. sind falsch.	Es werden keine inhärenten Beziehungen in der besonderen Situation bzw. im besonderen Beispiel erkannt.	Es werden keine Schlussfolgerungen gezogen. Die Aktivitäten führen nicht zu Konsequenzen, weder für einen Einzelfall noch für einen allgemeinen Fall.	Es werden lediglich Beispiele präsentiert, aber keine verallgemeinernden Argumente. Für die Repräsentation der Beispiele ist kein Transfer nötig. Sie erfolgt in derselben Form, in der die Beispiele bereits vorliegen oder in der Behauptung oder Vermutung enthalten sind.	Die Argumentation oder der Beweis wird nicht kommuniziert, weder mündlich noch schriftlich. Es fehlt eine an die Lerngemeinschaft gerichtete Erklärung, die über die möglicherweise getesteten Beispiele hinausgeht und auf die Behauptung oder Vermutung Bezug nimmt.

1: Sich selbst überzeugen	Es ist eine teilweise Unsicherheit erkennbar, aber es ist nicht weiter spezifiziert, worin diese Unsicherheit genau besteht. Sie bleibt diffus.	Es finden einige Erkundungen (z.B. Falsifikation) oder Verifikation zur Vermutung oder Behauptung statt, diese sind aber lokal und auf ein bis zwei (passende und korrekte) Beispiele beschränkt.	Es werden einzelne Aspekte des inhärenten Zusammenhangs am besonderen Fall bzw. an der besonderen Situation erkannt.	Es wird versucht, einzelne Schlussfolgerungen für den Einzelfall zu ziehen. Dies gelingt jedoch nicht oder die Schlussfolgerungen sind entweder nicht vollständig oder nicht gültig.	Es ist ein Versuch einer Argumentation zum gefundenen Zusammenhang erkennbar, aber die Repräsentation fehlt, ist unvollständig oder fehlerhaft.	Es wird versucht, den gefundenen Zusammenhang in einer Argumentation oder einem Beweis zu kommunizieren, aber es fehlen verschiedene entscheidende Schritte oder die Erklärung ist nicht nachvollziehbar und nicht oder zu wenig verständlich.
2: Eine befreundete Person überzeugen	Es ist Unsicherheit bezüglich der Gültigkeit der Behauptung oder Vermutung erkennbar, aber es fehlt eine genaue Sicht dessen, wie die Gültigkeit der Behauptung oder Vermutung geprüft werden könnte.	Eine Vermutung oder Behauptung wird an verschiedenen (mehr als zwei) passenden Beispielen getestet.	Der inhärente Zusammenhang in der besonderen Situation wird erkannt, aber es gelingt noch nicht, dies ausgehend von einem generischen Beispiel zu verallgemeinern.	Es werden entscheidende Schlussfolgerungen für den Einzelfall gezogen, aber noch nicht für den allgemeinen Fall.	Eine Verallgemeinerung erfolgt und wird zumindest teilweise korrekt ikonisch oder narrativ repräsentiert oder ist korrekt auf der Basis der Beispiele repräsentiert.	Die Problemlöseidee wird entweder auf semantisch-inhaltlicher oder syntaktisch-formaler Ebene kommuniziert und der gefundene Zusammenhang wird erklärt. Aber die Argumentation ist nicht vollständig oder teilweise inkorrekt, was ein lückenloses Nachvollziehen erschwert.
3: Eine skeptische Person überzeugen	Eine Begründungsnotwendigkeit bezüglich der Gültigkeit der Behauptung oder Vermutung ist erkennbar und es ist klar, dass eine solche Begründung über einen mathematischen Zusammenhang (Beweis) gefunden werden muss.	Eine Vermutung oder Behauptung wird an einem generischen Beispiel oder an einem allgemeinen Fall geprüft.	Es findet eine Verallgemeinerung statt und die entdeckten Zusammenhänge und Beziehungen werden an einem generischen Beispiel oder verallgemeinert erkannt.	Die gezogenen Schlussfolgerungen sind logisch korrekt, vollständig und für die Begründung des Einzelfalls sowie des allgemeinen Sachverhalts ausreichend.	Die vorgenommene Verallgemeinerung ist korrekt ikonisch (operativ), narrativ oder formal-deduktiv repräsentiert.	Die Problemlöseidee wird vollständig sowohl auf semantisch-inhaltlicher als auch auf syntaktisch-formaler Ebene formuliert und kommuniziert und der gefundene Zusammenhang wird dadurch vollständig erklärt.

Inwiefern sich dieser erste Entwurf eines Instruments zur Erfassung von kompetenz- und diskurspezifischen Dimensionen mathematischen Beweisens und Argumentierens auch empirisch einsetzen lässt und praktikabel sowie reliabel ist, wird sich erst noch zeigen müssen.

5 Qualitätsfragen am Beispiel des Mathematikunterrichts in unterschiedlichen Kontexten

Wie eingangs bereits erwähnt (vgl. Kapitel 1), wird Unterricht – und somit auch die Unterrichtsqualität – von kulturell geprägten Rahmenbedingungen beeinflusst, wie dies das Angebot-Nutzungs-Modell von Unterricht nahelegt (Abbildung 6). Die angebotsbezogenen Stützsysteme wirken diesem Modell zufolge zum einen direkt auf den Unterricht und zum anderen indirekt über die Lehrperson und ihre Merkmale und beeinflussen dadurch die Qualität des angebotenen Unterrichts. Als relevant betrachtete Einflussgrößen werden auf der Seite der angebotsbezogenen Stützsysteme 1) die Architektur des Systems (z. B. Gliederung des Schulsystems, Organisation des Unterrichts wie beispielsweise Jahrgangs- oder Mehrjahrgangsklassen), der Lehrplan, pädagogische Traditionen (z. B. erweiterte Unterrichtsformen im Mathematikunterricht der Schweiz, vgl. Pauli et al., 2010), 2) die Aus- und Weiterbildung der Lehrpersonen und ihre Qualifikation und 3) Merkmale der Einzelschule berücksichtigt.

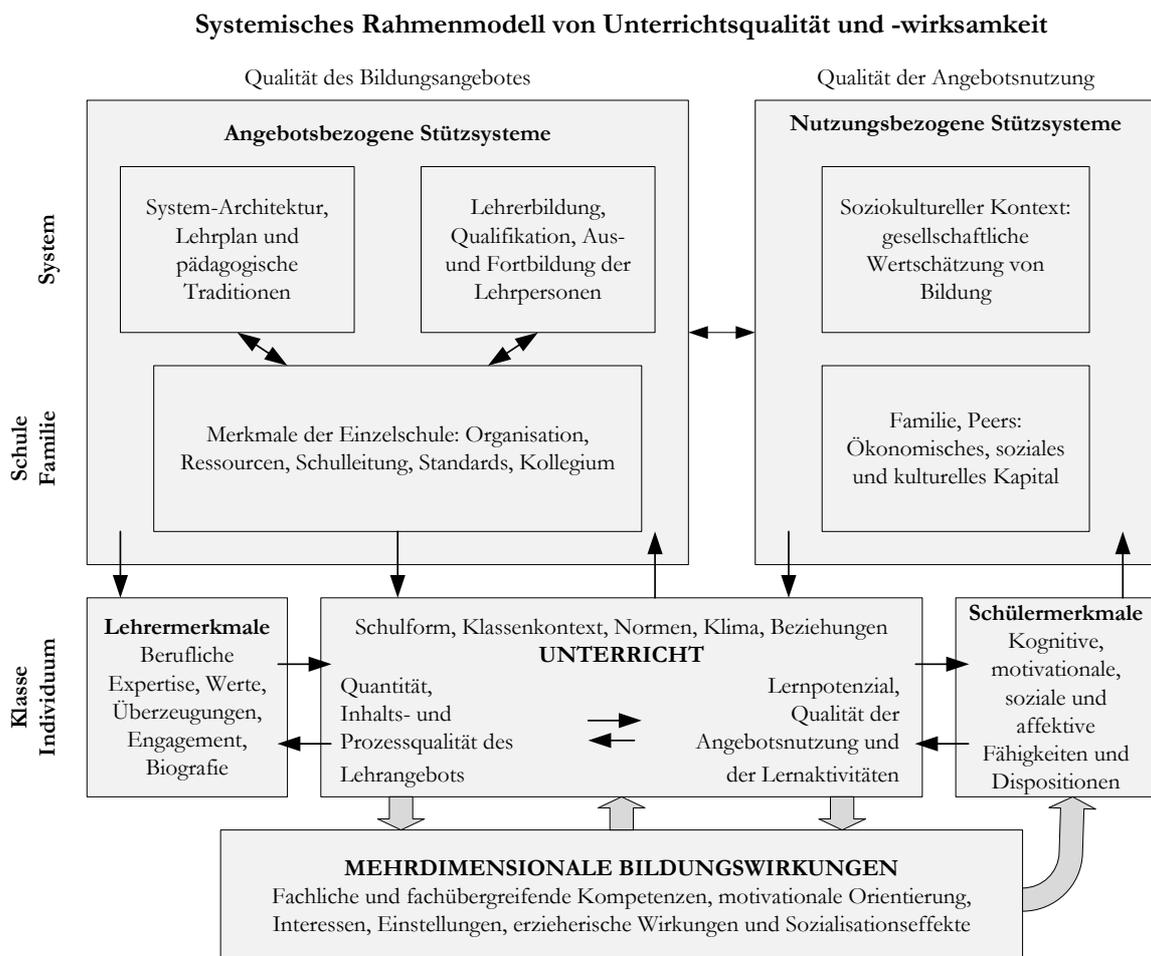


Abbildung 6: Systemisches Rahmenmodell von Unterrichtsqualität und -wirksamkeit (Reusser & Pauli, 2010, S. 18)

Zur Qualifikation, zum Fachwissen und zu relevanten Überzeugungen von Lehrpersonen im internationalen Vergleich liegen zahlreiche Erkenntnisse vor (z. B. Blömeke, Hsieh, Kaiser & Schmidt, 2014; Blömeke et al., 2009; Blömeke & Delaney, 2012; Kunter et al., 2011, 2013; Li & Moschkovich, 2013). Zum Einfluss des Fachwissens oder fachbezogener Überzeugungen auf die Gestaltung des Unterrichts und seine Qualität sind hingegen nur relativ wenige Befunde verfügbar. Relevante fachbezogene Überzeugungen von Lehrpersonen werden eher im Hinblick auf ihren Einfluss auf die Leistungsentwicklung der Lernenden diskutiert (z. B. Staub & Stern, 2002). Untersuchungen, die klar beobachtbare Unterschiede in der Gestaltung von Mathematikunterricht mit unterschiedlichen mathematikbezogenen Überzeugungen von Lehrpersonen in Verbindung bringen, fehlen hingegen weitgehend (siehe auch Studie 5, vgl. Kapitel 4.2).

Verschiedene große Videostudien (z. B. Givvin et al., 2005; Hiebert et al., 2003; Hugener et al., 2009; Kaur et al., 2013; Reusser & Pauli, 2003) konnten auf aufschlussreiche Weise darstellen, wie sehr sich Mathematikunterricht in verschiedenen Ländern und Kulturen unterscheidet. Nicht vollständig empirisch geklärt ist hingegen die Frage, inwiefern auch in räumlich sehr viel engeren Gebieten, beispielsweise innerhalb eines Landes oder auf lokaler Ebene, von systemisch bedingten Unterschieden auszugehen ist. Entsprechende Befunde aus einer breit angelegten schweizerischen Videostudie (Reusser & Pauli, 2003) belegen regional deutlich unterschiedliche Muster für den Mathematikunterricht der Sekundarstufe I der Schweiz. Diese Muster verlaufen allerdings entlang der Sprachregionen der Schweiz (deutschsprachige, französischsprachige und italienischsprachige Landesteile) und dürften somit auch kulturell geprägt sein. Offen ist daher, inwiefern systemisch bedingte Unterschiede auch innerhalb derselben Region – z. B. in verschiedenen Schweizer Kantonen oder in Kantonen, die derselben geografischen Region angehören in unterschiedlich organisierten Schulformen (z. B. Jahrgangsklasse – Mehrjahrgangsklasse) – im beobachteten Mathematikunterricht deutlich werden. Diese Frage hängt eng mit dem in Kapitel 2.4 thematisierten Verhältnis zwischen den Perspektiven der empirischen Bildungsforschung bzw. der Allgemeinen Didaktik und derjenigen der Mathematikdidaktik zusammen und verbindet diese verschieden ausgerichteten Perspektiven exemplarisch.

In den nachfolgenden Unterkapiteln werden Aspekte solcher nach wie vor offenen Fragen im Rahmen von kleineren Studien zum Mathematikunterricht der Primarschule genauer beleuchtet. Exemplarisch wird dies am Beispiel der Struktur der – insbesondere aus schulstrukturellen (Fickermann, Weishaupt & Zedler, 1998), schulpolitischen (Ministerium für Kultus, Jugend und Sport Baden-Württemberg, 2012) oder normativen Gründen (z. B. Monn, 2006; Thurn, 2010b) favorisierten – Mehrjahrgangsklassen bzw. des jahrgangübergreifenden Lernens aufgegriffen und untersucht. Mehrjahrgangsklassen bzw. jahrgangübergreifendes Lernen sind im Schweizer Kanton Thurgau, aber auch in anderen Deutschschweizer Kantonen sowie im süddeutschen Raum stark verbreitet und somit kein singuläres Phänomen. Es ergibt deshalb Sinn, zu fragen, inwiefern sich eine solche Schulstruktur auf das Lehren und Unterrichten von Mathematik und damit letztlich möglicherweise auch auf die Qualität des Mathematikunterrichts auswirken könnte.

Die erste vorzustellende Studie (Studie 8, vgl. Kapitel 5.1) befasst sich mit der Schulorganisation der jahrgangübergreifenden Klassen innerhalb eines Kantons und untersuchte die Gestaltungsmuster von Mathematikunterricht in Mehrjahrgangsklassen. Die zweite Studie (Studie 9, vgl. Kapitel 5.2) erweitert die Frage um diejenige nach Unterschieden und Gemeinsamkeiten in den Gestaltungsmustern von Mathematikunterricht der Primarschule für zwei verschiedene Schulorganisationsformen (Jahrgangs- bzw. Mehrjahrgangsklasse) in zwei unterschiedlichen, aber struktu-

rell vergleichbaren Schweizer Kantonen derselben geografischen Region, während sich die dritte Studie (Studie 10, vgl. Kapitel 5.3) mit Unterschieden in ausgewählten mathematikbezogenen Überzeugungen von Primarlehrpersonen aus Mehrjahrgangsklassen in unterschiedlichen, aber kulturell vergleichbaren Ländern befasst. Diese drei Studien bilden als inhaltlich auf die Schulstruktur bezogene Untersuchungen eine Einheit mit je anderer mathematikdidaktischer Fokussierung und befassen sich mit der übergeordneten Fragestellung des Einflusses von systembezogenen Merkmalen auf die Gestaltung und damit auf die Qualität von Mathematikunterricht.

5.1 Studie 8: Gestaltungselemente von Mathematikunterricht in Mehrjahrgangsklassen

Brunner, E. (2017). Mathematikunterricht in Mehrjahrgangsklassen der Primarschule: Eine Deskription entlang verschiedener Gestaltungselemente und Einschätzungen der Lehrpersonen. *Journal für Mathematik-Didaktik JMD*, 38(1), 57-91.

Ziel der zitierten Studie war eine Deskription des Mathematikunterrichts in Mehrjahrgangsklassen der Primarschulen entlang bestimmter Gestaltungselemente auf der Ebene der Sichtstruktur. In einer Stichprobe von 99 Lehrpersonen, die alle in einer Mehrjahrgangsklasse der Primarschule des Schweizer Kantons Thurgau unterrichten, wurde erfragt, wie die Lehrpersonen ihren Mathematikunterricht unter diesen besonderen strukturellen Rahmenbedingungen bezüglich verschiedener Gestaltungselemente beschreiben. Erfasst wurde zum einen die Häufigkeit des Einsatzes von bestimmten Unterrichts- und Sozialformen im Mathematikunterricht. Zum anderen interessierte, wie oft und in welchen Inhaltsbereichen und während welcher Lernphasen bestimmte Organisationsformen altersdurchmischten Lernens zu Einsatz kommen. Zudem wurden die Einschätzungen zu altersdurchmischem Lernen im Mathematikunterricht erhoben und zu vorhandenen Forschungsbefunden in Beziehung gesetzt. Die Studie griff damit ein aktuelles Thema der Schulentwicklung auf, welche in gewissen Regionen und für bestimmte Schulstufen altersdurchmisches oder jahrgangsübergreifendes Lernen als vorteilhaft propagiert (z. B. Laging, 2003; Thurn, 2010) und sich durch die Implementation jahrgangsübergreifender Klassen eine verstärkte Verbreitung erweiterter Lehr- und Lernformen sowie individualisierter Lernkonzepte erhofft. In der Studie wurde aus mathematikdidaktischer Perspektive nach der Gestaltung des Mathematikunterrichts innerhalb dieser Schulstruktur gefragt. Folgt man den Erwartungen diesbezüglicher Schulentwicklungsbestrebungen, müssten in Mehrjahrgangsklassen sehr oft erweiterte Lehr-Lernformen eingesetzt werden und grundsätzlich jahrgangsübergreifend gelernt werden anstelle von jahrgangsgetreutem Parallelunterricht für die einzelnen Jahrgänge der Abteilung.

Die Ergebnisse der Befragung der 99 Lehrpersonen aus Mehrjahrgangsklassen legen nahe, dass Mathematik auch in Mehrjahrgangsklassen überwiegend jahrgangsgetreut unterrichtet wird – und zwar für sämtliche Inhaltsbereiche („Zahl & Variable“, d. h. Arithmetik, „Form & Raum“, d. h. Geometrie, „Messen, Größen, Daten & Zufall“, d. h. Sachrechnen). Bevorzugt jahrgangsgetreut wurde in der Mehrzahl der Klassen bei der Einführung von neuem mathematischem Stoff gearbeitet, während Übung und Automatisierung vorwiegend in individualisierten Formen und Anwendungssituationen am häufigsten jahrgangsübergreifend organisiert beschrieben wurden.

Die Befragung bezog sich auf Qualitätsmerkmale, die ausschließlich auf der Sichtstrukturebene von Unterricht angesiedelt sind. Allerdings wurden diese Sichtstrukturmerkmale auf den

Fachunterricht bezogen und innerhalb dieses Fachunterrichts weiter aufgeteilt nach Inhaltsbereich und nach Phase und Funktion im Lernprozess. Die Ergebnisse selbst leisten deshalb zwar keine Bestimmung der Qualität des berichteten Mathematikunterrichts, weisen aber darauf hin, dass Schulreformen auf der Ebene von Schulorganisation und Sichtstruktur erstens ihre Wirkung auf den Fachunterricht zumindest teilweise verfehlt zu haben scheinen, indem nicht durch die Implementation einer bestimmten Schulorganisation bestimmte Unterrichtsformen evoziert und gefördert werden können und zweitens zu einer erheblichen Mehrbelastung der Lehrpersonen führen können. Denn wie die Befragung in Bezug auf Letzteres ergab, wurde die Belastung von den Lehrpersonen als hoch eingeschätzt und das Unterrichten von Mathematik wurde innerhalb der strukturellen Rahmenbedingungen einer Mehrjahrgangsklasse als sehr anspruchsvoll empfunden. Dies lässt vermuten, dass die Qualität des angebotenen Mathematikunterrichts längerfristig betrachtet leiden könnte, weil die Lehrpersonen innerhalb komplexer Rahmenbedingungen den Hauptfokus auf das Bewältigen der komplexen Struktur legen. Die Komplexität altersdurchmischter Klassen verlangt zum einen eine differenziertere Klassenführung und einen größeren Überblick über inhaltlich und sozial differente Programme als dies in einer Jahrgangsklasse mit Fokus auf ein reguläres inhaltliches Programm der Fall ist und führt zum anderen dazu, dass Lehrpersonen für verschiedene Lerngruppen und mehrere Jahrgänge gleichzeitig mehrere inhaltliche Parallelprogramme vorbereiten und verfügbar haben müssen. Angesichts der Tatsache, dass die Vorbereitungszeit der Lehrpersonen beschränkt ist, dürfte damit weniger Zeit für fachlich solides Planen und Durchführen von Mathematikunterricht für eine bestimmte Lerngruppe zu einem bestimmten Inhalt bleiben. Die Studie zeigt deshalb, dass Ansätze der Schulentwicklung bezüglich ihrer Wirkung nicht nur auf der allgemeinen generischen Ebene von Unterricht untersucht werden sollten, sondern zunehmend auch spezifisch für einen bestimmten Fachunterricht und aus fachspezifischer Sicht.

5.2 Studie 9: Gestaltungsmuster von Mathematikunterricht im Kantonsvergleich

Brunner, E. & Imhof, A. (2017). Gestaltungsmuster des Mathematikunterrichts in Jahrgangs- und Mehrjahrgangsklassen – Zwei Kantone im Vergleich. *Schweizerische Zeitschrift für Bildungswissenschaften SZBW*, 39(2), 353-374.

Diese Studie befasste sich wie bereits Studie 8 mit der Sichtstruktur von Unterricht und den bei Primarlehrpersonen erfragten Gestaltungsmustern ihres Mathematikunterrichts, nahm aber einen Vergleich bezüglich dreier Merkmale auf der Ebene des angebotsbezogenen Stützsystems vor. Es waren dies die Schulorganisation (Jahrgangs- bzw. Mehrjahrgangsklasse), die Region mit zwei unterschiedlichen Kantonen⁸ mit ihren pädagogischen Traditionen sowie der Ausbildungsstand

⁸ Die beiden ausgewählten Kantone Thurgau und Graubünden sind beide eher ländlich geprägt und gehören zu derselben geografischen Region der Ostschweiz. Beide Kantone bilden ihre Primarlehrpersonen an einer kantons-eigenen Pädagogischen Hochschule für den Unterricht in sieben bzw. acht Fächern aus. In beiden Kantonen ist die Verbreitung von Mehrjahrgangsklassen relativ hoch. Allerdings beruht diese Schulstruktur auf unterschiedlichen Voraussetzungen: Im Kanton Thurgau ist sie zum einen Ausdruck von aktiven Schulentwicklungsbemühungen und deshalb das Ergebnis einer offensiven Einführung mit finanziellen Anreizen, im Kanton Graubünden hingegen spiegelt sie primär die lokalen Gegebenheiten und dient beispielsweise dem Erhalt vieler kleiner Dorfschulen.

der Lehrpersonen (erfahrene Lehrpersonen vs. Berufseinsteigende). Geprüft wird in der Studie in den beiden Ostschweizer Kantonen Thurgau und Graubünden erstens, inwiefern sich der selbst berichtete Einsatz bestimmter Unterrichts- und Sozialformen im Mathematikunterricht in Jahrgangs- und Mehrjahrgangsklassen der Primarschule unterscheidet. Zweitens wird erfasst, inwiefern sich in den berichteten Gestaltungsmustern Unterschiede zwischen erfahrenen Lehrpersonen und Lehrpersonen, die am Beginn ihrer professionellen Laufbahn stehen, zeigen. Drittens wird für die beiden Kantone nach Unterschieden in den berichteten Gestaltungsmustern gesucht. Methodisch wurden die Daten aus der Befragung zunächst mittels Hauptkomponenten-Faktorenanalyse analysiert. Dadurch konnte die Komplexität der zahlreichen unterschiedlichen Gestaltungselemente reduziert werden. Es ließ sich eine sechsfaktorielle Lösung für die erfragten Unterrichtsformen beschreiben, die als sechs unterschiedliche Profile interpretiert werden konnten: traditionelle Unterrichtsformen, innovative Unterrichtsformen, diskursive Unterrichtsformen, offene Lernformen mit Wahlmöglichkeiten, Gruppenarbeit, Stillarbeit.

Die Ergebnisse zur selbst berichteten Gestaltung von Mathematikunterricht zeigen, dass die einzelnen Unterrichtsformen unterschiedlich stark berücksichtigt wurden und sich der Mathematikunterricht mehrheitlich als Einzel-/Stillarbeit, gefolgt von Tandemarbeit und Phasen des Lehrgesprächs, beschreiben lässt. Dieser Befund gilt mit Nuancen für beide Kantone sowie für beide Gruppen von Lehrpersonen (erfahrene Lehrpersonen vs. Berufseinsteigende). Generell ließen sich weit mehr Gemeinsamkeiten als Unterschiede finden. Die wenigen Unterschiede bezogen sich 1) auf den Einsatz traditioneller Unterrichtsformen, der im Kanton Thurgau für Jahrgangsklassen häufiger berichtet wurde als für Mehrjahrgangsklassen, im Kanton Graubünden hingegen in Mehrjahrgangsklassen stärker verbreitet ist als in Jahrgangsklassen, 2) auf innovative Unterrichtsformen, die im Kanton Thurgau verbreiteter zu sein scheinen als im Kanton Graubünden, und 3) auf Gruppenarbeiten, die im Kanton Graubünden häufiger von Novizinnen und Novizen genutzt werden als von Expertinnen und Experten, während sich dies im Kanton Thurgau umgekehrt verhält.

Insgesamt lässt sich feststellen, dass der Mathematikunterricht in einer bestimmten Region mehrheitlich Gestaltungsmustern folgt, die von Schulstruktur, Berufserfahrung der Lehrperson und dem Schulkanton unabhängig sind. Demnach hätten Unterschiede in angebotsbezogenen Stützsyste men nur einen geringen Einfluss auf die berichteten Sichtstrukturmerkmale von Mathematikunterricht. Unklar ist, inwiefern Personenmerkmale der Lehrpersonen in unterschiedlichen angebotsbezogenen Stützsyste men eine Rolle spielen. Damit befasst sich die nächste Studie.

5.3 Studie 10: Überzeugungen von Lehrpersonen zu jahrgangsgemischtem Mathematikunterricht im Ländervergleich

Brunner, E., Gasteiger, H., Lampart, J. & Schreieder, K. (2019). Mathematikunterricht in jahrgangsübergreifenden Klassen der Grundschule in der Schweiz und in Deutschland: Eine vergleichende Studie. *Schweizerische Zeitschrift für Bildungswissenschaften SZBW*, 41(1), 160-176.

Nicht nur die angebotsbezogenen Stützsyste me und die kulturell geprägten Rahmenbedingungen des Unterrichts beeinflussen die Gestaltungsmerkmale. Auch Überzeugungen von Lehrpersonen als Teil ihrer professionellen Kompetenz (Kunter et al., 2011) wird diesbezüglich ein erheblicher Einfluss zugeschrieben, wenngleich es unklar ist, inwiefern Überzeugungen tatsächlich auf die

Lehr-Lern-Leistung wirken und entsprechende Befunde disparat ausfallen (zusammenfassend Reusser & Pauli, 2014). In der nachfolgend vorgestellten Studie wurden die im Rahmen des Kantonsvergleichs (Studie 9, vgl. Kapitel 5.2) erarbeiteten Gestaltungsprofile für den Mathematikunterricht in Jahrgangs- und Mehrjahrgangsklassen der Primarschule mit den Überzeugungen der Lehrpersonen zur Wirkung von altersdurchmischem, jahrgangübergreifendem Lernen verbunden. Einen höheren Grad an Differenzierung im Unterricht und damit ein bestimmtes Bündel von Gestaltungsmustern konnten Carle und Berthold (2004) bei Lehrpersonen nachweisen, die jahrgangübergreifende Klassen *bevorzugen*. Dies legt die Vermutung nahe, dass weniger die Schulstruktur (jahrgangsbezogen bzw. jahrgangübergreifend) als vielmehr die Überzeugungen zum jahrgangübergreifenden Lernen von Lehrpersonen einen bedeutsamen Einfluss auf die Gestaltung des Unterrichts haben (Mason & Burns, 1995).

Das Ziel der vorliegenden Studie bestand darin, an zwei Stichproben aus Schweizer Primarschulen des Kantons Thurgau und deutschen Grundschulen in Bayern ($N_{CH} = 99$; $N_D = 78$), in denen jahrgangübergreifend unterrichtet wird, ländervergleichend Gestaltungsmuster in Verbindung mit Überzeugungen der Lehrpersonen zu beschreiben. Im Fokus standen kulturell differenziertere Strukturen als dies in Studie 9 (vgl. Kapitel 5.2) bei den beiden Schweizer Kantonen aus derselben geografischen Region der Fall war. Ein zentraler Unterschied zwischen den beiden Ländern besteht darin, dass im Schweizer Kanton Thurgau großflächige Umstellungen von Jahrgangs- zu Mehrjahrgangsklassen erfolgt sind, während dies in Bayern nicht in diesem Ausmaß der Fall ist. Daher unterscheiden sich die Rahmenbedingungen und somit möglicherweise auch die persönlichen Überzeugungen zu jahrgangübergreifendem Lernen in den beiden ausgewählten Regionen der beiden Länder, während die in beiden Ländern eingeführten Bildungsstandards auf ähnlichen Grundlagen beruhen. Ein weiterer Unterschied liegt in der Ausbildungsstruktur: In Deutschland erfolgt die Ausbildung zur Grundschullehrperson durch ein vertieftes Studium in drei Fächern, u. a. in Mathematik. In der Schweiz hingegen ist die Ausbildung zur Primarlehrperson als Lehrbefähigung für mindestens sieben Fächer angelegt und zielt stärker auf Generalistinnen und Generalisten denn auf Fachlehrpersonen ab.

Im Rahmen der schriftlichen länderübergreifenden Befragung wurden Gestaltungsmerkmale des Mathematikunterrichts sowie Überzeugungen zum jahrgangübergreifenden Lernen erhoben. Die Ergebnisse zeigen – im Gegensatz zum Vergleich zwischen zwei Schweizer Kantonen aus derselben Region in Studie 9 (vgl. Kapitel 5.2) – deutlich mehr Unterschiede als Gemeinsamkeiten hinsichtlich der Gestaltung von Mathematikunterricht in jahrgangübergreifenden Klassen der Grundschule sowie bezüglich der Überzeugungen zum jahrgangübergreifenden Lernen. Dies lässt die Vermutung zu, dass sowohl Überzeugungen als auch Gestaltungsmuster kulturell unterschiedlich geprägt sind und Teil kulturell unterschiedlich ausgestalteter Profile sind. Für fünf der sechs unterschiedlichen Gestaltungsmuster von Mathematikunterricht (vgl. Kapitel 5.2) zeigen sich Länderunterschiede in der Häufigkeit ihres Einsatzes. Die Lehrpersonen der Schweizer Stichprobe geben an, die Muster „Direkte Instruktion“ und „Still-Einzelarbeit“ deutlich häufiger einzusetzen, als dies die deutschen Grundschullehrkräfte berichten. Letztere beschrieben hingegen einen häufigeren Einsatz der Muster „Inhaltlich offen“, „Organisatorisch offen“ und „Diskursive Unterrichtsformen“. Kein statistisch signifikanter Länderunterschied besteht bei der berichteten Häufigkeit des Einsatzes des Gestaltungsmusters „Soziale Unterrichtsformen“.

Diese Länderunterschiede, die sich bei den selbst berichteten Gestaltungsmustern zeigen, setzen sich auch in den Überzeugungen zum jahrgangübergreifenden Lernen fort. Die deutschen

Grundschullehrpersonen schätzten jahrgangsübergreifendes Lernen bezüglich seiner Wirkung auf die Leistungen sowie auf die sozialen Kompetenzen der Lernenden deutlich positiver ein als die Lehrpersonen der Schweizer Stichprobe und sie waren zudem stärker davon überzeugt, dass jahrgangsübergreifendes Lernen besser sei für das Lernen der Kinder als der Unterricht in Jahrgangsklassen und dass diese Form des Lernens in erster Linie eine gute Organisation verlange. Übereinstimmend als höher im Vergleich zum Unterricht in Jahrgangsklassen schätzten alle Befragten die Herausforderung jahrgangsübergreifenden Lernens für die Lehrperson sowie die benötigte Vorbereitungszeit ein. Einig waren sich die Lehrpersonen beider Länder auch bezüglich der Notwendigkeit geeigneter Aufgabenstellungen für jahrgangsübergreifendes Lernen. Hingegen wurde weder ein höheres Potenzial für Kooperationen unter Lehrpersonen noch eine gesteigerte Interessantheit von jahrgangsübergreifendem Lernen für Lehrpersonen angenommen.

Des Weiteren bestehen Zusammenhänge zwischen einzelnen Überzeugungen und bestimmten Gestaltungsprofilen. So sind Lehrpersonen, die häufiger mit dem Muster „Organisatorisch offen“ unterrichten, auch stärker davon überzeugt, dass jahrgangsübergreifendes Lernen besser sei als jahrgangsgetrenntes und dass Ersteres zu höheren sozialen Kompetenzen der Lernenden führe und für die Lehrpersonen interessanter sei. Lehrpersonen, die vermehrt das Muster „Diskursive Formen“ einsetzen, stimmen eher der Aussage zu, dass jahrgangsübergreifendes Lernen zu höheren Leistungen der Lernenden, insbesondere der leistungsschwachen, führe und dass es für Lehrpersonen interessanter sei. Ähnliches gilt auch für diejenigen Lehrpersonen, die häufiger das Muster „Soziale Unterrichtsformen“ berichteten. Je häufiger Lehrpersonen hingegen Unterrichtsformen des Musters „Still-/Einzelarbeit“ berichteten, desto stärker sind sie davon überzeugt, dass jahrgangsübergreifendes Lernen das genaue Einhalten von Vereinbarungen erfordere, und desto geringer ausgeprägt ist ihre Überzeugung, dass jahrgangsübergreifendes Lernen besser sei oder bei Leistungsschwachen zu höheren Leistungen führe als jahrgangsgetrenntes Lernen.

Die zahlreichen Unterschiede lassen allerdings unterschiedliche Interpretationen zu und verlangen nach weiteren Analysen: Sie könnten tatsächlich Ausdruck von länderspezifisch unterschiedlichen strukturellen Voraussetzungen und damit einhergehenden unterschiedlichen Überzeugungen sein. Es wäre aber auch plausibel, anzunehmen, dass diese Unterschiede durch die systematisch unterschiedliche professionelle Erfahrung sowie das Alter der Lehrpersonen der beiden Teilstichproben erklärt werden könnten. Die Lehrpersonen der deutschen Stichprobe waren im Mittel nicht nur statistisch signifikant älter ($M_{CH} = 31.17$; $SD = 11.52$; $M_D = 43.26$; $SD = 9.98$; $t = -7.35$; $df = 175$; $p = .000$), sondern wiesen auch eine deutlich grössere Berufserfahrung auf ($M_{CH} = 2.57$; $SD = 1.88$; $M_D = 4.74$; $SD = 1.17$; $t = -9.45$; $df = 166.59$, inhom. Varianzen korr.; $p = .000$). Die Schweizer Lehrpersonen wurden anlässlich je einer für die entsprechende Gruppe obligatorischen Fortbildung befragt. Es handelt sich somit um eine Vollerhebung, während die deutsche Stichprobe nicht repräsentativ ist und sich auf 78 Lehrpersonen (von insgesamt 750 angeschriebenen Lehrpersonen) bezieht, die sich freiwillig an der Online-Befragung beteiligten.

Deutlich macht die Studie gleichwohl, dass weder Gestaltungsmuster noch Überzeugungen von Lehrpersonen zum Mathematikunterricht kontextfrei erforscht werden können, sondern sowohl mit (kulturell unterschiedlichen) angebotsbezogenen Stützsystemen als auch mit Merkmalen der Lehrpersonen zusammenhängen dürften. Während es zu Länderunterschieden eine reiche Forschungstradition in der Mathematikdidaktik gibt (z. B. Blömeke et al., 2014; Kaur et al., 2013;

Leung, 2006), fehlt eine solche bislang noch für stärker regional geprägte Muster und regional gefärbte kulturelle Kontexte.

5.4 Diskussion

Obwohl Mathematik als kulturumspannende Domäne betrachtet wird, deren Wurzeln auf Axiomen und bewiesenen Sätzen (Heintz, 2000) der Ideengeschichte der Mathematik (Winter, 1991) beruhen – was sich weltweit in weitgehend übereinstimmenden Curricula (Common Core State Standards Initiative, 2012; D-EDK., 2016; KMK, 2005; NCTM, 2000) und gemeinsamen Frameworks (Mullis, 2012; Mullis, Martin, Ruddock, O’Sullivan & Preuschoff, 2009; OECD, 2013) ausdrückt –, zeigt es sich, dass auch Mathematikunterricht nicht kontextfrei betrachtet werden kann. Mathematikunterricht spielt sich auf der Mikroebene zwar in der Interaktion zwischen der Lehrperson und den Schülerinnen und Schülern einer Klasse ab. Gleichwohl dürfen dabei auch die systembezogenen Faktoren, die den Spielraum der Gestaltung von Mathematikunterricht begrenzen, nicht außer Acht gelassen werden. Denn wie die in den vorhergehenden Kapiteln vorgestellten, auf einen eng gefassten, einheitlichen Bereich bezogenen drei Studien verdeutlichen, dass sowohl die Schulorganisationsform (Studien 8, 9 und 10) als auch die geografische Region bzw. das Land (Studie 9 und 10) eine bedeutsame Rolle spielen können und den Rahmen für das pädagogisch und fachdidaktisch Handeln und Unterrichten von Lehrpersonen innerhalb dieses Kontexts vorgeben. Unterrichtsqualität wird somit nicht nur von der Lehrperson zusammen mit ihrer Klasse *erzeugt*, sondern spiegelt auch gesellschaftliche und systembedingte Rahmenbedingungen, innerhalb derer ebendieser Unterricht stattfindet. Insofern ist die kulturumspannende Domäne Mathematik keineswegs mit einem für alle Kulturen einheitlich zu präsupponierenden Konzept von Unterrichtsqualität oder mit einem universellen Maßstab für Qualität von Mathematikunterricht gleichzusetzen. Es ist sehr viel mehr davon auszugehen, dass unterschiedliche – gesellschaftlich unterschiedlich gewichtete – Ausprägungen von Unterrichtsqualität bestehen, die durch je unterschiedliche systembedingte Merkmale beschränkt werden, was bei der regelmäßigen Veröffentlichung von Ergebnissen großer internationaler Leistungsmessungsstudien wie beispielsweise PISA allerdings sehr oft unberücksichtigt bleibt. Dieses Fazit legt bereits das in Kapitel 5 einführend dargelegte Angebot-Nutzungs-Modell (Reusser & Pauli, 2010) von Unterricht nahe. Die drei vorgestellten Studien nehmen dessen theoretische Grundannahmen auf und verankern diese exemplarisch in der Empirie.

6 Übergreifende Gesamtdiskussion

6.1 Kurze Zusammenfassung zentraler Ergebnisse

Ausgehend von der Feststellung eines Primats der Erziehungswissenschaften und der allgemeinen empirischen Bildungsforschung wurde anhand von fünf Aufgabenfeldern aufgezeigt, wo fachdidaktische Forschung neue Erkenntnisse generieren kann und wo deren Einbezug einen Mehrwert liefert (vgl. Kapitel 2.1–2.4). Es handelt sich dabei um die folgenden Aufgabenfelder: 1) das Aufgabenfeld „Schulfach und seine Inhalte und Ziele“, 2) das Aufgabenfeld „Fachliche Lehr-Lern-Prozesse“, 3) das Aufgabenfeld „Unterrichtspraxis mit Lehrmaterialien und Lernumgebungen“, 4) das Aufgabenfeld „Gesellschaftliche Anforderungen“ und schließlich 5) das Aufgabenfeld „Inter- und Transdisziplinarität“. Die Hauptergebnisse der vorliegenden Arbeit, in der zwei Aufsätze und zehn unterschiedliche Studien nach thematischen Schwerpunkten strukturiert miteinander verbunden wurden, werden nachfolgend kurz entlang der leitenden Fragestellung zusammenfassend dargestellt: Inwiefern kann fachdidaktische Forschung – hier die mathematikdidaktische Forschung – einen Mehrwert für die Erforschung von Lehr- und Lern-Prozessen bieten?

Thematischer Fokus „Unterrichtsqualität und der fachdidaktische Blick“

Im Zusammenhang mit der Frage der Unterrichtsqualität (vgl. Kapitel 3) wurde gezeigt, wie die fachdidaktische Perspektive die genuin erziehungswissenschaftliche oder generische Sicht (vgl. Kapitel 2.3–2.4) zu erweitern vermag und inwiefern der Einbezug einer fachdidaktischen Perspektive sogar unabdingbar ist, wenn bestimmte Kompetenz- und Inhaltsbereiche vertieft und präzise untersucht und beschrieben werden sollen (Aufgabenfeld 1), um davon ausgehend Qualitätsanforderungen definieren zu können. Qualität im Fachunterricht manifestiert sich u. a. auch darin, wie weit es im Unterrichtsgeschehen gelingt, die zentralen fachlichen Prozesse verknüpft mit inhaltlichen Aspekten des Lerngegenstandes zu adressieren (Aufgabenfeld 2, Aufgabenfeld 3), um den gesellschaftlichen Anforderungen aus fachlicher Sicht gerecht zu werden (Aufgabenfeld 4). Am Beispiel eines Aufsatzes einer interdisziplinär zusammengesetzten Gruppe von Autorinnen und Autoren (vgl. Kapitel 2.3) sowie einer empirischen Studie (vgl. Kapitel 3.3), die ebenfalls von einem interdisziplinär zusammengesetzten Team durchgeführt worden war, wurde des Weiteren das Aufgabenfeld der Inter- und Transdisziplinarität (Aufgabenfeld 5) konkretisiert. Im Kontext der Studie zur Qualität von unterschiedlichen Unterrichtsmethoden wurde insbesondere deutlich, dass eine fachdidaktische Perspektive für den Bezug von Unterrichtsmethoden auf bestimmte fachliche Inhalte und erlebte Emotionen sowie für die Interpretation der gewonnenen Erkenntnis gewinnbringend sein kann. Dass empirisch gezeigt werden konnte, dass es sich bei Skalen zur Erfassung von fachdidaktischen bzw. generischen Qualitätsmerkmalen um unterschiedliche Analyseinstrumente handelt (vgl. Kapitel 3.2) und dass ein gezieltes fachspezifisch-pädagogisches Coaching primär auf eine Entwicklung fachdidaktischer Qualität abzielt und diese sich auch – in Ansätzen – nachweisen lässt, verweist ebenfalls auf die Notwendigkeit der Berücksichtigung einer fachdidaktischen Perspektive und den Beitrag, den diese für das Forschungsfeld und letztlich auch für die Praxis zu leisten vermag.

Der Vorschlag für ein generisches Modell von Unterrichtsqualität sollte aufzeigen, in welchen Bereichen durch den Einbezug der Fachdidaktik eine Differenz in der Wahrnehmung entstehen kann und wie diese im Rahmen einer gemeinsamen Theoriebildung und einer Erweiterung

bzw. Weiterentwicklung bereits bestehender Instrumente aufgelöst werden könnte. Dazu soll das vorgestellte hierarchische Modell von fachlicher Unterrichtsqualität (vgl. Kapitel 3.4) einen konkreten Beitrag leisten, indem es Erkenntnisse aus der empirischen Lehr-Lern-Forschung aufgreift und diese auf den Mathematikunterricht bezieht.

Thematischer Fokus „Qualitätsfragen am Beispiel mathematischen Begründens, Argumentierens und Beweisens“

Die Erörterung von Qualitätsfragen am Beispiel bestimmter mathematischer Kompetenzen wie dem mathematischen Argumentieren und Beweisen (vgl. Kapitel 4) sollte deutlich machen, dass auf der Grundlage einer konsequenten fachdidaktischen Sicht weitere und teilweise andere Qualitätsmerkmale definiert und in der Folge operationalisiert werden können, als dies aus generischer Perspektive der Fall ist. Während der Fokus auf Gesprächsmuster noch nicht per se fachspezifisch ist, ist es deren Bezug auf bestimmte prototypische Vorgehensweisen beim Beweisen, das heißt auf bestimmte Beweistypen, die je unterschiedliche Voraussetzungen an die Gestaltung des didaktischen Gesprächs stellen und die darüber hinaus auch unterschiedliches Potenzial für die Partizipation der Lernenden am fachlichen Diskurs bieten (vgl. Kapitel 4.1). Hier verbinden sich Aspekte von Tiefen- und Oberflächenstrukturmerkmalen (Reusser, 2005) und es wird deutlich, dass diese sich gegenseitig bedingen (vgl. Kapitel 1).

Auch wenn nach Präferenzen von Lehrpersonen zu einem bestimmten Vorgehen bei der Bearbeitung von Beweisaufgaben gefragt wird (vgl. Kapitel 4.2), handelt es sich um Forschung aus einer konsequent fachspezifischen und sogar kompetenzspezifischen Perspektive. Gerade der Befund aus Studie 6 (vgl. Kapitel 4.2), wonach sich Lehrpersonen, die einen bestimmten Beweistyp mit ihren Klassen umsetzen, von Lehrpersonen unterscheiden, deren Beweissequenz entlang eines anderen Beweistyps erfolgt, verweist auf den Erkenntnisgewinn von Forschungsarbeiten, die fachdidaktisch ausgerichtet sind. Fachlehrpersonen orientieren sich im Gegensatz zu Allround-Lehrpersonen aufgrund ihrer fokussierten Ausbildung stärker an fachlichen Aspekten. Diese Fachperspektive wird den vorliegenden Befunden zufolge im Unterricht empirisch nachweisbar handlungsleitend und mündet in eine beobachtbare Unterrichtspraxis, die sich von derjenigen von Allround-Lehrpersonen unterscheidet.

Die beiden Studien zum frühen mathematischen Argumentieren (vgl. Kapitel 4.3–4.4) präzisieren zum einen eine bedeutsame mathematische Kompetenz für eine bestimmte Altersstufe und legten zum anderen prototypische didaktische Situationen vor, entlang derer diese Kompetenz im Kindergarten altersgerecht und in Anlehnung an die übliche Kindergartendidaktik und -praxis gefördert werden kann. Der Einblick in die Videos der vorgestellten Studie (vgl. Kapitel 4.4) sollte aufzeigen, dass es den teilnehmenden Lehrpersonen des Kindergartens gelungen war, mit den fünf- und sechsjährigen Kindern an anspruchsvollen mathematischen Kompetenzen zu arbeiten und die Kinder dabei auch als Produzierende von Argumenten bzw. von Teilen davon aktiv am Geschehen partizipieren zu lassen. Beide Studien zum Vorschulbereich belegen den Mehrwert einer fachdidaktischen Sicht zum einen durch den engen Fokus auf eine fachliche Kompetenz und zum anderen durch deren Verbindung mit gängiger Kindergartendidaktik und -praxis.

In der Diskussion zum Themenbereich (vgl. Kapitel 4.5) wurde zudem ein erster Entwurf eines Konzepts zur Beschreibung und Erfassung von fach-, inhalts- und kompetenzspezifischer Unterrichtsqualität vorgestellt, das die bisherigen Arbeiten und Überlegungen in einem operationalisierten Analyseraster zur Bestimmung der Qualität von Beweissequenzen zusammenfasst.

Inwiefern sich dieser Entwurf empirisch einsetzen lässt und ob sich mithilfe des Rasters Qualitätsmerkmale reliabel beschreiben lassen, müssen empirische Untersuchungen jedoch erst noch zeigen.

Der Mehrwert fachdidaktischer Forschung lässt sich beim Themenbereich des mathematischen Argumentierens und Beweisens inhaltlich wie folgt rekapitulieren: Es geht um eine fach-, kompetenz- und inhaltsfokussierte Sicht auf das Schulfach, seine Inhalte und Ziele (Aufgabenfeld 1), auf zentrale fachliche Lehr-Lern-Prozesse (Aufgabenfeld 2), auf die Unterrichtspraxis (Aufgabenfeld 3) sowie auf die gesellschaftlichen Anforderungen (Aufgabenfeld 4). Im Zusammenhang mit Aufgabenfeld 5, der Inter- und Transdisziplinarität, stand in diesem Zusammenhang weniger die Interdisziplinarität im Vordergrund als vielmehr die Transdisziplinarität, da der Prozess des mathematischen Argumentierens, Begründens und Beweisen über verschiedene Alters- und Bildungsstufen hinweg untersucht worden war. Bei den diesbezüglich involvierten anderen Disziplinen handelte es sich um die Entwicklungspsychologie und die Wissenserwerbsforschung.

Thematischer Fokus „Qualitätsfragen am Beispiel des Mathematikunterrichts in unterschiedlichen Kontexten“

In der Regel geht es bei der Auseinandersetzung mit Qualitätsfragen zu Mathematikunterricht in unterschiedlichen Kontexten in erster Linie um eine Fokussierung von Oberflächenstrukturmerkmalen (vgl. Kapitel 1), wobei die leitende Fragestellung lautet, inwiefern sich eine bestimmte Schul- oder Unterrichtsstruktur – in Kapitel 5 am Beispiel von jahrgangsübergreifenden bzw. jahrgangsgemischten Klassen illustriert – auf die Gestaltung von Mathematikunterricht auswirkt. Aber sowohl in Studie 8 als auch in Studie 9 (vgl. Kapitel 5.1–5.2) stellte sich heraus, dass ein Fokus auf die Oberflächenstruktur zwar erhellend sein kann, insgesamt aber nur wenig systematische Varianz nachzuweisen vermag. Wird hingegen – wie dies in Studie 10 (vgl. Kapitel 5.3) der Fall war – davon ausgegangen, dass auch Merkmale und Überzeugungen von Lehrpersonen eine Rolle spielen, dann lässt sich zeigen, dass diese ebenfalls einen nachweisbaren Einfluss auf die Umsetzung von bestimmten Oberflächenstrukturmerkmalen von Unterricht haben können. Diese drei thematisch eng miteinander verbundenen Studien beleuchteten insbesondere Aufgabenfeld 4 der fachdidaktischen Forschung (vgl. Kapitel 1), das heißt dasjenige der gesellschaftlichen Anforderungen an Unterricht und Bildung. Wenn es einer gesellschaftlich getragenen Wertvorstellung entspricht, Klassen nicht entlang von Jahrgängen, sondern jahrgangsübergreifend zu organisieren, so sollte diese Vorgabe hinsichtlich ihrer Wirksamkeit und Auswirkungen auch empirisch überprüft werden. Die Ergebnisse der vorgestellten Studien – insbesondere von Studie 8 und Studie 9 – weisen diesbezüglich deutlich auf Grenzen von gesellschaftlich und/oder bildungspolitisch überhöhten Vorstellungen zur Wirksamkeit von Schulreformen auf ausschließlich schulstruktureller Ebene hin. Denn es konnte kein grundsätzlich anderes Gestaltungsmuster von Mathematikunterricht für eine der beiden Unterrichtsstrukturen nachgewiesen werden. Dieser Befund mag – gesellschaftlich gesehen – nicht mit den allgemeinen Erwartungen korrespondieren, da er aufzeigt, dass die Unterrichtsgestaltung mit strukturellen Schulreformen allein noch nicht notwendigerweise verändert werden kann. Zugleich handelt es sich dabei – aus fachdidaktischer Sicht – aber auch um einen positiv zu wertenden Befund, da er darauf hindeutet, dass Lehrpersonen innerhalb des vorgegebenen Rahmens der Unterrichts- und Schulstruktur und davon relativ unabhängig versuchen, Mathematikunterricht durchzuführen. Dass die befragten Lehrpersonen im Zusammenhang mit Unterricht in jahrgangsübergreifenden Klassen von einem deutlichen Mehraufwand und höherer Belastung berichtet hatten, wie dies die Ergebnisse von

Studie 10 zeigen (vgl. Kapitel 5.3), muss gerade im Hinblick auf Fragen der Unterrichtsqualität sorgfältig berücksichtigt werden.

Nicht beleuchtet wurde in den drei vorliegenden Studien die Auswirkung einer bestimmten Schul- bzw. Organisationsstruktur auf das Lernen der Schülerinnen und Schüler und deren Leistungsentwicklung. Diesbezügliche empirische Untersuchungen sind zwingend notwendig und es dürfte ebenfalls zu den Aufgabenfeldern fachdidaktischer Forschung gehören, diese Notwendigkeit auch in einem bildungspolitischen und gesellschaftlichen Kontext herauszustreichen. Gerade im Zusammenhang mit den Ergebnissen großer Leistungsmessungsstudien wie den aktuellen PISA-Befunden zeigt sich, dass sich Leistungen früherer Kohorten nicht zwingend auf demselben Niveau fortschreiben (Hammer et al., 2016; OECD, 2014, 2016). Deshalb könnte der Einbezug von Kontext- sowie Lehrpersonenvariablen sinnvoll sein, um solche Unterschiede vertiefter analysieren zu können. Fruchtbar könnten zudem Studiendesigns sein, die stark differente bzw. ähnliche Kontexte gezielt zur Kontrastierung bestimmter Merkmale nutzen, wie dies in einer eben veröffentlichten Arbeit versucht wird (Gasteiger, Brunner & Chen, 2020), in der in einem ersten Schritt nach möglichen Gemeinsamkeiten und Unterschieden in den Rahmenbedingungen von Kindergartenlehrpersonen in Taiwan, Deutschland und der Schweiz gefragt wird, mit der Zielsetzung zu untersuchen, ob kulturelle oder strukturelle Unterschiede einen stärkeren Einfluss haben auf selbst berichtete Praxis früher mathematischen Bildungsaktivitäten im Kindergarten.

Im Zusammenhang mit den ersten drei Aufgabenfeldern fachdidaktischer Forschung (vgl. Kapitel 2.3, siehe auch oben), – das heißt der Bestimmung und der Beschreibung des Schulfachs mit seinen Inhalten und Zielen, der Untersuchung von fachlichen Lehr-Lern-Prozessen und der Erforschung der Unterrichtspraxis mit Lehrmaterialien und Lernumgebungen – führt der Einbezug einer konsequent fachdidaktischen Sicht auch im Themenfeld der Kontextvariablen von Unterrichtsqualität zu einem deutlichen Mehrwert: Ein bestimmter Kontext kann sich potenziell erst aus einer fachspezifischen Sicht heraus als besondere Herausforderung für das Unterrichten des betreffenden Fachs und damit als mögliche Einflussgröße für die Unterrichtsqualität erweisen. Die fachdidaktische Sicht kann beispielsweise vermuten und empirisch gestützt aufzeigen, dass die spezielle Organisationsstruktur von jahrgangsübergreifenden Klassen mit der relativ strengen Hierarchie, in der die fachlichen Inhalte und Grundvorstellungen strukturiert sind, in Konflikt geraten könnte oder dass vollständig individualisiertes Mathematiklernen ohne instruktionale Sequenzen bzw. ein Verschieben von Themen und Inhalten im Sinne eines Lernens am gemeinsamen Gegenstand (Feuser, 2013) zu fachlichen Verstehensschwierigkeiten führen können, da für das Verstehen eines bestimmten fachlichen Konzepts entsprechende Vorkenntnisse notwendig sind. Wer Brüche verstehen möchte, muss in der Lage sein, auf ein Relationszahlkonzept zurückzugreifen, die Teil-Ganzes-Beziehung verstanden haben und Brüche nicht primär als neue Zahlen, sondern als Beschreibung eines Verhältnisses auffassen können. Solche fachlichen Hierarchien von Grundvorstellungen und ihren mathematischen Voraussetzungen stellen den Rahmen für fachliches Lernen und Lehren dar und nicht primär die Schul- oder Organisationsstruktur.

Fazit und Thesen zu den Dimensionen offener Fragen

Im Eingangskapitel wurden sechs Dimensionen benannt, in denen im Zusammenhang mit Unterrichtsqualität offene Fragen bestehen. Zu jeder Dimension wird im Folgenden zusammenfassend in Form einer These Stellung bezogen:

- 1) *Fachspezifität vs. generische Qualität*: Unterrichtsqualität ist sowohl generisch als auch fachspezifisch zu erfassen. Die beiden Perspektiven sind komplementär und ergänzen sich gewinnbringend durch je andere Foki, Zielsetzungen und Instrumente.

Empirische Belege für diese These wurden in Kapitel 3 dargelegt.

- 2) *Fachspezifität vs. Inhaltsspezifität (bzw. Kompetenzspezifität)*: Unterrichtsqualität ist nicht nur fachspezifisch zu konzipieren, sondern je nach Inhalt bzw. je nach fokussierter Kompetenz, die aufgebaut werden soll, auch inhaltsspezifisch bzw. kompetenzspezifisch. Deshalb gibt es kein übergreifendes, allgemeingültiges fachspezifisches Konzept der Qualität von Mathematikunterricht.

Zur Begründung und Belegung dieser These kann auf die Ausführungen in Kapitel 4 verwiesen werden.

- 3) *Tiefen- vs. Sichtstruktur von Unterricht*: Sichtstrukturmerkmale können sowohl generischer als auch fachspezifischer Art sein und bilden den Möglichkeitsraum für die auf die Tiefenstruktur von (fachlichen) Lernprozessen abzielenden Qualitätsmerkmale.

Belege für diese These konnten in verschiedenen Studien (insbesondere Studie 2, Studie 5, Studie 8, Studie 9 und Studie 10) erbracht werden (vgl. Kapitel 3.3, Kapitel 4.2, Kapitel 5).

- 4) *Hierarchisierung von Qualitätsmerkmalen*: Generische Qualitätsmerkmale, wie sie beispielsweise in den Basisdimensionen von Unterrichtsqualität formuliert sind, stellen die Grundlage für einen wirkungsvollen Fachunterricht und damit für fachspezifische Unterrichtsqualität dar.

Zu dieser These wurde in Kapitel 3.4 auf der Grundlage der berichteten Studie ein entsprechender Diskussionsvorschlag unterbreitet.

- 5) *Wechselwirkungen von verschiedenen Qualitätsmerkmalen untereinander*: Die angenommenen Hierarchien sind empirisch noch zu überprüfen. Es ist jedoch davon auszugehen, dass eine schwache Qualitätsausprägung generischer Merkmale nicht durch hohe fachspezifische Qualität ausgeglichen werden kann. Da sich die beiden Skalen empirisch separieren lassen, kann es allerdings sein, dass die Einschätzung der generischen Qualität einer Unterrichtsstunde hoch und die Einschätzung der fachspezifischen Qualität gleichzeitig gering ausfällt. Sofern diese Einschätzungen reliabel erfolgt sind, lässt sich daraus ableiten, dass diese Differenzen möglicherweise zu Störungen der generischen Qualität führen können.

Die empirische Trennbarkeit von fachdidaktischen und generischen Qualitätsmerkmalen kann mit den Befunden von Studie 1 (vgl. Kapitel 3.2) belegt werden.

- 6) *Einfluss des kulturellen Hintergrundes und der systemisch bedingten Rahmenbedingungen auf die Unterrichtsqualität*: Die systemischen Rahmenbedingungen scheinen die Unterrichtsqualität auf der Ebene von Sichtstrukturmerkmalen kaum zu beeinflussen, wirken sich aber auf das Erleben der Lehrpersonen aus und dadurch möglicherweise langfristig auf die Merkmale der Tiefenstrukturebene bzw. auf die fachspezifischen Qualitätsmerkmale.

Die Ergebnisse der Studien 8 bis 10 (vgl. Kapitel 5) können als erste Anhaltspunkte zur empirischen Belegung dieser These angesehen werden.

6.2 Epistemisches Potenzial eines interdisziplinären Ansatz

Allen hier vorgestellten Arbeiten liegt ein interdisziplinärer Ansatz zugrunde: Ausgehend von entsprechenden theoretischen und methodischen Grundlagen aus dem Bereich der Erziehungswissenschaften, wurden die zu untersuchenden Fragestellungen konsequent auf den Fachbereich Mathematik bezogen und meist entlang eines quantitativen Forschungsparadigmata bearbeitet. Dabei zeigten sich in der Art der interdisziplinären Kooperation grundlegende Unterschiede. Interdisziplinäre Zusammenarbeit ist zum einen unabdingbar für einen fundierten Erkenntnisgewinn zu gemeinsamen Fragen, die Anteile von generischen und von fachspezifischen Aspekten aufweisen. Zum anderen sind verschiedene Formen der interdisziplinären Kooperation möglich und denkbar, die nach Forschungspersönlichkeiten, nach Fragestellungen und Forschungsdesigns unterschiedlich ausgestaltet werden können. In einzelnen Arbeiten – Studien 2 und 9 – wurde eine komplementäre Sichtweise durch den Einbezug der mathematikdidaktischen Perspektive in erziehungswissenschaftlich bzw. psychologisch angelegte Studien mit teils sehr elaborierten statistischen Verfahren erreicht. Der Ansatz der interdisziplinären Kooperation verlief in diesen Projekten im Wesentlichen gemäß dem Konzept der Arbeitsteilung zwischen unterschiedlichen Feldern der Expertise: Die fachdidaktische Sicht wurde von der Fachdidaktikerin eingebracht, während die Partnerinnen und Partner eine psychologische (Studie 2) Sicht vertraten oder sich mit der Umsetzung eines Studiendesigns in einem anderen Kanton befassten und sich intensiv an der Datenauswertung beteiligten (Studie 9).

Anderen präsentierten Arbeiten – Studien 3–7, 9 und 10 – liegt eine stärker eigenständig fachdidaktische Forschungsperspektive zugrunde, aus der heraus sowohl fachspezifische Fragestellungen formuliert als auch fachspezifische Analysen und Interpretationen vorgenommen wurden, die zugleich zwar Methoden und Konzepte aus dem Bereich der Erziehungswissenschaften einschlossen, jedoch ohne direkten Einbezug von Partnerinnen und Partnern aus dem Bereich der Erziehungswissenschaften, der Psychologie oder der empirischen Bildungsforschung durchgeführt wurden. Ein solcher Einbezug erfolgte personell durch die Autorin bzw. die Forschungsgruppe, da einzelne der beteiligten Personen sowohl über ein Studium in Mathematikdidaktik als auch über einen Studienabschluss in Kognitionspsychologie, Allgemeine Didaktik und Erziehungswissenschaft verfügten. Ein solch doppeltes bzw. mehrfaches wissenschaftliches Kompetenzprofil ist aber eher die Ausnahme als die Regel.

Eine Sonderstellung von Interdisziplinarität wird in Studie 1 (vgl. Kapitel 3.2) erkennbar. Hier arbeiteten zwei Forschende auf Augenhöhe partnerschaftlich zusammen – und zwar von Anfang an und während des gesamten Studienzeitraums – und brachten je ihre Expertise ein in die Untersuchung von Unterrichtsqualität aus unterschiedlichen Perspektiven im Zusammenhang mit dem Lernen im Praxisfeld von Studierenden, die eine Ausbildung zur Lehrperson für die Primarstufe absolvieren. Eine solche erkenntniserweiternde Verbindung unterschiedlicher Forschungsperspektiven mit ihren je unterschiedlichen Traditionen und teils auch Vorgehensweisen ist anspruchsvoll und gleichermaßen bereichernd. Notwendig für das Gelingen dieser Art der Kooperation sind das Herstellen eines geteilten Verständnisses zu einer Sachfrage, ein gemeinsam geteiltes Erkenntnisinteresse sowie Übereinstimmung in Bezug auf das Vorgehen. Dies ist zeitaufwendig und verlangt viele Absprachen sowie Klärungsprozesse und ist nur dann möglich,

wenn die jeweils andere Perspektive von beiden Seiten als prinzipiell aufschlussreich und in komplementärem Sinne bereichernd betrachtet wird und eine Bereitschaft besteht, sich vertieft auf die je andere Disziplin mit ihren Begrifflichkeiten, theoretischen Konzepten und Methoden einzulassen. Letzteres verfolgt nicht das Ziel, in den betreffenden Bereichen selbst Expertise zu erlangen, sondern soll es vielmehr ermöglichen, in einen fruchtbaren Arbeitsaustausch mit der beteiligten Expertin bzw. dem beteiligten Experten zu kommen.

6.3 Implikationen für die Praxis

Für die Unterrichtspraxis bedeuten die Erkenntnisse der vorgestellten Arbeiten insgesamt, dass „Unterrichtsqualität“ für viele unterrichtliche Prozesse, die Lehrpersonen unterschiedlicher Schulstufen in ihrer täglichen Praxis des Mathematikunterrichts umsetzen, eine zu grobkörnige Kategorie ist, und dies insbesondere dann, wenn zentrale Voraussetzungen für erfolgreiches Lehren und Lernen wie eine funktionierende Klassenführung (Ophardt & Thiel, 2017) gegeben sind. Wollen Lehrpersonen ihren eigenen Unterricht optimieren, möchten sie dies nicht selten bezogen auf ein bestimmtes Fach erreichen. Wenn zu diesem Zweck eine Orientierung an generischen Qualitätsmerkmalen erfolgt, wie dies beispielsweise in Aufsatz 2 (vgl. Kapitel 2.3) kritisch diskutiert wurde, besteht relativ bald nur noch wenig Spielraum für eine kontinuierliche Weiterentwicklung der Vermittlung und ko-konstruktiven Bearbeitung bestimmter Fachinhalte und bestimmter fachlicher Prozesse. Diesbezüglich dürften Analyseraster wie dasjenige, dessen Entwurf in der Diskussion der Arbeiten rund um das Thema „Qualität am Beispiel von mathematischem Begründen, Argumentieren und Beweisen“ erläutert wurde (Tabelle 2, Abschnitt 4.5), deutlich erfolgversprechender sein, weil sie einzelne notwendige psychologische Prozessschritte einer bestimmten Kompetenz aufgreifen und diese bei der inhaltlichen Bearbeitung mit Qualitätsanforderungen verbinden. Eine Lehrperson, die erkennt, dass in ihrem Unterricht bei Begründungssequenzen zentrale Verstehens- bzw. Prozessschritte (RC, „Robust Criteria“) fehlen, sieht in der Folge konkret auch entsprechendes Handlungs- und Verbesserungspotenzial, das sich durch den Einbezug dieser fehlenden Schritte ergibt. Insofern sind solche fach-, inhalts- und kompetenznahen Konkretisierungen von Unterrichtsqualität sinnvoll und möglicherweise auch sehr effektiv, weil das Optimierungspotenzial sehr konkret ausgewiesen wird.

Des Weiteren deuten die Erkenntnisse darauf hin, dass der Versuch, Unterrichtsqualität in Form eines allgemeinen Konstrukts zu erfassen, möglicherweise dem fachlichen Handeln von Lehrpersonen nicht gerecht wird und die Qualität ihres Fachunterrichts mittels eines solchen Ansatzes nicht angemessen eingeschätzt werden kann. Es wäre daher zu überlegen, ob bei einer Qualitätseinschätzung von Unterricht entweder immer eine erziehungswissenschaftlich-allgemeindidaktische und eine fachdidaktische Sicht einbezogen werden müssten oder ob die Perspektive aufgrund der Fragestellung zur Qualität des Unterrichts gezielt bestimmt und festgelegt werden müsste. Diese Überlegungen führen zu praktischen Konsequenzen auch auf der Ebene von Qualifikationsverfahren von Lehrpersonen. Denn das Konstatieren des Vorliegens allgemeiner Unterrichtsqualität, beispielsweise im Mathematikunterricht, sagt noch nichts darüber aus, wie gut eine Lehrperson den Aufbau einer ganz bestimmten mathematischen Kompetenz gestalten kann. In der Lehrenden- und Lehrerbildung wären deshalb Qualitätsaspekte aus verschiedenen Perspektiven heraus zu beleuchten, und zwar mit ihren je spezifischen Möglichkeiten und Grenzen.

Weil – wie dies in den Studien 8 und 9 deutlich wurde – Unterricht innerhalb bestimmter struktureller Rahmenbedingungen gestaltet wird, die sich möglicherweise weniger in unterschiedlichen Gestaltungsmustern auswirken als in den Überzeugungen und der Anstrengungsbereitschaft von Lehrpersonen (Studie 10), müssen auch system- und personenbezogene Voraussetzungen entsprechende Aufmerksamkeit erhalten, und zwar nicht unbedingt mit Blick auf die effektiv erreichte Unterrichtsqualität, sondern vielmehr mit Blick auf die Anstrengung und den Aufwand, den Lehrpersonen innerhalb bestimmter Strukturen und Kontexte erbringen müssen, um eine gute Qualität von Fachunterricht sicherstellen zu können. System- und personenbezogene Voraussetzungen sollten deshalb sowohl im Rahmen der Aus- und Weiterbildung von Lehrpersonen als auch im Zusammenhang mit Schulentwicklungsanliegen und Reformen als eigene relevante Faktoren eingehend diskutiert und berücksichtigt werden.

Für die Praxis bedeutsam ist auch die Erkenntnis aus Studie 7 (vgl. Kapitel 4.4), dass es in Interventionsstudien gelingen kann, Lehrpersonen mit anspruchsvollen fachlichen Konzepten und Kompetenzen vertraut zu machen und ihnen aufzuzeigen bzw. sie dazu anzuleiten, wie sie diese altersangemessen und in Übereinstimmung mit ihren vertrauten didaktischen Konzepten im Unterricht umsetzen können. Interventionsstudien wie die vorgestellte Pilotstudie unterliegen forschungsmethodisch diversen Einschränkungen und sind im Gegensatz zu klar kontrollierbaren Laborstudien nie vollständig kontrollierbar. Sie sind deswegen jedoch weder aus Forschungssicht minderwertig noch im Hinblick auf den fachdidaktischen Erkenntnisgewinn weniger wertvoll. Vielmehr dürfte je nach Studienkonzept sogar das Gegenteil der Fall sein: In solchen Studien kann vertieftes Theoriewissen praxisnah vermittelt und die Umsetzung entsprechend angeleitet, evaluiert und ausgewertet werden. Fachdidaktische Forschung weist mit ihren fünf unterschiedlichen Aufgabenfeldern eine sehr hohe Nähe zur Praxis und zum unterrichtlichen Alltag auf und versucht, Praxis zu verstehen, sie weiterzuentwickeln und Lösungen auszuarbeiten, die einen Beitrag zum Gelingen von Mathematikunterricht leisten können, sei dies durch das beispielhafte Ausarbeiten von prototypischen Situationen, wie dies in der Studie der Fall war, oder sei dies durch die Entwicklung und die Erprobung von geeigneten Lernmaterialien und Lernumgebungen. Aus der Nähe zur Praxis ergeben sich für die fachdidaktische Forschung aber auch neue Fragestellungen, die zusammen mit der Praxis oder für diese bearbeitet werden können, damit fachdidaktische Forschung zu einem Forschen und Lernen für, mit und von der Praxis wird.

6.4 Weiterer Forschungsbedarf

Weiterer Forschungsbedarf besteht auf mindestens vier Ebenen: 1) auf der Ebene der Bestimmung von Unterrichtsqualität aus unterschiedlichen Perspektiven mit dem Ziel, sowohl integrierende als auch komplementäre Ansätze zu identifizieren, zu erproben und zu evaluieren, 2) auf inhaltlicher Ebene zum mathematischen Argumentieren, Begründen und Beweisen und zur Erreichung entsprechender fachdidaktisch begründeter und empirisch belegbarer Unterrichtsqualität auf verschiedenen Schulstufen, 3) auf inhaltlicher Ebene zu den Auswirkungen von verschiedenen Kontextvariablen auf die Unterrichtsqualität im Fach und 4) auf der Ebene der Forschung im Hinblick auf eine verstärkte interdisziplinäre Sichtweise bzw. Zusammenarbeit innerhalb der inhaltlichen Forschungsbemühungen rund um das Thema der Unterrichtsqualität.

Auf allen drei inhaltlichen Ebenen ist zum einen vertiefte Theoriearbeit zu leisten und zum anderen muss gezielt nach methodischen Umsetzungen und möglichen Studiendesigns gesucht werden, die es ermöglichen, Qualität im Fachunterricht – auch sehr spezifisch für den Aufbau

einer bestimmten Kompetenz oder unter bestimmten strukturellen Voraussetzungen – systematisch zu erforschen und diese Kompetenz des Weiteren auch – beispielsweise durch Interventionsstudien – gezielt zu verbessern. Für alle drei inhaltlichen Ebene dürfte sich eine verstärkte interdisziplinäre Zusammenarbeit als besonders gewinnbringend erweisen, wobei die Partnerinnen und Partner bzw. die beteiligten Disziplinen je nach Fragestellung andere wären: Im Zusammenhang mit der Weiterentwicklung der Qualität von Begründungs- und Beweisequenzen und deren Auswirkung auf das Lernen der Schülerinnen und Schüler dürfte die Zusammenarbeit mit der Entwicklungspsychologie ebenso bedeutsam sein wie dann, wenn es darum geht, eine für ein bestimmtes Alter oder eine bestimmte Schulstufe adäquate Form der Bearbeitung mathematischer Begründungsprozesse zu entwickeln und zu erforschen. Im Zusammenhang mit Kontextvariablen wiederum wäre nicht nur der Einbezug der Sicht der empirischen Bildungsforschung sinnvoll, sondern ebenso sehr – je nach Fragestellung – auch die Berücksichtigung der Perspektive der Systemsteuerung oder einer kulturwissenschaftlichen Sicht. Inwiefern diese unterschiedlichen Disziplinen mit ihren Begriffen, Traditionen, Methoden und Theoriekonzepten kompatibel sind und sich gegenseitig verständigen können, um eine gemeinsame Fragestellung interdisziplinär zu erforschen, wird sich in der praktischen Umsetzung dieses kooperationsorientierten Ansatzes zeigen müssen. Die fachdidaktische Forschung ist diesbezüglich in ihrer Brückenfunktion (Reusser, 1991) in einer vielversprechenden Position: Für sie ist interdisziplinäres Arbeiten Notwendigkeit und Selbstverständlichkeit, und dies bei gleichzeitiger selbstbewusster Eigenständigkeit als eigene Disziplin.

Im Rahmen der vorliegenden Schrift wurde Unterrichtsqualität aus mathematikdidaktischer Sicht beleuchtet. Eine erste Grundlegung des Themas wurde erarbeitet (vgl. Kapitel 2), die anschließend an drei inhaltlichen Themen konkretisiert und teilweise auch empirisch überprüft wurde (vgl. Kapitel 3–5). Diese Darstellung müsste nun in einem weiteren Schritt nicht nur vertieft und umfassend empirisch überprüft werden, sondern zudem aus anderen Disziplinen heraus im interdisziplinären Diskurs kritisch rezipiert, ergänzt und diskutiert werden.

Literatur

- Aebli, H. (1981). *Denken. Das Ordnen des Tuns* (Bd. 2). Stuttgart: Klett-Cotta.
- Aebli, H. (1987). Mental development: Construction in a cultural context. In B. Inhelder, D. Caprona & A. Cornu-Wells (Hrsg.), *Piaget today* (S. 217–232). Frome/Somerset: Tradespools.
- Amt für Volksschule des Kantons Thurgau. (2016). *Lehrplan Volksschule Thurgau. Mathematik*. Frauenfeld: Amt für Volksschule des Kantons Thurgau.
- Arbinger, R. (1997). *Psychologie des Problemlösens: Eine anwendungsorientierte Einführung*. Darmstadt: Primus.
- Baer, M. (Hrsg.). (2006). *Didaktik auf psychologischer Grundlage: Von Hans Aebli's kognitionspsychologischer Didaktik zur modernen Lehr- und Lernforschung*. Bern: hep.
- Ball, D. L. & Forzani, F. M. (2009). The work on teaching and the challenge for teacher education. *Journal of Teacher Education*, 60(5), 497–511.
- Barzel, B., Holzäpfel, L., Leuders, T. & Streit, C. (2017). *Mathematik unterrichten: Planen, durchführen, reflektieren* (5. Aufl.). Berlin: Cornelsen.
- Berliner, D. C. (2005). The near impossibility of testing for teacher quality. *Journal of Teacher Education*, 56(3), 205–213.
- Biedermann, H., Weber, C., Herzog-Punzenberger, B. & Nagel, A. (2016). Auf die Mitschüler/innen kommt es an? Schulische Segregation – Effekte der Schul- und Klassenzusammensetzung in der Primarstufe und Sekundarstufe I. In M. Bruneforth, F. Eder, K. Krainer, C. Schreiner, A. Seel & C. Spiel (Hrsg.), *Nationaler Bildungsbericht Österreich 2015. 2. Fokussierte Analysen bildungspolitischer Schwerpunktthemen* (S. 133–174). Graz: Leykam.
- Bishop, A. J. (1997). *Mathematical enculturation: A cultural perspective on mathematics education*. Dordrecht: Kluwer Academic Publisher.
- Blömeke, S. & Delaney, S. (2012). Assessment of teacher knowledge across countries: A review of the state of research. *ZDM Mathematics Education*, 44(3), 223–247.
- Blömeke, S., Hsieh, F.-J., Kaiser, G. & Schmidt, W. H. (2014). *International perspectives on teacher knowledge, beliefs and opportunities to learn. TEDS-M results*. Heidelberg: Springer.
- Blömeke, S., Kaiser, G., Lehmann, R., König, J., Döhrmann, M., Buchholtz, C. & Hacke, S. (2009). TEDS-M: Messung von Lehrerkompetenzen im internationalen Vergleich. In R. Mulder, O. Zlatkin-Troitschanskaia, K. Beck, N. Reinhold & D. Sembill (Hrsg.), *Professionalität von Lehrenden – Zum Stand der Forschung* (S. 181–210). Weinheim: Beck.
- Brophy, J. E. & Good, T. L. (1986). Teacher behavior and student achievement. In M. C. Wittrock (Hrsg.), *Handbook of research on teaching* (3. Aufl., S. 328–375). London: Macmillan.
- Bruder, R., Leuders, T. & Büchter, A. (2016). *Mathematikunterricht entwickeln: Bausteine für kompetenzorientiertes Unterrichten* (4. Auflage). Berlin: Cornelsen.
- Bruner, J. (1971). Über kognitive Entwicklung. In J. Bruner, R. R. Olver & P. M. Greenfield (Hrsg.), *Studien zur kognitiven Entwicklung* (S. 21–96). Stuttgart: Klett.
- Brunner, E. (im Druck). Wie lassen sich schriftliche Begründungen von Schülerinnen und Schü-

- lern des 5. und 6. Schuljahrs beschreiben? In S. Krauss & K. Binder (Hrsg.), *Beiträge zum Mathematikunterricht 2019* (S. x-y). Münster: VTM.
- Brunner, E. (2013). *Innermathematisches Beweisen und Argumentieren in der Sekundarstufe I*. Münster: Waxmann.
- Brunner, E. (2014). *Mathematisches Argumentieren, Begründen und Beweisen: Grundlagen, Befunde und Konzepte*. Heidelberg: Springer.
- Brunner, E. (2015). Mathematikunterricht in einer integrativen Schule: Eine fachliche, fachdidaktische und sonderpädagogische Herausforderung. *Schweizerische Zeitschrift für Heilpädagogik*, 21(4), 13–19.
- Brunner, E. (2016). Bin ich ein Teil von dir? Kindergartenkinder lernen mathematisch argumentieren. *4 bis 8*, 2016(8), 37–39.
- Brunner, E. (2018a). *Mathematisches Argumentieren im Kindergarten fördern. Eine Handreichung*. Kreuzlingen: Pädagogische Hochschule Thurgau.
- Brunner, E. (2018b). Qualität von Mathematikunterricht: Eine Frage der Perspektive. *Journal für Mathematik-Didaktik JMD*, 39(2), 257–284.
- Brunner, E., Kreis, A., Staub, F. C., Schoy-Lutz, M. & Kosorok Labhart, C. (2014). Qualitätssteigerung von Mathematikunterricht dank Fachspezifischem Unterrichtscoaching. In J. Roth & J. Ames (Hrsg.), *Beiträge zum Mathematikunterricht 2014* (S. 273–276). Koblenz: Universität Koblenz-Landau.
- Carle, U. & Berthold, B. (2004). *Schuleingangsphase entwickeln. Leistungen fördern. Wie 15 staatliche Grundschulen in Thüringen die flexible, jahrgangsgemischte und integrative Schuleingangsphase einrichten*. Baltmannsweiler: Schneider Verlag Hohengehren.
- Carman, K. G. & Zhang, L. (2012). Classroom peer effects and academic achievement: Evidence from a Chinese middle school. *China Economic Review*, 23(2), 223–237.
- Charalambous, C. Y. & Praetorius, A.-K. (2018). Studying mathematics instruction through different lenses: Setting the ground for understanding instructional quality more comprehensively. *ZDM Mathematics Education*, 50(3), 355–366.
- Clements, D., Sarama, J. & Di Biase, A. (2004). *Engaging young children in mathematics: Standards for early childhood mathematics*. Mahwah, NJ: Erlbaum.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2. Aufl.). Mahwah, NJ: Erlbaum.
- Collins, A., Brown, J. & Newman, S. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In L. B. Resnick (Hrsg.), *Knowing, learning, and instruction: Essays in the honour of Robert Glaser* (S. 453-495). Mahwah, NJ: Erlbaum.
- Common Core State Standards Initiative. (2012). *Mathematics Standards*. Common Core State Standards Initiative. Abgerufen von: <http://www.corestandards.org/Math>
- D-EDK. (2016). *Lehrplan 21*. Abgerufen von: www.lehrplan21.ch
- D-EDK. (Hrsg.). (2016). *Lehrplan 21. Mathematik*. Abgerufen von: www.lehrplan21.ch
- Defila, R., Di Giulio, A. & Scheuermann, M. (2006). *Forschungsverbundmanagement. Handbuch zur Gestaltung inter- und transdisziplinärer Projekte*. Zürich: vdf Hochschulverlag AG.

- Ditton, H. (2000). Qualitätskontrolle und Qualitätssicherung in Schule und Unterricht. *Zeitschrift für Pädagogik*, 41(Beiheft. Qualität und Qualitätssicherung im Bildungsbereich: Schule, Sozialpädagogik, Hochschule), 73–92.
- Ditton, H. (2013). Kontexteffekte und Bildungsungleichheit: Mechanismen und Erklärungsmuster. In R. Becker & A. Schulze (Hrsg.), *Bildungskontexte. Strukturelle Voraussetzungen und Ursachen ungleicher Bildungschancen* (S. 173–206). Wiesbaden: Springer.
- Ditton, H. & Merz, D. (2000). *Qualität von Schule und Unterricht. Kurzbericht über erste Ergebnisse einer Untersuchung an bayrischen Schulen*. Osnabrück: Universität. Abgerufen von: <http://www.quassu.net/Bericht1.pdf>
- Dörner, D. (1974). *Problemlösen als Informationsverarbeitung*. Stuttgart: Kohlhammer.
- Drollinger-Vetter, B. (2011). *Verstehenselemente und strukturelle Klarheit: Fachdidaktische Qualität der Anleitung von mathematischen Verstehensprozessen im Unterricht*. Münster: Waxmann.
- Duncker, K. (1935). *Zur Psychologie des produktiven Denkens*. Berlin: Springer.
- Feuser, G. (2013). Die „Kooperation am Gemeinsamen Gegenstand“. *Behinderte Menschen, Zeitschrift für gemeinsames Leben, Lernen und Arbeiten*, 2, 17–35.
- Fickermann, D., Weishaupt, H. & Zedler, P. (Hrsg.). (1998). *Kleine Grundschulen in Europa: Berichte aus elf europäischen Ländern*. Weinheim: Dt. Studien-Verlag.
- Fischer, H. & Malle, G. (2004). *Mensch und Mathematik. Eine Einführung in didaktisches Denken und Handeln* (Nachdruck). Wien: Profil.
- Freeseemann, O. (2014). *Schwache Rechnerinnen und Rechner fördern. Eine Interventionsstudie an Haupt-, Gesamt- und Förderschulen*. Wiesbaden: Springer.
- Freudenthal, H. (1977). *Mathematik als pädagogische Aufgabe* (Bd. 1 & 2). Stuttgart: Klett.
- Funke, J. (2003). *Problemlösendes Denken* (1. Aufl.). Stuttgart: Kohlhammer.
- Garuti, R., Boero, P. & Lemut, E. (1998). Cognitive units of theorems and difficulty of proof. In A. Olivier & K. Newstead (Hrsg.), *Proceedings of the 22th International Conference of PME* (Bd. 2, S. 345–352). Stellenbosch, South Africa: University of Stellenbosch.
- Gasteiger, H. (2012). Fostering early mathematical competencies in natural learning situations—Foundation and challenges of a competence-oriented concept of mathematics education in kindergarten. *Journal für Mathematik-Didaktik JMD*, 33(2), 181–201.
- Gasteiger, H. (2015). Early mathematics in play situations: Continuity of learning. In B. Perry, A. Gervasoni & A. MacDonald (Hrsg.), *Mathematics and transition to school. International perspectives* (S. 255–272). Singapore: Springer.
- Gasteiger, H., Brunner, E. & Chen, Ch.-S. (2020). Basic conditions of early mathematics education—A comparison between Germany, Taiwan and Switzerland. *International Journal of Science and Mathematics Education* (Online first), 1–17.
- Gillies, R. (2015). Teacher dialog that supports collaborative learning in the classroom. In L. B. Resnick, C. S. C. Asterhan & S. N. Clarke (Hrsg.), *Socializing intelligence through academic talk and dialogue* (S. 335–345). Washington DC: American Educational Research Association AERA.
- Givvin, K. B., Hiebert, J., Jacobs, J. K., Hollingsworth, H. & Gallimore, R. (2005). Are there na-

- tional patterns of teaching? Evidence from the TIMSS 1999 video study. *Comparative Education Review*, 49(3), 311–343.
- Goetz, T., Lohrmann, K., Ganser, B. & Haag, L. (2005). Einsatz von Unterrichtsmethoden – Konstanz oder Wandel? *Empirische Pädagogik*, 19(4), 342–360.
- Greeno, J. G. (2006). Authoritative, accountable positioning and connected, general knowing: Progressive theme in understanding transfer. *The Journal of the Learning Sciences*, 15(4), 537–547.
- Greeno, J. G. (2015). Classroom talk sequences and learning. In L. B. Resnick, C. S. C. Asterhan & S. N. Clarke (Hrsg.), *Socializing intelligence through academic talk and dialogue* (S. 255–262). Washington DC: American Educational Research Association AERA.
- Grell, J. & Grell, M. (2010). *Unterrichtsrezepte* (12., neu ausgestattete Aufl.). Weinheim: Beltz.
- Hammer, S., Reiss, K., Lehner, M. C., Heine, J.-H., Sälzer, C. & Heinze, A. (2016). Mathematische Kompetenz in PISA 2015: Ergebnisse, Veränderungen und Perspektiven. In K. Reiss, C. Sälzer, A. Schiepe-Tiska, E. Klieme & O. Köller (Hrsg.), *PISA 2015. Eine Studie zwischen Kontinuität und Innovation* (S. 219–247). Münster: Waxmann.
- Heintz, B. (2000). *Die Innenwelt der Mathematik. Zur Kultur und Praxis einer beweisenden Disziplin*. Wien: Springer.
- Heinze, A. & Grüßing, M. (Hrsg.). (2009). *Mathematiklernen vom Kindergarten bis zum Studium. Kontinuität und Kohärenz als Herausforderung für den Mathematikunterricht*. Münster: Waxmann.
- Heitzmann, A. & Pauli, C. (2015). Professionalisierung in den Fachdidaktiken – Überlegungen zu einem zentralen, aber nicht unproblematischen Begriff. Einführung ins Themenheft. *Beiträge zur Lehrerinnen- und Lehrerbildung BzL*, 33(2), 183–199.
- Helmke, A. (2004). *Unterrichtsqualität erfassen, bewerten, verbessern*. Seelze: Kallmeyer.
- Helmke, A. (2006). Was wissen wir über guten Unterricht? *Pädagogik*, 58(2), 42–45.
- Helmke, A. & Weinert, F. E. (1997). Bedingungsfaktoren schulischer Leistungen. In F. E. Weinert (Hrsg.), *Psychologie des Unterrichts und der Schule* (Bd. 3, S. 71–176). Göttingen: Hogrefe.
- Henne, H. & Rehbock, H. (2001). *Einführung in die Gesprächsanalyse*. Berlin: De Gruyter.
- Hersh, R. (1993). Proving is convincing and explaining. *Educational Studies in Mathematics*, 24(2), 389–399.
- Hiebert, J., Carpenter, T. P., Fennema, E., Fuson, K. C., Wearne, D., Murray, H., Olivier, A. & Human, P. (1997). *Making sense. Teaching and learning mathematics with understanding*. Portsmouth, NH: Heinemann.
- Hiebert, J., Gallimore, R., Garnier, H., Givvin, K. B., Hollingsworth, H. & Jacobs, J. (2003). *Teaching mathematics in seven countries. Results from the TIMSS 1999 video study*. Washington, DC: U.S. Department of Education. Institute of Education Sciences.
- Hugener, I., Pauli, C., Reusser, K., Lipowsky, F., Rakoczy, K. & Klieme, E. (2009). Teaching patterns and learning quality in Swiss and German mathematics lessons. *Learning and Instruction*, 19(19), 66–78.
- Jahnke, H. N. & Ufer, S. (2015). Argumentieren und Beweisen. In R. Bruder, L. Hefendehl-Hebeker, B. Schmidt-Thieme & H.-G. Weigand (Hrsg.), *Handbuch der Mathematikdidaktik* (S.

- 331–355). Heidelberg: Springer Spektrum.
- Kaur, B., Anthony, G. & Ohtani, M. (2013). *Student voice in mathematics classrooms around the world*. Rotterdam: Sense Publisher.
- Klafki, W. (1991). *Neue Studien zur Bildungstheorie und Didaktik. Zeitgemäße Allgemeinbildung und kritisch-konstruktive Didaktik*. Weinheim: Beltz.
- Klieme, E., Lipowsky, F., Rakoczy, K. & Ratzka, N. (2006). Qualitätsdimensionen und Wirksamkeit von Mathematikunterricht. In M. Prenzel & L. Allolio-Näcke (Hrsg.), *Untersuchungen zur Bildungsqualität von Schule* (S. 127–146). Münster: Waxmann.
- Klieme, E., Pauli, C. & Reusser, K. (2009). The Pythagoras Study. In T. Janik & T. Seidel (Hrsg.), *The power of video studies in investigating teaching and learning in the classroom* (S. 137–160). Münster: Waxmann.
- Klieme, E., Schümer, G. & Knoll, S. (2001). Mathematikunterricht in der Sekundarstufe I. „Aufgabenkultur“ und Unterrichtsgestaltung. In Bundesministerium für Bildung und Forschung (Hrsg.), *TIMSS – Impulse für Schule und Unterricht* (S. 43–57). Bonn: BMBF.
- KMK. (2005). *Bildungsstandards der Kultusministerkonferenz. Erläuterungen zur Konzeption und Entwicklung*. München: Luchterhand.
- Kreis, A. (2012). *Produktive Unterrichtsbesprechungen: Lernen im Dialog zwischen Mentoren und angehenden Lehrpersonen*. Bern: Haupt.
- Kreis, A. & Staub, F. C. (2011). Fachspezifisches Unterrichtscoaching im Praktikum – eine quasi-experimentelle Interventionsstudie. *Zeitschrift für Erziehungswissenschaft*, 14(1), 61–83.
- Kron, F. W. (2008). *Grundwissen Didaktik* (5., überarb. Aufl.). München: Reinhardt.
- Krummheuer, G. (2008). Inskription, Narration und diagrammatisch basierte Argumentation. In H. Jungwirth & G. Krummheuer (Hrsg.), *Der Blick nach innen: Aspekte der alltäglichen Lebenswelt Mathematikunterricht* (Band 2, S. 7–37). Münster: Waxmann.
- Kunter, M. (2005). *Multiple Ziele im Mathematikunterricht*. Münster: Waxmann.
- Kunter, M., Baumert, J., Blum, W., Klusmann, U., Krauss, S. & Neubrand, M. (Hrsg.). (2011). *Professionelle Kompetenz von Lehrkräften – Ergebnisse des Forschungsprogramms COACTIV*. Münster: Waxmann.
- 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.
- Laging, R. (2003). Altersmischung – Eine pädagogische Chance. In R. Laging (Hrsg.), *Altersgemischtes Lernen in der Schule* (S. 1–29). Baltmannsweiler: Schneider Verlag Hohengehren.
- Learning Mathematics for Teaching Project. (2011). Measuring the mathematical quality of instruction. *Journal of Mathematics Teacher Education*, 14(1), 25–47.
- Leinhardt, G. & Steele, H. (2005). Seeing the complexity of standing to the side: Instructional dialogues. *Cognition and Instruction*, 23(1), 87–163.
- Leuders, T. (2003). Perspektiven von Mathematikunterricht. In T. Leuders (Hrsg.), *Mathematikdi-*

- daktik. Praxisbandbuch für die Sekundarstufe I und II* (S. 15–58). Berlin: Cornelsen.
- Leuders, T. & Prediger, S. (2016). *Flexibel differenzieren und fokussiert fördern im Mathematikunterricht*. Berlin: Cornelsen.
- Leung, F. K. S. (2006). Mathematics education in East Asia and the West: Does culture matters? In F. K. S. Leung, K. D. Graf & F. J. Lopez-Real (Hrsg.), *Mathematics education in different cultural traditions—A comparative study of East Asia and the West. New ICMI Study series* (Bd. 9, S. 21-46). Boston, MA: Springer.
- Li, Y. & Moschkovich, J. N. (2013). Proficiency and Beliefs in Learning and Teaching Mathematics—An Introduction. In Y. Li & J. N. Moschkovich (Hrsg.), *Proficiency and beliefs in learning and teaching mathematics. Learning from Alan Schoenfeld and Günter Törner* (S. 3–7). Rotterdam: Sense Publishers.
- Lipowsky, F. (2004). Was macht Fortbildungen für Lehrkräfte erfolgreich? Befunde aus der Forschung und mögliche Konsequenzen für die Praxis. *Die Deutsche Schule*, 96(4), 462–479.
- Lipowsky, F. (2015). Unterricht. In E. Wild & J. Möller (Hrsg.), *Pädagogische Psychologie* (S. 69–105). Heidelberg: Springer.
- Lipowsky, F., Drollinger-Vetter, B., Klieme, E., Pauli, C. & Reusser, K. (2018). Generische und fachdidaktische Dimensionen von Unterrichtsqualität. Zwei Seiten einer Medaille? In M. Martens, K. Rabenstein, K. Bräu, M. Fetzer, H. Gresch, I. Hardy & C. Schelle (Hrsg.), *Konstruktionen von Fachlichkeit. Ansätze, Erträge und Diskussionen in der empirischen Unterrichtsforschung* (S. 183–202). Bad Heilbrunn: Klinkhardt.
- Mason, D. A. & Burns, R. (1995). Teachers' views of combination classes. *The Journal of Educational Research*, 89(1), 36–45.
- Mason, J., Burton, L. & Stacey, K. (1982). *Thinking mathematically*. Reading, MA: Addison-Wesley.
- Mayring, P. (2002). *Einführung in die qualitative Sozialforschung* (5., überarb. u. neu ausgest. Aufl.). Weinheim: Beltz.
- Mehan, H. (1979). *Learning lessons. Social organization in the classroom*. Cambridge, MA: Harvard University Press.
- Meyer, H. (1987). *Unterrichtsmethoden II: Praxisband*. Berlin: Cornelsen.
- Meyer, H. (2009). *Was ist guter Unterricht?* Berlin: Cornelsen.
- Ministerium für Kultus, Jugend und Sport Baden-Württemberg. (2012). *Jahrgangübergreifendes Lernen in der Grundschule*. Stuttgart: Ministerium für Kultus, Jugend und Sport.
- Monn, X. (2006). Altersgemischte Lerngruppen – Umgang mit Heterogenität. In A. Tanner, H. Badertscher, R. Holzer, A. Schindler & U. Streckeisen (Hrsg.), *Heterogenität und Integration. Umgang mit Ungleichheit und Differenz in Schule und Kindergarten* (S. 266–275). Zürich: Seismo.
- Muijs, D., Kyriakides, L., van der Werf, G., Creemers, B., Timperley, H. & Earl, L. (2014). State of the art—teacher effectiveness and professional learning. *School Effectiveness and School Improvement*, 25(2), 231–256.
- Mullis, I. V. S. (2012). *TIMSS 2011 international results in mathematics*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center.

- Mullis, I. V. S., Martin, M. O., Ruddock, G. J., O'Sullivan, C. Y. & Preuschoff, C. (2009). *TIMSS 2011. Assessment Frameworks*. Amsterdam: International Association for the Evaluation of Educational Achievement (IEA).
- Mulryan-Kyne, C. M. (2004). Teaching and learning in multigrade classrooms: What the teachers say. *Irish Journal of Education*, 35(2004), 5–19.
- NCTM. (2000). *Principles and standards for school mathematics*. Reston, VA: NCTM.
- Neubrand, J. (2002). *Eine Klassifikation mathematischer Aufgaben zur Analyse von Unterrichtssituationen: Selbsttätiges Arbeiten in Schülerarbeitsphasen in den Stunden der TIMSS-Video-Studie*. Hildesheim: Franzbecker.
- Neubrand, M. (2006). Multiple Lösungswege für Aufgaben: Bedeutung für Fach, Lernen, Unterricht und Leistungserfassung. In W. Blum, C. Drüke-Noe, R. Hartung & O. Köller (Hrsg.), *Bildungsstandards Mathematik: Konkret*. (S. 162–177). Berlin: Cornelsen.
- OECD. (2013). *PISA 2012 assessment and analytical framework: Mathematics, reading, science, problem solving and financial literacy*. Paris: OECD.
- OECD. (2014). *PISA 2012 Results: What students know and can do. Student performance in mathematics, reading and science* (Bd. 1, rev. Aufl.). Paris: OECD.
- OECD (Hrsg.). (2016). *PISA 2015 Results: Excellence and Equity in Education* (Bd. 1). Paris: OECD.
- Ophardt, D. & Thiel, F. (2017). Kompetenzen des Klassenmanagements (KODEK) – Entwicklung und Evaluation eines Fortbildungsprogramms für Lehrpersonen zum Klassenmanagement. In C. Gräsel & K. Trempler (Hrsg.), *Entwicklung von Professionalität des pädagogischen Personals. Interdisziplinäre Betrachtungen, Befunde und Perspektiven* (S. 133–152). Springer Online.
- Pauli, C. & Lipowsky, F. (2007). Mitmachen oder zuhören? Mündliche Schülerinnen- und Schülerbeteiligung im Mathematikunterricht. *Unterrichtswissenschaft*, 35(2), 101–124.
- Pauli, C. & Reusser, K. (2015). Discursive cultures of learning in (everyday) mathematics teaching: A video-based study on mathematics teaching in German and Swiss classrooms. In L. B. Resnick, C. S. C. Asterhan & S. N. Clarke (Hrsg.), *Socializing intelligence through academic talk and dialogue* (S. 181–193). Washington, DC: American Educational Research Association AERA.
- Pauli, C., Reusser, K. & Grob, U. (2010). Reformorientierter Mathematikunterricht in der Deutschschweiz. In K. Reusser, C. Pauli & M. Waldis (Hrsg.), *Unterrichtsgestaltung und Unterrichtsqualität. Ergebnisse einer internationalen und schweizerischen Videostudie zum Mathematikunterricht* (S. 309–339). Münster: Waxmann.
- Piaget, J. (2003). *Meine Theorie der geistigen Entwicklung*. Weinheim: Beltz.
- Pianta, R. C., La Paro, K. M. & Hamre, B. K. (2008). *Classroom assessment scoring system*. Baltimore, MD: Paul H. Brookers Publishing Co.
- Pianta, R. & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher*, 38(2), 109–119.
- Pöhler, B., Prediger, S. & Wessel, L. (2016). Realizing Macro-Scaffolding for Percentages in Mathematics Classrooms—A field experiment. In C. Csíkos, A. Rausch, & J. Sztányi (Hrsg.),

- Proceedings of the 40th Conference of the International Group for the Psychology of Mathematics Education* (Bd. 1, S. 221). Szeged, Hungary: PME.
- Polya, G. (1949). *Schule des Denkens* (1. Aufl.). Bern: Francke.
- Polya, G. (1995). *Schule des Denkens* (4. Aufl.). Bern: Francke.
- Pool Maag, S. & Moser Opitz, E. (2014). Inklusiver Unterricht – grundsätzliche Fragen und Ergebnisse einer explorativen Studie. *Empirische Sonderpädagogik*, 6(2), 133–149.
- Praetorius, A.-K. (2014). *Messung von Unterrichtsqualität durch Ratings*. Münster: Waxmann.
- 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(1), 2–12.
- Prediger, S. (2016). Inklusion im Mathematikunterricht: Forschung und Entwicklung zur fokussierten Förderung statt rein unterrichtsmethodischer Bewältigung. In J. Menthe, D. Höttecke, T. Zabka, M. Hammann & M. Rothgangel (Hrsg.), *Befähigung zu gesellschaftlicher Teilhabe. Beiträge der fachdidaktischen Forschung* (S. 316–372). Münster: Waxmann.
- Radatz, H., Schipper, W., Ebeling, A. & Dröge, R. (1996). *Handbuch für den Mathematikunterricht – Schuljahr 1*. Hannover: Schroedel.
- Rakoczy, K., Klieme, E., Lipowsky, F. & Drollinger-Vetter, B. (2010). Strukturierung, kognitive Aktivität und Leistungsentwicklung im Mathematikunterricht. *Unterrichtswissenschaft*, 38(3), 229–246.
- Rakoczy, K. & Pauli, C. (2006). Hoch inferentes Rating: Beurteilung der Qualität unterrichtlicher Prozesse. In I. Hugener, C. Pauli & K. Reusser (Hrsg.), *Videoanalysen. Dokumentation der Erhebungs- und Auswertungsinstrumente zur schweizerisch-deutschen Videostudie „Unterrichtsqualität, Lernverhalten und mathematisches Verständnis“* (Teil 3, S. 206–233). Frankfurt: DIPF.
- Reiss, K. (2002). *Argumentieren, Begründen, Beweisen im Mathematikunterricht. Projektserver SINUS*. Bayreuth: Universität.
- Reiss, K. & Ufer, S. (2009). Fachdidaktische Forschung im Rahmen der Bildungsforschung. Eine Diskussion wesentlicher Aspekte am Beispiel der Mathematikdidaktik. In R. Tippelt & B. Schmidt (Hrsg.), *Handbuch Bildungsforschung* (2. überarb. u. erw. Aufl., S. 199–213). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Resnick, L. B. (1992). From protoquantities to operators: Building mathematical competence on a foundation of everyday knowledge. In G. Leinhardt, R. Putnam & R. Hatrup (Hrsg.), *Analysis of arithmetic for mathematics teaching* (S. 373–429). Hillsdale, NJ: Erlbaum.
- Reusser, K. (1984). *Problemlösen in wissenstheoretischer Sicht. Problematisches Wissen, Problemformulierung und Problemverständnis*. Bern: Universität.
- Reusser, K. (1989). *Vom Text zur Situation zur Gleichung. Kognitive Simulation von Sprachverständnis und Mathematisierung beim Lösen von Textaufgaben. Habilitationsschrift*. Bern: Universität Bern.
- Reusser, K. (1991). Plädoyer für die Fachdidaktik und für die Ausbildung von Fachdidaktiker/innen für die Lehrerbildung. *Beiträge zur Lehrerbildung BzL*, 9(2), 193–215.
- Reusser, K. (2005). Problemorientiertes Lernen – Tiefenstruktur, Gestaltungsformen, Wirkung. *Beiträge zur Lehrerbildung BzL*, 23(2), 159–182.

- Reusser, K. (2006). Konstruktivismus – vom epistemologischen Leitbegriff zur Erneuerung der didaktischen Kultur. In M. Baer, M. Fuchs, P. Füglistler, K. Reusser & H. Wyss (Hrsg.), *Didaktik auf psychologischer Grundlage. Von Hans Aebli's kognitionspsychologischer Didaktik zur modernen Lehr-Lernforschung* (S. 151–168). Bern: hep.
- Reusser, K. & Pauli, C. (2010). Unterrichtsgestaltung und Unterrichtsqualität – Ergebnisse einer internationalen und schweizerischen Videostudie zum Mathematikunterricht: Einleitung und Überblick. In K. Reusser, C. Pauli & M. Waldis (Hrsg.), *Unterrichtsgestaltung und Unterrichtsqualität* (S. 9–32). Münster: Waxmann.
- Reusser, K. & Pauli, C. (2011). *Didaktische Kommunikation und Bildungswirkungen im problemorientierten Mathematikunterricht* (SNF-Projekt Nr. 113971). Zürich: Universität Zürich, Pädagogisches Institut.
- Reusser, K. & Pauli, C. (2014). Berufsbezogene Überzeugungen von Lehrerinnen und Lehrern. In E. Terhart, H. Bennewitz & M. Rothland (Hrsg.), *Handbuch der Forschung zum Lehrerberuf* (2. üb. und erw., S. 642–661). Münster: Waxmann.
- Reusser, Kurt & Pauli, C. (2003). *Mathematikunterricht in der Schweiz und in weiteren sechs Ländern. Bericht über die Ergebnisse einer internationalen und schweizerischen Video-Unterrichtsstudie*. Zürich: Universität Zürich, Pädagogisches Institut.
- Rott, B. (2014). Mathematische Problembearbeitungsprozesse von Fünftklässlern – Entwicklung eines deskriptiven Phasenmodells. *Journal für Mathematik-Didaktik JMD*, 35(2), 251–282.
- Schindler, V., Moser Opitz, E., Cadonau-Bieler, M. & Ritterfeld, U. (2019). Überprüfung und Förderung des mathematischen Fachwortschatzes der Grundschulmathematik – eine empirische Studie. *Journal für Mathematik-Didaktik JMD*, 40(1), 1–35.
- Schlesinger, L. & Jentsch, A. (2016). Theoretical and methodological challenges in measuring instructional quality in mathematics education using classroom observations. *ZDM Mathematics Education*, 48(1), 29–40.
- Schoenfeld, A. (1985). *Mathematical problem solving*. San Diego, CA: Academic Press.
- Schoenfeld, A. (1994). *Mathematical thinking and problem solving. Studies in mathematical thinking and learning*. Hillsdale, NJ: Erlbaum.
- Schoenfeld, A. H. (2013). Classroom observations in theory and practice. *ZDM Mathematics Education*, 45(4), 607–621.
- Schoenfeld, A. H., Floden, R. E., the Algebra Teaching and Mathematics Assessment Project Berkeley & Lansing, E. (2014). *The TRU Math Scoring Rubric*. Abgerufen von: <http://ats.berkeley.edu/tools.html>
- Schukajlow, S., Leiss, D., Pekrun, R., Blum, W., Müller, M. & Messner, R. (2012). Teaching methods for modelling problems and students' task-specific enjoyment, value, interest and self-efficacy expectations. *Educational Studies in Mathematics*, 79(2), 215–237.
- Schukajlow, S. & Leiss, D. (2011). Selbstberichtete Strategienutzung und mathematische Modellierungskompetenz. *Journal für Mathematik-Didaktik*, 32(1), 53–77.
- Seidel, T. (2014). Angebots-Nutzungs-Modelle in der Unterrichtspsychologie: Integration von Struktur- und Prozessparadigma. *Zeitschrift für Pädagogik*, 60(6), 850–866.

- Sfard, A. (2007). When the Rules of discourse change, but nobody tells you: Making sense of mathematics learning from a commognitive standpoint. *The Journal of the Learning Sciences*, 16(4), 565–613.
- Shimizu, Y., Kaur, B., Huang, R. & Clarke, D. (Hrsg.). (2010). *Mathematical tasks in classrooms around the world*. Rotterdam: Sense Publishers.
- Sommerhoff, D., Brunner, E. & Ufer, S. (2019). Appraisals for different types of proof. In M. Graven, H. Venkat, A. A. Essien & P. Vale (Hrsg.), *Proceedings of the 43rd Conference of the International Group for the Psychology of Mathematics Education* (Bd. 4, S. 101). Pretoria, South Africa: PME.
- Staub, F. C. & Stern, E. (2002). The nature of teachers' pedagogical content beliefs matters for students' achievement gains: Quasi-experimental evidence from elementary mathematics. *Journal of Educational Psychology*, 94(2), 344–355.
- Stylianides, A. J. (2015). The role of mode of representation in students' argument construction. In K. Krainer & N. Vondrová (Hrsg.), *Proceedings of the Ninth Conference of the European Society for Research in Mathematics Education CERME 9* (S. 213–210). Prag: Charles University in Prague, Faculty of Education and ERME.
- Stylianides, A. J. (2016). *Proving in the elementary mathematics classroom*. Oxford University Press.
- Terhart, E. (2002). Fremde Schwestern. Zum Verhältnis von Allgemeiner Didaktik und empirischer Lehr-Lern-Forschung. *Zeitschrift für Pädagogische Psychologie*, 16(2), 77–86.
- Terhart, E. (2014). Unterrichtstheorie: Einführung in den Thementeil. *Zeitschrift für Pädagogik*, 60(6), 813–81.
- Thurn, S. (2010a). Altersmischung in der Schule: Voneinander lernen – Miteinander leben in der Laborschule Bielefeld. In R. Laging (Hrsg.), *Altersgemischtes Lernen in der Schule* (4. unveränd. Aufl., S. 135–150). Baltmannsweiler: Schneider Verlag Hohengehren.
- Thurn, S. (2010b). Leben, lernen, leisten in jahrgangsübergreifenden Gruppen. In A. Buholzer & A. Kummer Wyss (Hrsg.), *Alle gleich – Alle unterschiedlich! Zum Umgang mit Heterogenität in der Schule* (S. 28–39). Zug: Klett & Balmer.
- Toulmin, S. E. (1996). *Der Gebrauch von Argumenten* (2. Aufl.). Weinheim: Beltz.
- Van de Pol, J., Brindley, S. & Higham, R. J. E. (2017). Two secondary teachers' understanding and classroom practice of dialogic teaching: A case study. *Educational Studies in Mathematics*, 43(5), 497–515.
- Van de Pol, J., Volman, M. L. L. & Beishuizen, J. (2010). Scaffolding in teacher-student interaction: A decade of research. *Educational Psychology Review*, 22(3), 271–297.
- Van den Heuvel-Panhuizen, M. (2001). *Children learn mathematics—A learning-teaching trajectory with intermediate and attainment targets for calculation with whole numbers in primary school*. Utrecht: Freudenthal Institute.
- Van den Heuvel-Panhuizen, M. & van den Boogaard, S. (2008). Picture books as an impetus for kindergartners' mathematical thinking. *Mathematical Thinking and Learning*, 10(4), 341–373.
- Van Oers, B. (2010). Emergent mathematical thinking in the context of play. *Educational Studies of Mathematics*, 74(1), 23–37.

- Van Eemeren, F. H. & Grootendorst, R. (2004). *A systematic theory of argumentation. The pragma-dialectical approach*. Cambridge, MA: University Press.
- Verschaffel, L., De Corte, E., Lasure, S., Van Vaerenbergh, G., Bogaerts, H. & Ratinckx, E. (1999). Learning to solve mathematical application problems: A design experiment with fifth graders. *Mathematical Thinking and Learning*, 1(3), 195–229.
- Verschaffel, L., Van Dooren, W., Greer, B. & Mukhopadhyay, S. (2010). Reconceptualising word problems as exercises in mathematical modelling. *Journal für Mathematik-Didaktik JMD*, 31(1), 9–29.
- Walkowiak, T. A., Berry, R. Q., Meyer, J. P., Rimm-Kaufman, S. E. & Ottmar, E. R. (2014). Introducing an observational measure of standards-based mathematics teaching practices: Evidence of validity and score reliability. *Educational Studies in Mathematics*, 85(1), 109–128.
- Weinert, F. E., Schrader, F.-W. & Helmke, A. (1990). Unterrichtsexpertise – Ein Konzept zur Verringerung der Kluft zwischen zwei theoretischen Paradigmen. In L. M. Alisch, J. Baumert & K. Beck (Hrsg.), *Professionswissen und Professionalisierung* (Bd. 28, S. 173–206). Braunschweig: Technische Universität.
- Wellenreuther, M. (2005). *Lehren und Lernen - aber wie? Empirisch-experimentelle Forschungen zum Lehren und Lernen im Unterricht* (2., korr. und überarb. Aufl.). Baltmannsweiler: Schneider Verlag Hohengehren.
- Wernet, J. & Lepak, J. (2014). *TRU Math conversation guide. Modul A: Contextual algebraic tasks*. Berkeley, CA & East Lansing, MI: Graduate School of Education, University of California, Berkeley & College of Education, Michigan State University. Abgerufen von: <http://ats.berkeley.edu/tools.html> and/or <http://map.mathshell.org/materials/pd.php>.
- Wertheimer, M. (1964). *Produktives Denken* (2. Aufl.). Frankfurt: Kramer.
- Wessel, L. (2015). *Fach- und sprachintegrierte Förderung durch Darstellungsvernetzung und Scaffolding: Ein Entwicklungsforschungsprojekt zum Anteilbegriff*. Wiesbaden: Springer Spektrum.
- West, L. & Staub, F. C. (2003). *Content-focused coaching: Transforming mathematics lessons*. Reston, VA: Portsmouth, NH: Heinemann.
- Winter, H. (1975). Allgemeine Lernziele für den Mathematikunterricht? *Zentralblatt für Didaktik der Mathematik ZDM*, 7(3), 106–116.
- Winter, H. (1991). *Entdeckendes Lernen im Mathematikunterricht. Einblicke in die Ideengeschichte und ihre Bedeutung für die Pädagogik* (2. verb. Aufl.). Braunschweig: Vieweg.
- Wittmann, E. C. & Müller, N. G. (1988). Wann ist ein Beweis ein Beweis? In P. Bender (Hrsg.), *Mathematikdidaktik – Theorie und Praxis. Festschrift für Heinrich Winter* (S. 237–258). Berlin: Cornelsen.
- Yackel, E. & Cobb, P. (1996). Sociomathematical norms, argumentation and autonomy in mathematics. *The Journal of Research in Mathematics Education*, 27(4), 458–477.
- Zöttl, L., Ufer, S. & Reiss, K. (2010). Modelling with heuristic worked examples in the KOMMA learning environment. *Journal Für Mathematik-Didaktik JMD*, 31(1), 143–165.

Verzeichnis diskutierter Arbeiten

- Brunner, E. (2014). Verschiedene Beweistypen und ihre Umsetzung im Unterrichtsgespräch. *Journal für Mathematik-Didaktik JMD*, 35(2), 229-249.
- Brunner, E. (2015). Mathematikdidaktische Forschung. Eine notwendige vertiefende Perspektive. *Beiträge zur Lehrerinnen- und Lehrerbildung BzL*, 33(2), 235-245.
- Brunner, E. (2017). Mathematikunterricht in Mehrjahrgangsklassen der Primarschule: Eine Deskription entlang verschiedener Gestaltungselemente und Einschätzungen der Lehrpersonen. *Journal für Mathematik-Didaktik JMD*, 38(1), 57-91.
- Brunner, E. (2018). Qualität von Mathematikunterricht: Eine Frage der Perspektive. *Journal für Mathematik-Didaktik JMD*, 39(2), 257-284.
- Brunner, E. (2019). Förderung mathematischen Argumentierens im Kindergarten: Erste Erkenntnisse aus einer Pilotstudie. *Journal für Mathematik-Didaktik JMD*, 40(2), 323–356.
- Brunner, E., Kreis, A., Staub, F. C., Schoy-Lutz, M. & Kosorok Labhart, C. (2014). Qualitätssteigerung von Mathematikunterricht dank Fachspezifischem Unterrichtscoaching. In J. Roth & J. Ames (Hrsg.), *Beiträge zur Mathematikdidaktik 2014* (S. 273-276). Koblenz: Universität Koblenz-Landau.
- Brunner, E. & Imhof, A. (2017). Gestaltungsmuster des Mathematikunterrichts in Jahrgangs- und Mehrjahrgangsklassen – Zwei Kantone im Vergleich. *Schweizerische Zeitschrift für Bildungswissenschaften SZBW*, 39(2), 353-374.
- Brunner, E., Gasteiger, H., Lampart, J. & Schreieder, K. (2019). Mathematikunterricht in jahrgangsübergreifenden Klassen der Grundschule in der Schweiz und in Deutschland: Eine vergleichende Studie. *Schweizerische Zeitschrift für Bildungswissenschaften SZBW*, 41(1), 160-176.
- Brunner, E. & Reusser, K. (2019). Type of Mathematical Proof: Personal Preference or Adaptive Teaching Behavior? *ZDM Mathematics Education*, 51(5), 747-758.
- Bieg, M., Goetz, T., Sticca, F., Brunner, E., Becker, E. S., Morger, V. & Hubbard, K. (2017). Teaching methods and their impact on students' emotions in mathematics: an experience-sampling approach. *ZDM Mathematics Education*, 49(3), 411-422.
- Lindmeier, A., Brunner, E. & Grüßing, M. (2018). Early mathematical reasoning – theoretical foundations and possible assessment. In E. Bergqvist, M. Österholm, C. Granberg & L. Sumpter (Hrsg.), *Proceedings of the 42th Conference of the International Group for the Psychology of Mathematics Education* (Bd. 3, S. 315–322). Umea: PME.
- Praetorius, A., Klieme, E., Kleickmann, T., Brunner, E., Lindmeier, A., Taut, S. & Charalambous, C. (2020). Towards Developing a Theory of Generic Teaching Quality: Origin, Current Status, and Necessary Next Steps Regarding the Three Basic Dimensions Model. *Zeitschrift für Pädagogik, Beiheft 66*, 15-36.