



**Mehrdimensionale Bildungsziele in internationalen Large-
Scale Assessments:
Konzeptualisierung, Entwicklung und
Entstehungsbedingungen am Beispiel der MINT-Fächer**

Habilitation an der TUM School of Education

Dr. Anja Schiepe-Tiska

Juni 2019

Inhalt

1. Thematische Einordnung und Relevanz.....	3
1.1. Mehrdimensionale Bildungsziele im Kontext des deutschen Bildungsverständnisses und allgemeiner internationaler Grundbildungskonzepte.....	4
1.2. Schulische und unterrichtliche Bedingungen des Erreichens mehrdimensionaler Bildungsziele.....	8
1.3. Rahmenmodell zur Untersuchung mathematisch-naturwissenschaftlicher mehrdimensionaler Bildungsziele.....	12
2. Zusammenfassung der Forschungsarbeiten und ihres Erkenntnisgewinns.....	15
2.1. Forschungsbereich I - Ausprägung, Entwicklung und Zusammenspiel mehrdimensionaler Bildungsziele.....	18
2.2. Forschungsbereich II – Schulische und unterrichtliche Bedingungen des Erreichens mehrdimensionaler Bildungsziele.....	27
3. Eigenleistungen im Kontext der Fachpublikationen.....	39
4. Literaturverzeichnis.....	41

1. Thematische Einordnung und Relevanz

Sowohl gesellschaftlich als auch auf das Individuum bezogen beschränkt sich der Bildungsauftrag der Schule nicht nur auf die Vermittlung von (Fach-)Wissen und kognitiven Fertigkeiten. Zwar bilden diese eine wichtige Grundlage, um selbstbewusst und verantwortlich mit den Herausforderungen der heutigen Zeit umgehen zu können, sie reichen allein jedoch nicht aus. Allgemeine und fachbezogene positive Einstellungen zum Lernen ergänzen Fachwissen und kognitive Fertigkeiten und werden als wichtige Voraussetzungen für lebenslanges Lernen und die aktive Teilhabe an der Gesellschaft gesehen (vgl. Fend, 2008; Prenzel, 2012; Schiepe-Tiska & Schmidtner, 2013). Das Herausbilden von Interessen, Wertvorstellungen und einer positiven Wahrnehmung der eigenen Fertigkeiten sind zentral für weiterführende Ausbildungsentscheidungen und beeinflussen die Bereitschaft, sich auf neue Sachverhalte einzulassen sowie lebenslang zu lernen.

Auch in Bildungspolitik und Forschung rückt die Auffassung, dass schulische Bildung mehrere Ziele verfolgt, (wieder) verstärkt in den Mittelpunkt der Betrachtung (Blossfeld et al., 2015; Danner, Lechner & Spengler, 2019). Bereits in den achtziger Jahren betonten Treiber und Weinert (1982), dass die Frage nach dem Erreichen multipler Zielkriterien eine der theoretisch spannendsten und gleichzeitig praktisch bedeutsamsten Fragen darstellt. Dementsprechend wurde die Forderung nach der Berücksichtigung verschiedener Bildungsziele auch in empirischen Untersuchungen regelmäßig wiederholt (z. B. Baumert, 1997, Möller, 2016). Auch wenn sich hier langsam ein Wandel abzeichnet, folgert Möller (2016), dass in aktuellen Untersuchungen nach wie vor selten das Erreichen multipler Ziele thematisiert wird.

Daran knüpft die vorliegende publikationsbasierte Habilitation an. Ihr Gegenstand ist es am Beispiel des *Programme for International Student Assessments* (PISA) die Berücksichtigung und Erfassung mehrdimensionaler Bildungsziele in groß angelegten Schulleistungsstudien (Large-Scale Assessments) zu betrachten, diese empirisch für Deutschland in den internationalen Vergleich einzuordnen und darüber hinaus den Zusammenhang mit Effekten des Kontextes wie der Institution oder der Klassenkomposition sowie Qualitätsmerkmalen des Unterrichts als zentrale Bedingungen für das Erreichen mehrdimensionaler

Bildungsziele zu untersuchen. Dabei werden vor allem die Fächer Mathematik und Naturwissenschaften in den Blick genommen. Beide Bereiche sind für die aktive und selbständige gesellschaftliche Teilhabe an der heutigen durch Globalisierung, technologischen Wandel und Digitalisierung geprägten Kultur von zentraler Bedeutung. Außerdem bieten Berufe in den Bereichen Mathematik, Informatik, Naturwissenschaften und Technik (MINT) gute Zukunftschancen und es gilt, fähige Jugendliche für diese Bereiche zu interessieren (vgl. Schiepe-Tiska, Simm & Schmidtner, 2016).

Die Arbeit verknüpft erziehungswissenschaftliche und pädagogisch-psychologische Perspektiven der Schul- und Unterrichtsforschung miteinander. Sie setzt sich zum einen mit normativ durchwirkten Rahmenkonzeptionen internationaler Vergleichsstudien wie PISA auseinander, die für die Auswahl und Operationalisierung mehrdimensionaler Ziele in Large-Scale Assessments wichtig sind. Zum anderen wird eine stärker pädagogisch-psychologische Perspektive eingenommen, wenn qualitätsvolle Lehr-Lernprozesse im Hinblick auf unterschiedliche Bildungsziele untersucht werden. Durch die Fokussierung auf Mathematik und Naturwissenschaften wird darüber hinaus die jeweilige fachdidaktische Perspektive mitberücksichtigt.

1.1. Mehrdimensionale Bildungsziele im Kontext des deutschen Bildungsverständnisses und allgemeiner internationaler Grundbildungskonzepte

Der Begriff mehrdimensionale Bildungsziele verweist bereits auf die normative Dimension des deutschen Bildungssystems. Die Frage danach, welche Normen erfüllt und Ziele erreicht werden sollen, steht unter Berücksichtigung der aktuellen kulturellen und gesellschaftlichen Verhältnisse im Mittelpunkt bildungspolitisch-ethischer Diskurse. Bildung als ein in der Person ablaufender Prozess des Sich Herausbildens eines zeitlich überdauernden Selbst- und Wertebewusstseins (vgl. Seel & Hanke, 2015), impliziert bereits die Annahme eines mehrdimensionalen Bildungsverständnisses, welches in der Tradition nach Humboldt (1903) sowohl individuelle als auch gesellschaftliche Erfordernisse erfüllt: die Entwicklung einer verhaltenssicheren und lebensfähigen Persönlichkeit, die sich nicht nur an ihren

individuellen Interessen orientiert, sondern zeitgleich ihren Beitrag zu einem Zusammenleben in der globalisierten Welt im Sinne von aktiver Teilhabe leistet (Blossfeld et al., 2015). Demnach ist die Idee einer mehrdimensionalen Allgemeinbildung, deren Inhalte nicht primär auf den Lebensvollzug und die Verwendbarkeit ausgerichtet sind, fester Bestandteil des deutschen Schulkanons und wird in Schulgesetzen, Lehrplänen und Schulprogrammen explizit aufgegriffen.

Die Teilnahme an internationalen Vergleichsstudien wie PISA um die Jahrtausendwende hat in Deutschland erheblich dazu beigetragen, erneut die Frage zu stellen, welche Lernergebnisse am Ende eines Schulabschnittes erzielt werden sollen. Das für viele überraschende, im internationalen Vergleich schlechte Abschneiden Deutschlands führte zu einem Perspektivwechsel im Bildungswesen – weg von einer Input- zu einer stärkeren an Bildungsergebnissen orientierten Betrachtungsweise, die eine systematische und Indikatoren gestützte Beobachtung des deutschen Bildungssystems erlaubt. Ziel der Teilnahme an Vergleichsstudien ist es, professionellen Akteuren, Verantwortlichen und der Öffentlichkeit ein, in den internationalen Vergleich eingebettetes (Benchmarking), realistisches Bild von tatsächlichen, an Standards orientierten (Monitoring), Lernergebnissen zu vermitteln (Seidel & Prenzel, 2008). Die Lernergebnisse beschränken sich dabei nicht nur auf (Fach-) Wissen, sondern beziehen motivationale Orientierungen, Selbstbilder und Einstellungen mit ein.

Die Grundbildungskonzepte, die den internationalen Vergleichsstudien zugrunde liegen, versuchen unabhängig von nationalen Curricula oder Bildungssystemen einen international vergleichbaren Standard von Basiskompetenzen zu definieren, die zu einer umfassenden Persönlichkeitsbildung und einer anschlussfähigen Grundbildung von Jugendlichen beitragen. Konzeptionen von Grundbildung orientieren sich am angelsächsischen *Literacy*-Konzept und beschreiben funktionale Merkmale und Ansprüche, die für die derzeitige und zukünftige gesellschaftliche Teilhabe Jugendlicher anwendbar sowie anschlussfähig für kontinuierliches, lebenslanges Lernen sind und von allen Schülerinnen und Schülern erreicht werden sollen. Dabei wird immer auch eine normative Perspektive einbezogen und die Frage gestellt, welche zentralen Gegenstände der Kultur wichtig für das aktuelle Weltverstehen sind (vgl. Baumert, Stanat & Demmrich, 2001).

Die Rahmenkonzeptionen von PISA

PISA selbst verfolgt den Ansatz, relevante Aspekte von Grundbildung mit Blick auf die Allgemeinbildung zu definieren (Identifikation anschlussfähiges Wissen in der Rahmenkonzeption) und zu erfassen (Umsetzung in konkrete Aufgaben/Testentwicklung). Die Definition der relevanten Aspekte spiegelt sich in pragmatischen und zugleich elaborierten Rahmenkonzeptionen der berücksichtigten Domänen wider. Die Studie fokussiert dabei einen Ausschnitt aus dem Spektrum allgemeiner Bildung und allgemeinbildender Schulfächer und begann 2000 zunächst die Fachbereiche Lesen, Mathematik und Naturwissenschaften in den Blick zu nehmen. Neben der Lesekompetenz als Befähigung zur Teilhabe an einer Kultur, deren Wissen in Texten vorliegt, ermöglicht ein sicherer Umgang mit mathematischen Symbolen und Modellen, welche die Grundlagen für eine formalisierte, gemeinsame Weltsprache bilden, eine aktive Teilhabe an der durch Globalisierung und technischen Wandel geprägten Gesellschaft. Analog dazu werden ebenfalls naturwissenschaftlich-technische Basiskompetenzen als Kulturwerkzeuge betrachtet (OECD, 2013a, 2017). Mathematik und Naturwissenschaften weisen eine Anwendungsbreite auf, die über die entsprechenden Schulfächer hinausgeht und relevante Aspekte von Grundbildung erfasst, die im Sinne einer Allgemeinbildung zu verstehen sind.

Die Ausarbeitung der Rahmenkonzeption wird durch die Fragen geleitet, welches konzeptuelle und prozedurale Wissen für die gesellschaftliche Teilhabe bedeutsam sind und in welchen Situationen dieses Wissen genutzt werden sollte. Diese Wissensanforderungen werden im jeweiligen Domänenkontext untersucht und auf unterschiedlich anspruchsvollen Niveaus definiert. In diesem Zusammenhang spielt der Kompetenzbegriff eine wichtige Rolle. Kompetenzen werden definiert als „(...) bei Individuen verfügbare oder von ihnen erlernbare kognitive Fähigkeiten und Fertigkeiten, bestimmte Probleme zu lösen, sowie die damit verbundenen motivationalen, volitionalen und sozialen Bereitschaften und Fähigkeiten, die Problemlösungen in variablen Situationen erfolgreich und verantwortungsvoll nutzen zu können“ (Weinert, 2002, S. 27). PISA orientiert sich dabei an der Funktion von Kompetenzen im kulturellen und gesellschaftlichen Zusammenhang und konzentriert sich auf solche Kompetenzen, die bezugnehmend auf die aktuelle empirische Befundlage als notwendige Voraussetzungen für weiterführende Lern- und

Bildungsprozesse angesehen werden (Prenzel, Carstensen, Frey, Drechsel & Rönnebeck, 2007). Die in den Rahmenkonzeptionen entwickelten Kompetenzmodelle beschreiben Strukturen, Ausprägungen und den Aufbau der entsprechenden Kompetenzen. Die Kompetenzmodelle zeichnen sich dabei bereits durch ein mehrdimensionales Bildungsverständnis aus, denn sie nehmen vor allem solche Lernergebnisse in den Blick, die für eine problembewusste Wissensanwendung wichtig sind und die Persönlichkeit des Einzelnen bilden. Die Auswahl der Lernergebnisse ist jedoch bisher stark durch Wissensaspekte geprägt. Darüberhinausgehende fachspezifische Orientierungen, Selbstbilder und Verhaltensweisen sowie domänenübergreifende Persönlichkeitsmerkmale werden aktuell eher selektiv berücksichtigt, gewinnen jedoch immer stärker an Bedeutung und sollen eine eng gefasste, rein kognitive Leistungsorientierung überwinden (OECD, 2018).

Die Berücksichtigung unterschiedlicher Bildungsziele wird bei PISA in einer Gegenüberstellung kognitiver vs. nicht-kognitiver Lernergebnisse systematisiert (OECD, 2017). Fachliche und übergeordnete Bildungsziele beeinflussen sich wechselseitig und sagen gemeinsam Erfolg und Lebenszufriedenheit vorher (vgl. z. B. Almlund, Duckworth, Heckman & Kautz, 2011; Murphy & Hall, 2011; Pinquart & Sörensen, 2000; Robbins, Lauver, Le, Davis, Langley & Carlstrom, 2004). Nicht-kognitive Lernergebnisse dienen zum einen als Prädiktoren für Leistungsunterschiede; sie sind zum anderen aber auch als eigenständige Zielbereiche von Schule und Unterricht zu verstehen. Die Rahmenkonzeptionen für Mathematik und Naturwissenschaften sprechen fachbezogenen motivationalen Orientierungen, Einstellungen und Verhaltensweisen eine wichtige Rolle zu, sich produktiv mit dem jeweiligen Inhalt auseinanderzusetzen (OECD, 2013a, 2017). Diese Lernergebnisse beeinflussen neben dem deklarativen und prozeduralen Wissen, ob und wie Jugendliche ihre Kompetenzen in spezifischen Problemsituationen nutzen.

Grundbildungsmodelle wie das von PISA bieten somit die Möglichkeit, normativ definierte, mehrdimensionale Bildungsziele konzeptuell zu fassen und diese auf vorhandene, in der Forschung belegte, theoretische Konstrukte zu beziehen und am Ende die Ergebnisse mit Blick auf diese Kompetenzmodelle zu reflektieren. Sie bilden außerdem die Grundlage für die Operationalisierung der

ausgewählten Konstrukte und zur Entwicklung internationaler Testkonzeptionen. Allerdings fokussieren auch die entwickelten Tests und die Umsetzung in konkrete Aufgaben zur empirischen Erfassung der Kompetenzen vor allem kognitiv-fachliche Aspekte, da die Entwicklung von Assessments im Bereich nicht-kognitiver Lernergebnisse noch nicht soweit fortgeschritten ist wie die Entwicklung kognitiver Tests. Bisher liegen für viele Konstrukte noch keine reliablen und validen Testverfahren vor, die auch kulturübergreifend eingesetzt werden könnten. Deshalb wird zur Messung nicht-kognitiver Lernergebnisse vorwiegend auf Fragebögen und Einschätzungsskalen zurückgegriffen. Die Rahmenkonzeptionen zu den Kontextfragebögen greifen allgemeine sowie fachspezifische emotionale und motivationale Orientierungen, Einstellungen und Verhaltensweisen noch einmal gesondert auf und sortieren sie neben Lesen, Mathematik und Naturwissenschaften als wichtige Lernergebnisse ein (OECD, 2013a, 2017).

Seit PISA 2012 gibt man zumindest den methodischen Herausforderungen der kulturvergleichenden Erfassung mehr Gewicht und versucht ihnen durch verschiedene neue Fragebogenformate wie zum Beispiel Ankervignetten oder Forced-Choice-Formaten (bisher allerdings mit mäßigem Erfolg) zu begegnen (Marksteiner, Kuger & Klieme, 2018; OECD, 2014). Durch die wiederkehrende Erhebung im Dreijahreszyklus, die zum Teil auch nicht-kognitive Lernergebnisse einschließt, kann jedoch zumindest innerhalb eines Bildungssystems die Entwicklung ausgewählter motivational affektiver Indikatoren, Selbstbilder und Einstellungen einigermaßen reliabel über die Zeit hinweg untersucht werden. Diese Betrachtungen sind unbeeinflusst durch kulturelle Einflüsse auf die Definition und Messung nicht-kognitiver Lernergebnisse.

1.2. Schulische und unterrichtliche Bedingungen des Erreichens mehrdimensionaler Bildungsziele

PISA erhebt die Ausprägung mehrdimensionale Bildungsziele 15-jähriger Schülerinnen und Schüler am Ende ihrer Pflichtschulzeit. Gerade in dieser Phase besitzen mehrdimensionale Bildungsziele einen hohen Stellenwert, da sie für Jugendliche eine entscheidende Phase der Identitätsbildung darstellt. Sie entwickeln klare Vorstellungen von der eigenen Person und klären ihre Beziehung zu anderen

Personen und der Welt im Allgemeinen. Damit gewinnt neben Fragen nach den eigenen Interessen oder Vorstellungen zur Berufswahl auch die Auseinandersetzung mit gesellschaftlicher und politischer Teilhabe an Relevanz. Dieser Zeitpunkt markiert außerdem in den meisten Ländern das Ende eines institutionalisierten Bildungs- und Erziehungsauftrags. Bis dahin bietet die Schule zahlreiche Möglichkeiten, Potenziale zu entwickeln. Umgekehrt können bis zu diesem Zeitpunkt entstandene Versäumnisse schwieriger aufgefangen werden. PISA geht dabei der Frage nach, inwieweit bis zum Testzeitpunkt schulische und außerschulische Lerngelegenheiten wahrgenommen wurden, um teilhabe- und handlungsfähig zu werden. Dies bekräftigt die besondere Bedeutung von Schule und Unterricht für das Erreichen mehrdimensionaler Bildungsziele, da gerade der Fachunterricht systematische Lerngelegenheiten bietet, sich mit Mathematik und Naturwissenschaften auseinanderzusetzen. Dabei wird die Art des Unterrichts auch durch das Zusammenspiel mit Institutionseffekten wie die besuchte Schulart (Baumert, Stanat & Watermann, 2006) oder Kompositionseffekten wie die Zusammensetzung der Klasse nach Merkmalen der Leistung sowie der sozialen Herkunft beeinflusst (Dumont, Neumann, Maaz & Trautwein, 2013). Schularten unterscheiden sich hinsichtlich der Zusammensetzung ihrer Schülerschaft, der Curricula, den zugrundeliegenden pädagogischen und didaktischen Traditionen sowie den Arbeits- und Lernbedingungen. Sie bilden daher differentielle Lern- und Entwicklungsumgebungen, die im Zusammenspiel mit dem angebotenen Fachunterricht das Erreichen mehrdimensionaler Bildungsziele beeinflussen.

In Bezug auf den Unterricht selbst spielen normative, pädagogische Prinzipien und aktuelle Standards eine wichtige Rolle und beschreiben, was einen „guten“ Unterricht ausmacht (Berliner, 2005). Beispielsweise orientiert sich ein guter Naturwissenschaftsunterricht am Modell des forschend-entdeckenden Lernens, welches auch in der PISA-Erhebung Berücksichtigung findet. Im forschend-entdeckenden Unterricht werden alltagsnahe, authentische Problemstellungen, oft in Kleingruppen, bearbeitet. Mithilfe des zur Verfügung gestellten Lehrmaterials sollen Schülerinnen und Schüler naturwissenschaftliche Prinzipien beispielhaft nachvollziehen und entsprechende Konzepte aufbauen (Bruner, 1961). Obwohl ein solcher Unterricht aus normativer Perspektive den aktuellen Vorstellungen von sinnvollem Lernen entspricht, ist dieses Unterrichtskonzept nicht unumstritten und

die empirische Befundlage in Bezug auf das Erreichen verschiedener Lernergebnisse gemischt (z. B. Alfieri, Brooks, Aldrich & Tenenbaum, 2011; Furtak, Seidel, Iverson & Briggs, 2012; Kirschner, Sweller & Clark, 2006; Lazonder & Harmsen, 2016; Schiepe-Tiska, Schmidtner, Müller, Heine, Neumann & Lüdtke, 2016). Daher ist „guter“ Unterricht nach Berliner (2005) nicht das alleinige Bewertungskriterium, sondern wird durch ein zweites, für die vorgelegten Arbeiten ebenfalls zentrales Kriterium, ergänzt, nämlich inwieweit der Unterricht das Erreichen mehrdimensionaler Bildungsziele ermöglicht („effektiver“ Unterricht). Eine der effektivsten Unterrichtsmethoden gerade für das Erreichen kognitiver Lernergebnisse ist die direkte Instruktion, bei der der Unterricht stark durch die Lehrkraft gesteuert wird, sowie eher kleinschrittig und stark vorstrukturiert ist (Hattie, 2009). Auch solche Unterrichtsmerkmale finden in PISA Berücksichtigung (OECD, 2013a, 2017). Ein Unterricht, der beide Prinzipien vereint und sowohl aktuelle Normen und Wertvorstellungen berücksichtigt, sowie gleichzeitig auf das Erreichen unterschiedlicher Lernziele fokussiert, wird als qualitativvoller Unterricht bezeichnet (Berliner, 2005). In Bezug auf das forschend-entdeckende Lernen weisen die aktuellen Forschungsergebnisse auf die Stärke der Kombination beider Kriterien hin: Ein forschend-entdeckender Unterricht, der durch die Lehrkraft strukturiert und gesteuert wird, ermöglicht das Erreichen sowohl kognitiver als auch motivational-affektiver Lernziele (Alfieri et al., 2011; Furtak, et al., 2012; Lazonder & Harmsen, 2016; Schiepe-Tiska, Schmidtner et al., 2016).

Qualitätvoller Unterricht lässt sich demnach auch dadurch charakterisieren, dass nicht einzelne Unterrichtsmerkmale, sondern deren Arrangement und Zusammenspiel für das Erreichen mehrdimensionaler Ziele entscheidend sind (Oser & Baeriswyl, 2001). Dabei haben sich in der Forschung vor allem Tiefenstrukturen als Merkmale der Qualität der Interaktion zwischen Lehrenden und Lernenden als bedeutsam erwiesen (vgl. Kunter & Ewald, 2016). In der deutschsprachigen Unterrichtsforschung hat sich im Rahmen der Teilnahme an internationalen Schulleistungsstudien eine sparsame Systematik zur Beschreibung von Tiefenstrukturen etabliert, nämlich die Basisdimensionen Klassenführung, konstruktive Unterstützung und kognitive Aktivierung (vgl. Praetorius, Klieme, Herbert & Pinger, 2018). Deren Wirksamkeit wurde vor allem in Bezug auf das Erreichen kognitiver Lernergebnisse untersucht. Mathematikleistungen Jugendlicher

sind umso besser, je mehr die Lernenden kognitiv herausgefordert werden und die Lehrperson zeitgleich eine effiziente Klassenführung durchsetzt (Klieme, Schümer & Knoll, 2001; Kunter, 2005; Rakoczy, 2008). Im Bereich nicht-kognitiver Lernergebnisse wurde vor allem Interesse untersucht, welches mit einer effizienten Klassenführung und einer positiv ausgeprägten konstruktiven Unterstützung zusammenhängt (z. B. Kunter & Voss, 2011). In der Grundschule zeigte sich für die Naturwissenschaften zusätzlich ein positiver Einfluss kognitiv aktivierenden Unterrichts auf das situationale Interesse (Fauth, Decristan, Rieser, Klieme & Büttner, 2014).

Die Erfassung schulischer und unterrichtlicher Rahmenbedingungen in PISA

In PISA erfolgt die Erfassung schulischer und unterrichtlicher Rahmenbedingungen ebenfalls mithilfe der Kontextfragebögen. Die Rahmenkonzeptionen der Fragebögen in PISA 2012 und 2015 orientiert sich an den, in der internationalen Schuleffektivitätsforschung verbreiteten, Kontext-Input-Prozess-Ergebnis-Modellen (z.B. Purves, 1987). Die Stärke dieser Modelle ist, dass sie empirisch generierte Rahmenmodelle sind, welche die Integration und Bündelung komplexer Variablengruppen ermöglichen und so versuchen, Forschungsbefunde zu systematisieren und zu vereinheitlichen. Dabei berücksichtigen sie den Mehrebenen-Charakter des Bildungssystems, ermöglichen das Zusammenspiel von Merkmalen auf unterschiedlichen Ebenen zu beschreiben und erlauben auch die Modellierung von Prozessen auf Schul- und Klassenebene. Diese vereinfachte Darstellung soll aber nicht darüber hinwegtäuschen, dass die dahinterliegenden Zusammenhänge, Wirkungen und Wechselwirkungen komplexer und komplizierter sind, wie zum Beispiel das manche Merkmale (z. B. Lernstrategien) als Input, Prozess oder Lernergebnis betrachtet werden können.

Die Rahmenkonzeption von PISA unterscheidet die Aggregationsebenen Bildungssystem, Schulen, Klassen sowie Schülerinnen und Schüler (Jude, 2016; OECD, 2013a; 2017). Individuelle Schülervoraussetzungen wie die soziale Herkunft sowie die leistungs- und herkunftsbezogene Zusammensetzung von Klassen werden als Inputkriterien eingeordnet. Auf der Bildungssystemebene verortete institutionelle Effekte wie Schulartzuweisung sowie Aspekte der Qualität des Lehrens und Lernens auf der Klassenebene werden als Prozesse für das Erreichen mehrdimensionaler

Lernergebnisse betrachtet (OECD, 2013a, 2017). In Bezug auf den Unterricht finden dabei zum einen domänenübergreifende Qualitätsmerkmale (z. B. Disziplin im Klassenzimmer, Unterstützung durch die Lehrkraft) und zum anderen für die jeweilige Schwerpunktdomäne fachspezifische Charakteristika von Unterricht (z. B. Vertrautheit mit mathematischen Begriffen, forschend-entdeckender Unterricht) Berücksichtigung, welche vor allem aus der Sicht der 15-Jährigen im internationalen Schülerfragebogen erhoben werden (Schiepe-Tiska, Reiss, Obersteiner, Heine, Seidel & Prenzel, 2013; Schiepe-Tiska, Schmidtner et al., 2016). Dem liegt die Auffassung zugrunde, dass Unterricht ein Angebot bereitstellt, welches von den Lernenden wahrgenommen und aktiv genutzt werden muss. Dieses Verständnis berücksichtigt konstruktivistische Lerntheorien, die die Bedeutung individueller Lernvoraussetzungen wie Vorwissen, Einstellungen und subjektiver Erfahrungen für den Lernprozess hervorheben.

1.3. Rahmenmodell zur Untersuchung mathematisch-naturwissenschaftlicher mehrdimensionaler Bildungsziele

Das Ziel der vorgelegten Forschungsarbeiten war es, mehrdimensionale Bildungsziele am Beispiel der Mathematik und den Naturwissenschaften im internationalen Vergleich zu betrachten, sowie deren schulische und unterrichtliche Rahmenbedingungen zu untersuchen. Dieses Ziel wurde im Rahmen groß angelegter Schulleistungsstudien, insbesondere der PISA-Studien 2012 und 2015, verfolgt. Abbildung 1 systematisiert aufbauend auf dem Rahmenmodell schulischer und außerschulischer Lernumwelten von Doll und Prenzel (2003) die in der vorliegenden Arbeit untersuchten Merkmale und Zusammenhänge. Dieses Rahmenmodell wurde als Grundlage für die vorliegende Arbeit ausgewählt, da es verschiedene Ansätze aus der Unterrichts-, Schulentwicklungs- und Sozialisationsforschung aufgreift und sich dabei an der Rahmenkonzeption von PISA orientiert. Es unterscheidet verschiedene miteinander vernetzte Lernumwelten, die direkt oder indirekt zum einen die individuellen Lernvoraussetzungen und Verarbeitungsprozesse von Schülerinnen und Schülern beeinflussen und zum anderen auf das Unterrichtsgeschehen und die Unterrichtsprozesse wirken. Im Zusammenspiel beeinflussen diese Merkmale die Ausprägung und Entwicklung

fachlicher sowie fachübergreifender kognitiver und nicht-kognitiver Lernergebnisse (Doll & Prenzel, 2003).

Die vorgelegten Forschungsarbeiten fokussieren im Bereich der kognitiven Lernergebnisse vor allem auf die Ausprägung und Entwicklung mathematischer und naturwissenschaftlicher Kompetenz. Im Bereich nicht-kognitiver mathematisch-naturwissenschaftlicher Lernergebnisse werden vor allem das fachbezogene Interesse, Selbstwirksamkeitserwartungen sowie Berufserwartungen in den Blick genommen. Die einzelnen Beiträge untersuchen darüber hinaus unterschiedliche Aspekte der Lernumwelten wie die besuchte Schulart oder die leistungs- und herkunftsbezogene Zusammensetzung einer Klasse sowie das Zusammenspiel mit domänenübergreifenden (Disziplin im Klassenzimmer, konstruktive Unterstützung) und fachspezifischen (kognitive Aktivierung in der Mathematik, forschend-entdeckender Unterricht in den Naturwissenschaften) Merkmalen der Unterrichtsqualität. Die Arbeiten gliedern sich in zwei Teilbereiche, die sowohl auf konzeptioneller als auch empirischer Grundlage bearbeitet werden:

1. Die Ausprägung und Entwicklung mehrdimensionaler mathematisch-naturwissenschaftlicher Bildungsziele sowie deren Zusammenspiel
2. Schulische und unterrichtliche Bedingungen des Erreichens mehrdimensionaler mathematisch-naturwissenschaftlicher Bildungsziele

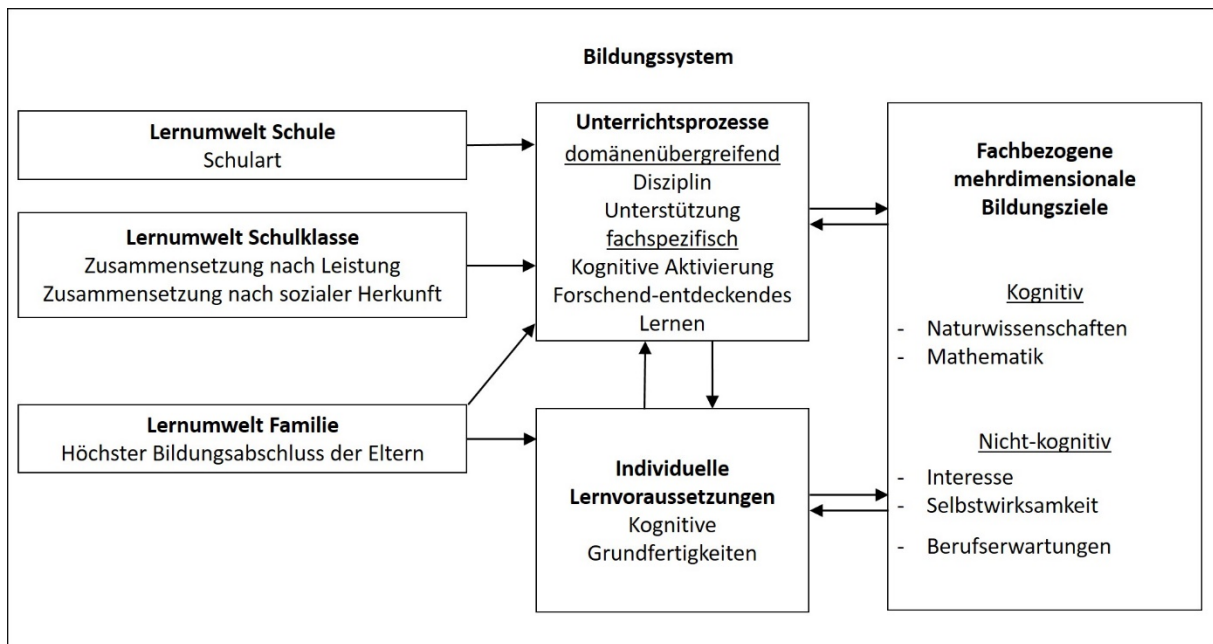


Abbildung 1. Rahmenmodell der Habilitation

Disziplinär ist der Forschungsgegenstand einschlägig für die empirische Bildungsforschung und darin insbesondere für die Erziehungswissenschaft und die Pädagogische Psychologie. Das ausgewählte Forschungsthema deckt somit die Breite der in der Habilitation angestrebten Venia Legendi in Erziehungswissenschaft/Pädagogischer Psychologie ab.

Im Folgenden werden für jeden der beiden Teilbereiche die einzelnen Forschungsarbeiten zusammengefasst und ihr Erkenntnisgewinn abgeleitet. Dabei wird auch kurz auf Besonderheiten und Herausforderungen des Designs von PISA als Large-Scale Assessment eingegangen und es werden in Deutschland als Ergänzung durchgeführte nationale Studien skizziert, die zum Teil Grundlage für die vorliegenden Arbeiten waren.

2. Zusammenfassung der Forschungsarbeiten und ihres Erkenntnisgewinns

Die Habilitation setzt sich aus einer Reihe von Fachpublikationen¹ zusammen, die in Tabelle 1 zusammengefasst sind. Die ersten fünf Beiträge sind der Fragestellung zugeordnet, wie mehrdimensionale Bildungsziele in Deutschland im internationalen Vergleich ausgeprägt sind, wie sie sich über die Zeit entwickeln und welche Zusammenhänge zwischen fachlichen Komponenten und motivational-affektiven Merkmalen sowie fachbezogenen Einstellungen bestehen. Sechs weitere Beiträge lassen sich der zweiten Fragestellung zuordnen, welche übergreifenden und fachspezifischen Qualitätsmerkmale von Schule und Unterricht mit dem Erreichen mehrdimensionaler Bildungsziele zusammenhängen.

Im Folgenden werden diese beiden Forschungsbereiche kurz skizziert und die Ergebnisse der zugehörigen Publikationen sowie ihr Erkenntnisgewinn zusammengefasst. Darüber hinaus werden eigene weiterführende Forschungsarbeiten umrissen. Die eigene Leistung im Kontext der Fachpublikationen wird im anschließenden Abschnitt beschrieben.

¹ Gemäß § 9 (2) 2 der Habilitationsordnung der Technischen Universität München vom 9. Dezember 2003, in der Fassung der Änderungssatzung vom 13. Dezember 2005

Tabelle 1. Publikationen im Kontext der Habilitation

*Forschungsbereich I – Ausprägung, Entwicklung und Zusammenspiel
mehrdimensionaler Bildungsziele*

1. Schiepe-Tiska, A., Roczen, N., Müller, K., Prenzel, M. & Osborne, J. (2016). Science-related outcomes: attitudes, motivation, value beliefs, strategies. In Kuger, S., Klieme, E., Jude, N., & Kaplan, D. (Eds.), *Assessing context of learning. An international perspective* (S. 301-330). New York: Springer.
2. Schiepe-Tiska, A., Simm, I. & Schmidtner, S. (2016). Motivationale Orientierungen, Selbstbilder und Berufserwartungen in den Naturwissenschaften in PISA 2015. In K. Reiss, C. Sälzer, A. Schiepe-Tiska, E. Klieme & O. Köller (Hrsg.), *PISA 2015. Eine Studie zwischen Kontinuität und Innovation* (S. 99-132). Münster: Waxmann.
3. Schiepe-Tiska, A. & Schmidtner, S. (2013). Mathematikbezogene emotionale und motivationale Orientierungen, Einstellungen und Verhaltensweisen von Jugendlichen in PISA 2012. In M. Prenzel, C. Sälzer, E. Klieme, & O. Köller (Hrsg.), *PISA 2012. Fortschritte und Herausforderungen in Deutschland* (S. 99-122). Münster: Waxmann.
4. Schiepe-Tiska, A., Rönnebeck, S., Schöps, K., Neumann, K., Schmidtner, S., Parchmann, I. & Prenzel, M. (2016). Naturwissenschaftliche Kompetenz in PISA 2015 – Ergebnisse des internationalen Vergleichs mit einem modifizierten Testansatz. In K. Reiss, C. Sälzer, A. Schiepe-Tiska, E. Klieme & O. Köller (Hrsg.), *PISA 2015. Eine Studie zwischen Kontinuität und Innovation* (S. 45-98). Münster: Waxmann.
5. Schiepe-Tiska, A., Rönnebeck, S., Heitmann, P., Schöps, K., Prenzel, M. & Nagy, G. (2017). Die Veränderung der naturwissenschaftlichen Kompetenz von der 9. zur 10. Klasse bei PISA und den Bildungsstandards unter Berücksichtigung geschlechts- und schulartspezifischer Unterschiede sowie der Zusammensetzung der Schülerschaft. *Zeitschrift für Erziehungswissenschaft*, 20, 151–176. <https://doi.org/10.1007/s11618-017-0754-2>

6. Müller, K., Prenzel, M., Seidel, T., Schiepe-Tiska, A., & Kjaernsli, M. (2016). Science teaching and learning in schools. In S. Kuger, E. Klieme, N. Jude & D. Kaplan (Hrsg.), *Assessing context of learning. An international perspective* (S. 423-446). New York: Springer.
 7. Schiepe-Tiska, A., Schmidtner, S., Müller, K., Heine, J.-H., Neumann, K. & Lüdtke, O. (2016). Naturwissenschaftlicher Unterricht in Deutschland in PISA 2015 im internationalen Vergleich. In K. Reiss, C. Sälzer, A. Schiepe-Tiska, E. Klieme & O. Köller (Hrsg.), *PISA 2015. Eine Studie zwischen Kontinuität und Innovation* (S. 133-176). Münster: Waxmann.
 8. Schiepe-Tiska, A., Reiss, K., Obersteiner, A., Heine, J.-H., Seidel, T. & Prenzel, M. (2013). Mathematikunterricht in Deutschland: Befunde aus PISA 2012. In M. Prenzel, C. Sälzer, E. Klieme & O. Köller (Hrsg.), *PISA 2012. Fortschritte und Herausforderungen in Deutschland* (S. 123-154). Münster: Waxmann.
 9. Schiepe-Tiska, A., Heine, J.-H., Lüdtke, O., Seidel, T. & Prenzel, M. (2016). Mehrdimensionale Bildungsziele im Mathematikunterricht und ihr Zusammenhang mit Basisdimensionen der Unterrichtsqualität. *Unterrichtswissenschaft*, 44, 211-225. <https://doi.org/10.3262/UW1603211>
 10. Schiepe-Tiska, A. (2019). School Tracks as Differential Learning Environments Moderate the Relationship between Teaching Quality and Multidimensional Learning Goals in Mathematics. *Frontiers in Education*, 4, 4. <https://doi.org/10.3389/educ.2019.00004>
 11. Kuger, S., Klieme, E., Lüdtke, O., Schiepe-Tiska, A. & Reiss, K. (2017). Mathematikunterricht und Schülerleistung in der Sekundarstufe. Zur Validität von Schülerbefragungen in Schulleistungsstudien. *Zeitschrift für Erziehungswissenschaft*, 20, 61–98. <https://doi.org/10.1007/s11618-017-0750-6>
-

2.1. Forschungsbereich I - Ausprägung, Entwicklung und Zusammenspiel mehrdimensionaler Bildungsziele

Als wichtige Facette der vergleichenden Erziehungswissenschaft ermöglichen internationale Vergleichsstudien wie PISA ausgewählte Indikatoren mehrdimensionaler Bildungsziele in verschiedenen Ländern zu untersuchen und miteinander zu vergleichen (vgl. Seel & Hanke, 2015). Im Mittelpunkt der vergleichenden Analyse stehen einerseits Unterschiede und Besonderheiten in der Ausprägung mehrdimensionaler Bildungsziele (melioristische Funktion) sowie andererseits Gemeinsamkeiten und allgemeine Trends (evolutionistische Funktion) (Hörner, 1997). Als Querschnittsstudie angelegt, ermöglicht PISA zunächst eine Momentaufnahme über das Erreichen mehrdimensionaler Bildungsziele am Ende der Pflichtschulzeit im internationalen Vergleich. Im Rahmen dessen können auch geschlechts-, herkunfts- sowie national schulartspezifische Disparitäten untersucht werden. Durch die wiederkehrende Erhebung im Dreijahreszyklus ermöglicht PISA außerdem, Veränderungen mehrdimensionaler Bildungsziele über die Zeit im Sinne von Trendanalysen auf der Ebene der Population zu untersuchen, die innerhalb eines Landes unbeeinflusst durch kulturelle Einflüsse der Messung sind. Darüber hinaus ermöglicht eine als nationale Ergänzungsstudie zu PISA 2012 durchgeführte Längsschnittstudie (Paneldesign), die Untersuchung der Veränderung der mathematischen und naturwissenschaftlichen Kompetenz am Ende der zehnten Klassenstufe (vgl. Reiss, Klieme, Köller & Stanat, 2017). Diese Studie lässt außerdem durch die überlappende klassenbasierte Stichprobe mit dem Ländervergleich 2012 (Pant, Stanat, Schroeders, Roppelt, Siegle & Pöhlmann, 2013) vergleichende Analysen zwischen der fächerübergreifend angelegten naturwissenschaftlichen Kompetenz in PISA und der stärker curriculumsorientierten Kompetenz, wie sie in den Bildungsstandards erfasst wird, zu.

Die vorgelegten Arbeiten fokussieren demnach drei Fragestellungen:

a) welche Gemeinsamkeiten und Unterschiede in den mathematisch-naturwissenschaftlichen kognitiven und nicht-kognitiven Lernergebnissen zeigen sich für Deutschland im Vergleich zu anderen Bildungssystemen, b) gibt es geschlechts- und schulartspezifische Disparitäten in der Ausprägung mehrdimensionaler Lernergebnisse und c) wie verändern sich mehrdimensionale

Bildungsziele über die Zeit (Trend- und Panelanalysen) auch unter der Berücksichtigung der Determinanten Geschlecht, Schulart sowie Kompositionseffekten der Leistung und sozialen Herkunft. Darüber hinaus erfolgt im Rahmen der Beiträge jeweils eine kritische Betrachtung des zugrundeliegenden Grundbildungsmodells in Bezug auf die Berücksichtigung normativ definierter mehrdimensionaler Bildungsziele sowie deren Operationalisierung und Erfassung im Test oder Fragebogen. Die einzelnen Beiträge zielen darauf ab, erstens konzeptionell den aktuellen Stand zu fachbezogenen motivational-affektiven Bildungszielen und Einstellungen herauszuarbeiten und weiterzuentwickeln, wichtige Ziele, die im Rahmen von PISA erhoben werden können, auszuwählen, sowie Skalen für deren Operationalisierung zu entwickeln (Aufsatz 1). Zweitens sollte eine vertiefte Beschreibung und Analyse mehrdimensionaler Bildungsziele erfolgen, die den Blick klar auf Deutschland spezifische Besonderheiten lenkt, diese unter Berücksichtigung der deutschen Bildungspraxis und Bildungspolitik diskutiert und einordnet sowie offene Handlungsfelder aufzeigt (Aufsatz 2, 3 und 4). Dabei lag ein spezieller Fokus auf der Veränderung kognitiver und nicht-kognitiver Lernergebnisse, welche gerade bei PISA 2015 auf der Basis eines modifizierten Testansatzes – der Umstellung von papierbasiertem auf computerbasiertes Testen – (kritisch) betrachtet wurde. Für die Naturwissenschaften wurde zusätzlich zu den Tendaussagen der Verlauf der PISA-Kompetenz innerhalb eines Schuljahres von der neunten zur zehnten Klasse unter Berücksichtigung der Determinanten Geschlecht und Schulart sowie der leistungs- und herkunftsbezogenen Komposition der Klasse untersucht und zur Veränderung der stärker am Curriculum orientierten Kompetenz der Bildungsstandards in Bezug gesetzt (Aufsatz 5).

Damit leisten die, für die Habilitation, vorgelegten Arbeiten (1) einen Beitrag zur Konzeptionierung und Konsolidierung theoretischer Ansätze zu mehrdimensionalen Bildungszielen in internationalen Vergleichsstudien sowie deren Operationalisierung und vertiefen (2) durch empirische Analysen den Kenntnisstand zur Ausprägung und Entwicklung sowie zum Zusammenspiel relevanter Indikatoren mehrdimensionaler Bildungsziele in Deutschland.

Aufsatz 1 (Schiepe-Tiska, Roczen, Müller, Prenzel & Osborne, 2016) setzt an der internationalen Rahmenkonzeption für die Naturwissenschaften in PISA an und befasst sich konzeptionell mit der Frage, wie die dort normativ definierten motivationalen Orientierungen und Einstellungen in konkrete Ziele und Indikatoren übersetzt und operationalisiert werden können. Ziel des Beitrags war, einen theoretischen Rahmen zur Beschreibung bildungsrelevanter Indikatoren fachspezifischer nicht-kognitiver Lernergebnisse zu entwickeln und darauf aufbauend Indikatoren zu identifizieren und zu operationalisieren, die sich für eine Erfassung im Rahmen internationaler Schulleistungsstudien wie PISA eignen. Der Aufsatz fasst den aktuellen internationalen Forschungsstand zu naturwissenschaftsbezogenen motivational-affektiven Orientierungen, Selbstbildern und Einstellungen zusammen. Er berücksichtigt dabei die große Heterogenität theoretischer und disziplinärer Zugänge und versucht, verschiedene theoretische Ansätze zu integrieren und systematisch zu verbinden. Außerdem wird auf den Begriff „nicht-kognitive“ Lernergebnisse eingegangen und das Verhältnis unterschiedlicher Bildungsziele zueinander thematisiert. Unter der Berücksichtigung spezifischer Besonderheiten und Grenzen von Large-Scale Assessments werden Auswahlkriterien definiert und die theoretischen Konstrukte dahingehend geprüft, inwieweit sie sich als Indikatoren für eine international vergleichende Betrachtung in PISA eignen. Der Erkenntnisgewinn wird anschließend dahingehend zusammengefasst, konkrete Vorschläge für die konzeptuelle Präzisierung und Auswahl fachspezifischer motivationaler Orientierungen, Selbstbilder und Einstellungen für PISA zu unterbreiten. Darauf aufbauend wurden Skalen entwickelt, die im Feldtest von PISA 2015 zum Einsatz kamen (vgl. Kuger, Jude & Klieme, 2016). Der vorgeschlagene Ansatz wurde in der Hauptstudie von der OECD umgesetzt und liefert die Grundlage für die internationale (OECD, 2016) und nationale Berichterstattung (Reiss, Sälzer, Schiepe-Tiska, Klieme & Köller, 2016). Insgesamt leistet der Aufsatz einen wichtigen Beitrag zur systematischen sowie stärker theoriebasierten Auswahl und Operationalisierung nicht-kognitiver Lernergebnisse in internationalen Large-Scale-Assessments.

In eigenen weiterführenden Arbeiten werden darüber hinaus vorgeschlagene Indikatoren untersucht, die zwar international nicht in der Hauptstudie zum Einsatz kamen, jedoch als nationale Ergänzung in Deutschland umgesetzt werden konnten

(z. B. schulfachbezogenes Interesse, Umgang mit Ungewissheit). Des Weiteren wurden entwickelte Skalen im Rahmen eines parallel laufenden Forschungsprojektes der TÜV Süd Stiftung² eingesetzt.

Aufsatz 2 (Schiepe-Tiska, Schmidtner et al., 2016) knüpft direkt an den ersten Beitrag an und greift die dort identifizierten Indikatoren sowie die entwickelten Skalen auf. Der Aufsatz geht den Fragestellungen nach, wie in Deutschland naturwissenschaftsbezogene nicht-kognitive Lernergebnisse im Vergleich zu anderen Staaten ausgeprägt sind, ob es differentielle Unterschiede in Bezug auf das Geschlecht und die Schulart gibt und inwieweit sich diese Merkmale zwischen PISA 2006 und 2015 verändert haben. Darüber hinaus wird der Zusammenhang kognitiver und nicht-kognitiver Lernergebnisse mit naturwissenschaftsbezogenen Berufserwartungen untersucht. Dabei wurde bewusst auf einen Vergleich mit allen Staaten, wie die OECD (2016) ihn berichtet, verzichtet, um mögliche kulturspezifische Einflüsse bei der Beantwortung von Einschätzungsskalen zu berücksichtigen (vgl. van de Vijver & He, 2016). Stattdessen wurde eine Reihe von Staaten ausgewählt, deren Kulturräume gewisse Ähnlichkeiten mit Deutschland aufweisen.

Sowohl der Vergleich mit anderen Bildungssystemen als auch die Entwicklung über die Zeit machten deutlich, dass es in Deutschland einige Herausforderungen zu meistern gilt. Freude und Interesse an den Naturwissenschaften sowie die instrumentelle Motivation waren im Vergleich zu anderen Staaten unterdurchschnittlich ausgeprägt und nahmen darüber hinaus seit PISA 2006 ab. Dies gilt sowohl für Jugendliche an den Gymnasien als auch in den nicht gymnasialen Schularten. Auch bei den Selbstwirksamkeitserwartungen, welche zwar im internationalen Durchschnitt lagen, war im Trend eine Abnahme gerade bei den Mädchen zu verzeichnen. Ausgeprägte Geschlechterunterschiede zugunsten der Jungen zeigten sich auch bei den Berufserwartungen. Zwar nahm die Tendenz einen naturwissenschaftlichen Beruf zu ergreifen im Vergleich zu PISA 2006 zu, sie lag jedoch noch immer unterhalb des OECD-Durchschnitts. Weiterführende Analysen des Zusammenspiels mehrdimensionaler Lernergebnisse zeigten, dass für die Berufswartung weniger die naturwissenschaftliche Kompetenz entscheidend ist,

² Projekt „Ergonomie für Schulen – Ein „Koffer der TÜV-Süd-Stiftung zur Verbesserung der Arbeitsplätze von Schülerinnen und Schülern“ (Projektnummer S212/10025/2012)

sondern Freude und Interesse sowie die instrumentelle Motivation wichtige Prädiktoren sind. Im Einklang mit anderen Forschungsarbeiten identifiziert der Beitrag folgende Handlungsfelder: (a) die Notwendigkeit einer stärkeren Berücksichtigung geschlechterspezifischer Unterschiede in den naturwissenschaftsbezogenen motivational-affektiven Lernergebnissen, (b) die bessere Förderung gerade kompetenzstarker Jugendlicher sowie (c) den Ausbau des bestehenden Angebots der Berufsinformation und –orientierung an Schulen.

Im Rahmen eines von mir betreuten Promotionsvorhabens (Doroganova, 2019) wird aktuell der Frage nachgegangen, wie realistisch die Berufserwartungen 15-Jähriger in Deutschland gemessen an ihren Bildungsaspirationen sind und wie diese mit schulischen und unterrichtlichen Rahmenbedingungen zusammenhängen. Dabei sollen auch Mediationsmodelle, die den vermittelnden Zusammenhang über andere motivational-affektive Lernergebnisse in den Blick nehmen, geprüft werden.

Aufsatz 3 (Schiepe-Tiska & Schmidtner, 2013) nimmt emotionale und motivationale Orientierungen, Selbstbilder sowie Einstellungen und Verhaltensweisen in der Domäne Mathematik in PISA 2012 in den Blick. Er untersucht deutschlandspezifische Besonderheiten mathematikbezogener nicht-kognitiver Lernergebnisse im internationalen Vergleich unter der Berücksichtigung geschlechter- und schulartspezifischer Disparitäten. Darüber hinaus analysiert er die Entwicklung dieser Lernergebnisse von PISA 2003 zu PISA 2012. Wie in Aufsatz 2 wurde zur Einordnung des internationalen Vergleichs nur eine Auswahl von Staaten mit ähnlichem Kulturraum herangezogen.

Die Ergebnisse zeigten, dass sowohl Freude und Interesse als auch die instrumentelle Motivation ähnlich wie in den Naturwissenschaften im Vergleich zum OECD-Durchschnitt geringer ausgeprägt waren und im Vergleich zu PISA 2003 abnahmen. Beides ist jedoch zentral für die Bereitschaft, sich auch über den Unterricht und die Schulzeit hinaus mit Mathematik zu beschäftigen. Dabei waren ebenfalls deutliche Geschlechterunterschiede zugunsten der Jungen zu finden. Die Selbstbilder wiederum waren im internationalen Vergleich überdurchschnittlich positiv ausgeprägt und gerade die mathematikbezogenen Selbstwirksamkeitserwartungen verbesserten sich im Vergleich zu PISA 2003. Diese Zunahme ließ sich auf eine höhere Sicherheit im Umgang mit einfachen Anwendungsaufgaben zurückführen. Hier scheinen Maßnahmen wie die Einführung

der länderübergreifenden Bildungsstandards (KMK, 2003) den Anwendungsbezug im Mathematikunterricht gestärkt zu haben. Durch das häufigere Vorkommen fühlten sich Jugendliche auch sicherer im Umgang mit diesen Aufgaben.

Vertiefend zu diesen Ergebnissen wird in einem weiterführenden Projekt der Frage nachgegangen, ob es kulturspezifische Unterschiede im Beantwortungsverhalten von Einschätzungsskalen zwischen westlichen (Deutschland) und östlichen (Taiwan) Staaten gibt (Müller, Schiepe-Tiska & Strohmaier, 2017), welche die Ergebnisse beeinflussen. Dazu wurde das Blickbewegungsverhalten während des Bearbeitens der Fragebögen mithilfe eines Eye-Trackers erhoben. Ein von mir betreutes Promotionsvorhaben (Tupac-Yupanqui, Schiepe-Tiska, Weis, Heine & Reiss, 2019) beschäftigt sich darüber hinaus mit dem Vergleich unterschiedlicher Antwortformate (Ratingskalen vs. Forced-Choice), die eingesetzt wurden, um die kulturelle Vergleichbarkeit nicht-kognitiver Lernergebnisse zu erhöhen.

Die Aufsätze 4 und 5 befassen sich schwerpunktmäßig mit kognitiven Lernergebnissen – hier speziell mit der naturwissenschaftlichen Kompetenz. **Aufsatz 4** setzt sich zum einen kritisch mit der veränderten Rahmen- und Testkonzeption der naturwissenschaftlichen Kompetenz in PISA 2015 auseinander. Zum anderen ordnet der Beitrag die Ergebnisse der naturwissenschaftlichen Kompetenz für Deutschland in den internationalen Vergleich ein und untersuchte geschlechter- und schulartspezifische Disparitäten. Eine wichtige Besonderheit von PISA 2015 war, dass erstmals der Erhebungsmodus von einem papierbasierten auf einen computerbasierten Test wechselte, mit dem Ziel, dass sowohl Rahmenkonzeption als auch Test in Anbetracht der ständig voranschreitenden Entwicklungen in Naturwissenschaft und Technik ihre Aktualität und Relevanz behalten. Zudem bestand durch den Moduswechsel die Möglichkeit auch neue Aufgabenformate mit interaktiven Elementen zu entwickeln und einzusetzen. Diese Änderungen sind jedoch gerade für die Untersuchung der Veränderung naturwissenschaftlicher Kompetenz problematisch, da nicht nur eine aktualisierte Rahmenkonzeption, sondern ein veränderter Test zum Einsatz kamen. Es stellte sich daher außerdem die Frage, inwieweit die Ergebnisse mit denen vorheriger Zyklen vergleichbar sind.

Der internationale querschnittliche Vergleich zeigte, dass Deutschland sich innerhalb einer Gruppe von Staaten positioniert, die den Spitzenstaaten Japan,

Estland, Finnland und Kanada folgt. Erstmals waren Kompetenzunterschiede zugunsten der Jungen zu finden, die sich in detaillierteren Analysen auf eine homogenere Leistung in verschiedenen Wissenssystemen und einen deutlichen Vorsprung in der Teilkompetenz *Phänomene naturwissenschaftlich erklären* zurückführen ließen. Betrachtet man diese Ergebnisse in der Zusammenschau mit denen nicht-kognitiver Lernergebnisse scheint der Vorsprung der Jungen wenig überraschend, da sie auch über günstiger ausgeprägte motivationale Orientierungen und Selbstbilder berichteten. Allerdings ist die Befundlage zu Geschlechterunterschieden in den Naturwissenschaften in Deutschland nicht einheitlich. So fand der Bildungstrend 2012 vor allem Vorsprünge zugunsten der Mädchen (Schroeders, Penk, Jansen & Pant, 2013).

In Bezug auf die Entwicklung der naturwissenschaftlichen Kompetenz seit PISA 2006 ließ sich in Deutschland insgesamt keine signifikante Veränderung feststellen auch wenn die Ergebnisse zusätzlicher nationaler Analysen mit Feldtestdaten darauf hinwiesen, dass die computerbasierten Naturwissenschaftsaufgaben etwas schwieriger waren als die papierbasierten Aufgaben (Robitzsch, Lüdtke, Köller, Kröhne, Goldhammer & Heine, 2017). Darüber hinaus zeigten eine ganze Reihe Staaten, die zum Teil bei PISA 2006 Spitzenleistungen erbrachten, einen deutlichen Abfall der naturwissenschaftlichen Kompetenz. Hier scheint es Deutschland gelungen zu sein, die Ergebnisse auf einem hohen Niveau zu stabilisieren, auch wenn sich der Einfluss der veränderten Testkonzeption nicht direkt überprüfen ließ.

Differenzierte Analysen nach Schularten offenbarten, dass jedoch am Gymnasium ein Rückgang der naturwissenschaftlichen Kompetenz zu finden war. Auch wenn ungeklärt bleibt, ob dies auf die veränderte Testanlage zurückzuführen ist, weisen die Ergebnisse zu den naturwissenschaftsbezogenen Selbstwirksamkeitserwartungen auf einen weiteren möglichen Erklärungsansatz hin. Gymnasiastinnen und Gymnasiasten fühlten sich in 2015 auch weniger als bei PISA 2006 in der Lage, naturwissenschaftliche Phänomene zu erkennen, zu beschreiben und zu erklären. Andere Studien verdeutlichen den engen Zusammenhang zwischen Selbstwirksamkeitserwartungen und naturwissenschaftlicher Kompetenz (Jansen, Scherer & Schroeders, 2015) und bekräftigen, dass eine gemeinsame

Berücksichtigung und Förderung unterschiedlicher Bildungsziele im Fokus des Unterrichts stehen sollte.

Aufsatz 5 untersucht ebenfalls die Entwicklung naturwissenschaftlicher Kompetenz. Im Gegensatz zu den Trendanalysen in Aufsatz 4 wird hier jedoch die Kompetenzentwicklung im Verlauf eines Schuljahres empirisch überprüft. Dazu wurden Daten der nationalen Erweiterungsstudie PISA Plus 2012/2013 ausgewertet. Im Rahmen dieser Studie kam auch ein stärker am Curriculum orientierter Test zur Überprüfung der Bildungsstandards zum Einsatz, die in bisherigen Längsschnittstudien vor allem im Mittelpunkt der Betrachtung standen. Der Aufsatz analysiert Gemeinsamkeiten und Unterschiede in den Rahmenkonzeptionen der naturwissenschaftlichen Kompetenz in PISA und den Bildungsstandards. Ziel des Beitrags ist es, zum einen die Veränderung der breit angelegten naturwissenschaftlichen Kompetenz in PISA in Abhängigkeit der besuchten Schulart, des Geschlechts und der Klassenzusammensetzung (Kompetenz und soziale Herkunft) zu untersuchen. Zum anderen soll die Veränderung der PISA-Kompetenz in Bezug zur Veränderungen der in den Bildungsstandards verankerten und curricular orientierten Kompetenzbereiche Fachwissen und Erkenntnisgewinnung in den Fächern Biologie, Chemie und Physik gesetzt und analysiert werden.

Die Ergebnisse weisen darauf hin, dass der in den deutschen Curricula intendierte Wissensaufbau von der neunten zur zehnten Klasse nicht auf die bei PISA erhobene breite naturwissenschaftliche Kompetenz bezogen ist, sondern das lehrplanbezogene Wissen im jeweiligen Fach stärker im Blickpunkt steht. Hier scheint auch eine stärkere Homogenisierung des Wissens erreicht zu werden, die sich in einer Abnahme der Leistungsvarianz zwischen den Schulklassen widerspiegelte. Diese Befunde stehen in Einklang mit bisherigen Studien, die vor allem curricular verankerte naturwissenschaftliche Kompetenzen untersuchten (z. B. Baumert et al. 1997; Köller & Baumert 2002; National Center for Education Statistics 1995; Walter, Senkbeil, Rost, Carstensen & Prenzel, 2006).

Differenzierte Analysen zu Schulart und Geschlecht belegten für die PISA-Kompetenz Zunahmen vor allem an den Gymnasien und für Mädchen, die am Ende der zehnten Klasse den Vorsprung der Jungen aus der neunten Klasse aufholten. Der Geschlechterbefund steht im Gegensatz zu früheren Studien, die stärker

lehrplanorientierte naturwissenschaftliche Kompetenz untersuchten und eine Zunahme der Geschlechterunterschiede zugunsten der Jungen fanden (Ivanov 2011; Ivanov & Nikolova 2010). Kompositionseffekte über die besuchte Schulart hinaus zeigten sich nur in PISA und nur für das leistungsspezifische Ausgangsniveau, nicht jedoch für die soziale Herkunft. Letzteres steht in Einklang mit bisherigen Befunden, die keinen oder einen eher kleinen Effekt fanden (vgl. Dumont et al., 2013). Insgesamt weisen die Befunde darauf hin, dass der Unterricht in Deutschland eher das Ziel verfolgt, der Funktion der Bildungsstandards zu entsprechen und dafür zu sorgen, dass möglichst viele Schülerinnen und Schüler mit ihrer Kompetenz den Regelstandards gerecht werden (KMK, 2005 a, b, c). Eigene weiterführende Arbeiten beschäftigen sich mit dem Einfluss des naturwissenschaftlichen Unterrichts auf die Veränderung der PISA- beziehungsweise Bildungsstandardkompetenz.

Zusammengefasst liefern die vorgelegten theoretischen und empirischen Forschungsarbeiten wichtige Erkenntnisse zur Systematik und Operationalisierung mehrdimensionaler Bildungsziele in groß angelegten internationalen Schulleistungsstudien. Die Arbeiten liefern zudem einen wichtigen Erkenntnisgewinn zur Ausprägung, dem Zusammenspiel und der Entwicklung mehrdimensionaler Bildungsziele in Mathematik und den Naturwissenschaften in Deutschland und ordnen diese in die internationale Staatengemeinschaft ein. Besonders die Ergebnisse zu den nicht-kognitiven Lernergebnissen und der Vergleich der Entwicklung der naturwissenschaftlichen Kompetenz, wie sie in PISA und den Bildungsstandards konzeptualisiert und erfasst wird, verdeutlichen, dass Schule und Unterricht in Deutschland nach wie vor vor allem die Vermittlung von Fachwissen fokussieren. Andere Lernergebnisse scheinen, wenn überhaupt, eher implizit angestrebt oder als Nebeneffekt zum Erreichen kognitiver Lernergebnisse betrachtet zu werden. Dies ist kritisch einzuschätzen, da somit der nachhaltige Beitrag von Schule und Unterricht zu einer allgemeinen Bildung, die über eine reine Leistungsorientierung hinausgeht, die für die gesellschaftliche Teilhabe bedeutsam ist sowie auf lebenslanges Lernen vorbereitet, in Frage gestellt werden kann. Darüber hinaus zeigt das Zusammenspiel verschiedener Zielindikatoren, dass durch eine Fokussierung der Wissensaspekte Ausbildungs-, Studien- und Berufswahlentscheidungen frühzeitig eingeeengt werden, wenn Jugendliche keine

Freude und Interesse an der Auseinandersetzung mit Mathematik und Naturwissenschaften haben, sie deren Relevanz für ihre Zukunft nicht sehen und sich selbst wenig zutrauen. Die Ergebnisse betonen damit die Relevanz der Berücksichtigung breiter angelegter Zielkriterien von Schule und Unterricht und machen deutlich, wie wichtig es ist, breite Bildungsziele zu adressieren.

Ein wichtiges Ziel für die weitere Forschung und Praxis bleibt das Zusammenspiel verschiedener Bildungsziele zu untersuchen, um deren Vereinbarkeit stärker in den Blick zu nehmen. Dazu braucht es empirische Arbeiten, die sich damit befassen, wie mehrdimensionale Bildungsziele stärker in der professionellen Wahrnehmung der Lehrkräfte berücksichtigt und konkret im Unterricht umgesetzt werden können. Large-Scale Assessments wie PISA, die auch die Möglichkeit der Teilnahme an einem internationalen Lehrerfragebogen bieten, könnten Einstellungen und Verhaltensweisen von Lehrkräften im Umgang mit mehrdimensionalen Bildungszielen erfassen. Der spezifische Blick auf die Umsetzung im Unterricht muss jedoch mit ergänzenden Forschungsdesigns untersucht werden.

Mit Blick auf die Erfassung nicht-kognitiver Lernergebnisse in internationalen Large-Scale Assessments ist die Frage nach der kulturellen Vergleichbarkeit von Selbstberichten zentral. Nur wenn diese reliabel und valide vergleichbar sind, lassen sich die von der OECD (2013b, 2016) vorgelegten Tabellen sinnvoll interpretieren. Hier bietet PISA gute Anknüpfungsmöglichkeiten, die Möglichkeiten und Grenzen von Selbstberichten genauer zu untersuchen.

2.2. Forschungsbereich II – Schulische und unterrichtliche Bedingungen des Erreichens mehrdimensionaler Bildungsziele

Für das Erreichen mehrdimensionaler Zielkriterien nehmen die Schule und vor allem der Fachunterricht eine zentrale Rolle ein. Hier werden systematisch Lerngelegenheiten geboten, die nicht nur die Entwicklung kognitiver Kompetenzen, sondern auch die motivational-affektiver Lernergebnisse, Selbstbilder und Einstellungen ermöglichen. Auch PISA erhebt Merkmale auf der Ebene des Unterrichts aus der Perspektive der Schülerinnen und Schüler und ermöglicht eine

empirische kulturvergleichende Untersuchung hinsichtlich des Vorkommens domänenübergreifender und fachspezifischer Qualitätsmerkmale des Unterrichts. Weiterführende internationale Vergleichsanalysen sind jedoch dadurch eingeschränkt, dass keine klassenbasierte sondern eine auf der Schulebene gezogene altersbasierte Stichprobe erhoben wird (vgl. Heine et al., 2016). Auf Schulebene gemittelte Schülerurteile weisen jedoch zum Teil unbefriedigende Reliabilitäten auf (Wenger, Lüdtke & Brunner, 2018) und die Klasse als geteilte Lernumgebung für das Erreichen mehrdimensionaler Bildungsziele wird nicht berücksichtigt. Für solche Untersuchungen erweitert Deutschland deshalb das internationale PISA-Designs und erhebt traditionell zusätzlich eine klassenbasierte Stichprobe von Neuntklässlerinnen und Neuntklässlern, die zum Teil Grundlage für die im folgenden dargestellten Arbeiten war.

Ziel der im zweiten Forschungsbereich vorgelegten Beiträge ist es, domänenübergreifende und fachspezifische Qualitätsmerkmale des MINT-Unterrichts auf konzeptioneller sowie empirischer Ebene im Kontext internationaler Large-Scale Assessments in den Blick zu nehmen und ihre Wirkung in Bezug auf multiple Zielkriterien zu untersuchen. Dazu soll zum einen am Beispiel der Naturwissenschaften der aktuelle Stand der Unterrichtsforschung herausgearbeitet und systematisiert sowie Vorschläge zur Operationalisierung in PISA erarbeitet werden (Aufsatz 6). Zum anderen soll der Zusammenhang zwischen sowohl pädagogisch-psychologischen als auch fachspezifischen, fachdidaktischen Unterrichtsmerkmalen und mehrdimensionalen Bildungszielen empirisch untersucht werden (Aufsatz 7, 8 und 9). Dem zugrunde liegt das Verständnis, dass nicht einzelne Merkmale sondern deren Arrangement und Zusammenspiel für das Erreichen multipler Ziele entscheidend sind (Oser & Baeriswyl, 2001). Darüber hinaus wird vertiefend untersucht inwieweit der wahrgenommene Unterricht mit der besuchten Schulart als differentielle Lernumgebung interagiert und wie beides mit mehrdimensionalen Bildungszielen zusammenhängt (Aufsatz 10). Da die in PISA im Querschnitt erhobenen Unterrichtsmerkmale in der Regel nur eine begrenzte Interpretation der Wirksamkeit von Unterricht erlauben, untersucht der letzte Beitrag unter der Verwendung längsschnittlicher Daten der PISA Plus Studie 2012/2013 inwieweit die Bedeutung einzelner Unterrichtsmerkmale für die Entwicklung der mathematischen Kompetenz überschätzt wird (Aufsatz 11).

Die für die Habilitation vorgelegten Beiträge erweitern damit zum einen den Kenntnisstand des internationalen Bildungsmonitorings um Aspekte der Unterrichtswahrnehmung. Darüber hinaus ergänzen sie den Stand der empirischen Unterrichtsforschung in zweierlei Hinsicht: (a) in der Konsolidierung theoretischer Ansätze zu effektivem MINT-Unterricht sowie dessen Operationalisierung in Large-Scale Assessments und (b) in der detaillierten, anspruchsvollen Analyse der Wirkzusammenhänge mit verschiedenen Lernergebnissen.

Aufsatz 6 (Müller, Prenzel, Seidel, Schiepe-Tiska & Kjaernsli, 2016) arbeitet systematisch den aktuellen Forschungsstand der Unterrichtsforschung in den Naturwissenschaften auf. Ziel des Beitrags ist es, auf der Grundlage aktueller Forschungsarbeiten für ein Bildungssystem relevante Indikatoren des (Naturwissenschafts-) Unterrichts zu identifizieren, diese konzeptuell zu präzisieren und zu operationalisieren. Er bildet die Grundlage für die internationale (OECD, 2016) und nationale PISA-Berichterstattung (Reiss et al., 2016).

Der konzeptionell angelegte Beitrag fasst die Entwicklung von Modellen zur Unterrichtseffektivität zusammen und unterscheidet zwischen domänenübergreifenden sowie naturwissenschaftsspezifischen Unterrichtsmerkmalen. Bei letzterem lag der Schwerpunkt auf forschend-entdeckendem Lernen als eines der zentralen Unterrichtskonzepte in den Naturwissenschaften. Unter der Berücksichtigung spezifischer Besonderheiten und Grenzen von Large-Scale Assessments wurden Auswahlkriterien definiert und die theoretischen Konstrukte dahingehend geprüft, inwieweit sie sich als Indikatoren für eine international vergleichende Betrachtung in PISA eignen. In Bezug auf die Auswahlkriterien wurde neben der praktischen Relevanz, Einschränkungen durch das Fehlen einer klassenbasierten Stichprobe sowie Bildungssystem-Unterschiede im naturwissenschaftlichen Curriculum diskutiert. Der Erkenntnisgewinn wurde dahingehend zusammengefasst, dass ein Modell des naturwissenschaftlichen Unterrichtens und Lernens für PISA 2015 entwickelt wurde, das zwischen instruktionalen Kontextmerkmalen sowie naturwissenschaftsbezogenen Aktivitäten unterscheidet und diese mit verschiedenen Lernergebnissen in Zusammenhang bringt. Für die ausgewählten Indikatoren wurden Skalen geprüft und neu entwickelt, die im Feldtest zum Einsatz kamen (vgl. Kuger, Jude & Klieme, 2016). Da bei PISA

2015 zum ersten Mal international auch ein Lehrerfragebogen eingesetzt wurde, konnte sowohl die Perspektive der Schülerinnen und Schüler als auch die Perspektive der Lehrkräfte berücksichtigt werden.

Aufsatz 7 (Schiepe-Tiska, Schmidtner et al., 2016) baut auf dem vorherigen konzeptionellen Beitrag auf und untersucht die Ausprägung ausgewählter domänenübergreifender und fachspezifischer Indikatoren des deutschen Naturwissenschaftsunterrichts im internationalen Vergleich. Für Merkmale des forschend-entdeckenden Unterrichts als zentrales fachspezifisches Element des Naturwissenschaftsunterrichts (vgl. Bruner, 1961) wird darüber hinaus zum einen die Veränderung des Vorkommens dieser Merkmale im Vergleich zu PISA 2006 untersucht und zum anderen für Deutschland in Anlehnung an Kobarg und Kollegen (2011) empirisch Unterrichtsmuster gebildet, die das Zusammenspiel einzelner Facetten zeigen. Für diese Unterrichtsmuster wird sowohl der Zusammenhang mit domänenübergreifenden Unterrichtsmerkmalen als auch der Zusammenhang mit mehrdimensionalen Bildungszielen untersucht. Ähnlich wie bei den nicht-kognitiven Lernergebnissen, werden für den internationalen Vergleich Staaten mit ähnlichem Kulturraum herangezogen sowie Staaten, die zum einen unterschiedliche Unterrichtstraditionen aufweisen und denen zum anderen eine Förderung in der Breite sowie in der Leistungsspitze gelingt. Spezifisch wird auch auf Unterschiede zwischen gymnasialen und nicht gymnasialen Schularten eingegangen.

Die Ergebnisse zu den domänenübergreifenden Unterrichtsmerkmalen zeigten, dass in Deutschland Verbesserungsbedarf in der wahrgenommenen Unterstützung, der Differenzierung sowie der individuellen Rückmeldung durch die Lehrkraft besteht. Ein forschend-entdeckender Unterricht kommt häufiger vor als im OECD-Durchschnitt. In Einklang mit Befunden von Börlin (2012) spielt jedoch das Anwenden von Prinzipien auf naturwissenschaftliche Phänomene sowie das Herstellen eines Alltagsbezugs nur eine geringe Rolle. Experimente scheinen eher der Erarbeitung fachlicher Inhalte zu dienen. Dabei zeigten sich Unterschiede zwischen den Schularten: In nicht gymnasialen Schularten scheinen Experimente als didaktisches Instrument zur Untersuchung und Erklärung von Phänomenen herangezogen zu werden, wohingegen an Gymnasien das Experiment zur Theoriebildung und Prüfung von theoretisch hergeleiteten Hypothesen verwendet wird. Im Zusammenspiel zeichnete sich ein guter Naturwissenschaftsunterricht in

Deutschland ähnlich wie bei Kobarg und Kollegen (2011) durch eine Kombination hoher kognitiver Anregung und regelmäßiger „Hands-on“ Aktivitäten, wie Experimente selbstentwickeln oder strukturierte Laborexperimente durchführen, aus. In Zusammenhang mit den domänenübergreifenden Unterrichtsmerkmalen ging ein Unterricht, in dem eher strukturierte Laborexperimente durchgeführt wurden, mit einer höheren Disziplin einher. Ein Unterricht, in dem selbst Experimente entwickelt und durchgeführt wurden, zeichnete sich durch eine häufigere individuelle Rückmeldung zu Stärken und Schwächen der Jugendlichen aus. In Einklang mit Befunden aus Metaanalysen (Alfieri et al., 2011; Furtak, et al., 2012; Lazonder & Harmsen, 2016) gingen beide Unterrichtsmuster an Gymnasien mit einer hohen naturwissenschaftlichen Kompetenz und einer ausgeprägten Freude und Interesse an den Naturwissenschaften einher. Die Untersuchung der Veränderung des forschend-entdeckenden Unterrichts seit PISA 2006 weist darauf hin, dass sich der naturwissenschaftliche Unterricht in Deutschland kaum verändert hat und gerade die für mehrdimensionale Bildungsziele zentrale epistemische Lernaktivität des Schlussfolgerns (Kjærnsli & Lie, 2011; Lavonen & Laaksonen, 2009) weniger häufig vorkam. In Zusammenhang mit der fehlenden Veränderung naturwissenschaftlicher Kompetenz sowie der Verschlechterung motivational-affektiver Lernergebnisse und Einstellungen zeigte sich hier deutlicher Handlungsbedarf. Ein Naturwissenschaftsunterricht, der sich durch häufigere strukturierte und selbstständige „Hands on“ Lernaktivitäten der Schülerinnen und Schülern auszeichnet, wurde als Entwicklungsfeld, auch gestützt durch Befunde aus Metanalysen zum forschend-entdeckenden Unterricht (Alfieri et al., 2011; Furtak, et al., 2012; Lazonder & Harmsen, 2016), diskutiert.

In weiterführenden Arbeiten wird der Frage nachgegangen, wie leistungsstarke und leistungsschwache Jugendliche aus Staaten mit unterschiedlicher durchschnittlicher naturwissenschaftlicher Kompetenz den forschend-entdeckenden Unterricht in ihrem Land wahrnehmen und wie dieser mit anderen domänenübergreifenden Unterrichtsmerkmalen zusammenhängt (Forbes, Neumann & Schiepe-Tiska, 2019).

Aufsatz 8 (Schiepe-Tiska et al., 2013) befasst sich schwerpunktmäßig mit dem Mathematikunterricht und untersucht domänenübergreifende und fachspezifische Qualitätsmerkmale des Mathematikunterrichts im internationalen

Vergleich sowie deren Zusammenhang mit mehrdimensionalen Bildungszielen. Auch dieser Beitrag baut auf theoretischen Überlegungen auf, dass sich ein qualitätsvoller Unterricht durch das Arrangement und Zusammenspiel verschiedener Unterrichtsmerkmale auszeichnet (Oser & Baeriswyl, 2001) und die Lehrkraft ein Lernangebot bereitstellt, das von den Schülerinnen und Schülern auf der Grundlage ihrer individuellen Voraussetzungen und Motivation wahrgenommen und genutzt werden muss (vgl. Seidel & Reiss, 2014).

Die Ergebnisse des internationalen Vergleichs identifizierten als Handlungsfelder ähnlich wie in den Naturwissenschaften einen stärkeren Fokus auf die wahrgenommene Unterstützung durch die Mathematiklehrkraft zu legen, verstärkt individuelle Förderung anzubieten und den Unterricht konkreter an den Bedürfnissen der 15-Jährigen auszurichten. Gerade diese Unterrichtsmerkmale sind von großer Bedeutung für das Erreichen motivational-affektiver Lernergebnisse (z. B. Kunter & Voss, 2011). In Bezug auf die fachspezifischen Merkmale ließ sich der Erkenntnisgewinn dahingehend verdichten, dass trotz angestoßener Maßnahmen zur Weiterentwicklung des Mathematikunterrichts (z. B. KMK-Bildungsstandards, SINUS Programm) aus Sicht der Jugendlichen nach wie vor eher innermathematische Aufgaben oder Aufgaben mit offensichtlichen Lösungswegen bearbeitet werden. Im Zusammenspiel verschiedener Unterrichtsmerkmale zeigten die Ergebnisse, dass ein Mathematikunterricht, der sich durch ein hohes Maß an Unterstützung und kognitiver Aktivierung auszeichnete, an Gymnasien mit einer hohen Mathematikkompetenz und einer ausgeprägten Freude an Mathematik einherging. Durch den Fokus des Beitrags auf den internationalen Vergleich wurde nicht berücksichtigt, dass sich die Fünfzehnjährigen in unterschiedlichen Klassen befinden, und somit die Beobachtungseinheiten nicht unabhängig voneinander sind (vgl. Geiser, 2010). Darüber hinaus konnte nicht untersucht werden, auf welcher Ebene (Schüler- oder Klassenebene) die Unterrichtsmerkmale mit mehrdimensionalen Bildungszielen zusammenhängen.

Dieser Fragestellung geht **Aufsatz 9** (Schiepe-Tiska, Heine, Lüdtke, Seidel & Prenzel, 2016) nach. Bisherige Studien verweisen darauf, dass gerade für die mathematische Kompetenz die durch die Klasse geteilte Wahrnehmung eines störungsarmen und kognitiv aktivierenden Unterrichts entscheidend ist (vgl. Kunter, Baumert, Blum, Klusmann, Krauss & Neubrand, 2011). Für motivational-affektive

Lernergebnisse hingegen scheint eher die individuelle Wahrnehmung eines störungsarmen Unterrichts in dem die Lehrkraft individuell unterstützt, positive Auswirkungen zu haben (Rakoczy, 2008). Der Aufsatz verwendet Daten der klassenbasierten PISA-Stichprobe und untersucht den Zusammenhang zwischen den als „Basisdimensionen der Unterrichtsqualität“ bezeichneten Merkmalen kognitive Aktivierung, Disziplin im Klassenzimmer sowie konstruktiver Unterstützung und den multiplen Zielkriterien mathematische Kompetenz und Freude und Interesse an Mathematik auf der Individual- und Klassenebene. Erstmals werden zusätzlich mathematikbezogene Selbstwirksamkeitserwartungen als wichtiges Bildungsziel mitberücksichtigt.

Die Ergebnisse zur mathematischen Kompetenz sind konsistent mit bisherigen Studien, die einen Zusammenhang mit einem störungsarmen Unterricht auf der Klassenebene berichteten (vgl. Kunter, et al., 2011). Gleiches zeigte sich für die Selbstwirksamkeitserwartungen. Ein positiver Zusammenhang mit kognitiver Aktivierung konnte hingegen sowohl für die Kompetenz als auch für die Selbstwirksamkeitserwartungen nur auf der Individualebene gefunden werden, die wiederum neben der Disziplin auch wichtig für das Interesse an Mathematik der Neuntklässlerinnen und Neuntklässler war. Die Ergebnisse zu Freude und Interesse stehen in Einklang mit bisheriger Forschung, die die Bedeutung der individuellen Wahrnehmung betonen und zeigten, dass eine höher wahrgenommene Unterstützung mit einer ausgeprägteren Freude an Mathematik einherging (z. B. Rakoczy, 2008). Dieses Ergebnis ist vor allem unter der Perspektive bedeutsam, dass der internationale Vergleich eine Abnahme von Freude und Interesse im Vergleich zu PISA 2003 belegte, sowie einen Förderbedarf in der konstruktiven Unterstützung skizzierte. Unter der Perspektive des Erreichens mehrdimensionaler Bildungsziele scheint eine verstärkte individuelle Unterstützung durch die Lehrkraft zu einer Verbesserung des Mathematikunterrichts beitragen zu können.

Aufsatz 10 (Schiepe-Tiska, 2019) verfolgt das Ziel, das Zusammenspiel zwischen Qualitätsmerkmalen des Mathematikunterrichts und der besuchten Schulart im Hinblick auf mehrdimensionale Bildungsziele zu untersuchen. Schularten unterscheiden sich in Bezug auf die Zusammensetzung der Schülerschaft, ihrer Curricula, der Arbeits- und Lernbedingungen und der Ausbildung, die ihre Lehrkräfte durchlaufen (vgl. Baumert, Stanat & Watermann, 2006). Aktuelle Ergebnisse aus

PISA sowie der TIMMS-Videostudie belegen, dass es demnach auch Unterschiede in der Unterrichtsqualität zwischen Gymnasien und nicht gymnasialen Schularten gibt (Schiepe-Tiska et al, 2013; Klieme et al., 2001). Obwohl dieses Zusammenspiel gut bekannt ist, verfolgen weiterführende Studien eher den Ansatz gerade bei komplexeren Modellierungen von Zusammenhängen die Schulart nur als Kontrollvariable zu berücksichtigen, die Unterschiede aber nicht konkret zu untersuchen. Gerade mit Blick auf die Annahme, dass mit der Zuweisung von Schülerinnen und Schülern auf unterschiedliche Schularten homogenere Lerngemeinschaften entstehen, in der die individuellen Bedürfnisse stärker berücksichtigt werden können (vgl. Betts, 2011), stellt sich die Frage, ob Mathematikunterricht in den Schularten unterschiedlich zu gestalten ist, um das Erreichen mehrdimensionale Bildungsziele zu fördern. Um der Frage nachzugehen wie die Interaktion zwischen Schulart und Unterrichtmerkmalen mit mehrdimensionalen Bildungszielen zusammenhängt, muss jedoch zunächst analysiert werden, inwieweit Schülerangaben zur Wahrnehmung des Unterrichts über Schularten hinweg vergleichbar sind.

Die Ergebnisse belegen, dass Schülerinnen und Schüler in allen Schularten zwischen den Dimensionen der Unterrichtsqualität kognitive Aktivierung, Disziplin im Klassenzimmer und wahrgenommener Unterstützung unterscheiden können, und dass diese Struktur zwischen beiden Gruppen vergleichbar ist. Neuntklässlerinnen und Neuntklässler am Gymnasium berichteten dabei in Einklang mit anderen Studien über eine stärker ausgeprägte Disziplin im Mathematikunterricht, fühlten sich aber gleichzeitig weniger durch ihre Lehrkräfte unterstützt (Gruehn, 2000; Klieme et al., 2001; Kunter et al., 2005). Für kognitive Aktivierung fanden sich im Gegensatz zu bisherigen Forschungsbefunden keine Unterschiede. Innerhalb der Schularten zeigten sich außerdem differentielle Zusammenhangsprofile mit mehrdimensionalen Bildungszielen sowohl auf der Klassen- als auch auf der Individualebene. In beiden Schularten ging eine ausgeprägte Disziplin mit einer höheren Kompetenz einher. Für nicht gymnasiale Schularten ließ sich der Befund jedoch nur auf der Individualebene finden. Das widerspricht der Annahme, dass Klassenführung generell ein Merkmal der geteilten Lernumwelt ist (Aldrup, Klusmann, Lüdtke, Göllner & Trautwein, 2018). Vielmehr scheint aufgrund des heterogeneren Hintergrunds der Schülerinnen und Schüler sowie der allgemein als

geringer wahrgenommenen Disziplin die individuelle Wahrnehmung eines störungsarmen Unterrichts entscheidender für deren Kompetenz zu sein. Zusätzlich ließ sich für die nicht gymnasialen Schularten ein Zusammenhang zwischen Freude und Interesse an Mathematik und kognitiv aktivierendem Unterricht finden. Für das Gymnasium fanden sich differentielle Effekte eines positiven Zusammenhangs zwischen den mathematikbezogenen Selbstwirksamkeitserwartungen und der wahrgenommenen Unterstützung am Gymnasium. In Einklang mit den Ergebnissen von Murdock und Miller (2003) trauen sich Jugendliche, die ihre Lehrkräfte als unterstützend wahrnehmen, im akademischen Bereich mehr zu. Dieser Befund unterstreicht einmal mehr die Notwendigkeit gerade Mathematiklehrkräfte an gymnasien dahingehend aus- und weiterzubilden, dass sie sich in Bezug auf das Erreichen mehrdimensionaler Bildungsziele nicht nur als Wissensvermittler verstehen, sondern gleichzeitig die Lehrer-Schülerbeziehung expliziter in den Blick nehmen.

Aufsatz 11 (Kuger, Klieme, Lüdtke, Schiepe-Tiska & Reiss, 2017) befasst sich mit der Fragestellung inwieweit Analysen, die auf den Daten internationaler Vergleichsstudien basieren und querschnittliche Zusammenhänge zwischen Unterrichtsmerkmalen und mehrdimensionalen Bildungszielen untersuchen, die Bedeutung dieser Merkmale in Bezug auf ihre prognostische Validität überschätzen und die Befunde überinterpretieren. Obwohl auch die OECD mit ihren eigenen Analysen immer wieder den Eindruck erweckt, dass sich aus solchen querschnittlichen Analysen Wirkmechanismen von Unterrichtsmerkmalen auf die Schülerleistung finden lassen würden (z. B. OECD, 2016), können diese Untersuchungen „nur“ eine Zustandsbeschreibung der Unterrichtsrealität abbilden (vgl. Kuger & Klieme, 2016). Für diesen Aufsatz wurden Daten der nationalen Erweiterungsstudie PISA Plus 2012/2013 ausgewertet und mithilfe des Längsschnittdesigns untersucht, inwieweit solche querschnittlichen Ergebnisse die Effektivität der untersuchten Merkmale überschätzen. Dazu wurde für das Fach Mathematik die prognostische Validität von domänenübergreifenden sowie fachspezifischen Merkmalen der Unterrichtsqualität, des Unterrichtsinhaltes und der Unterrichtsaktivitäten auf die Leistung der Schülerinnen und Schüler am Ende der zehnten Klasse untersucht und mit klassischen Querschnittsmodellen verglichen.

Die Befunde zeigten, dass von vier Unterrichtsmerkmalen, die querschnittlich mit der mathematischen Kompetenz zusammenhingen, drei auch eine prognostische Validität für die Schülerleistung am Ende der zehnten Klasse besaßen: Leistungsstarke Jugendliche waren in störungsärmeren Klassen zu finden und berichteten über einen häufigeren Einsatz von Aufgaben, die algebraische Inhalte einüben. In leistungsschwächeren Klassen wurde häufiger über ein individuelles Lehrerfeedback berichtet. Zusammenhänge mit einem stärker schülerorientierten Unterricht ließen sich im Längsschnitt nicht mehr nachweisen. Der Beitrag weist damit auf eine positive Bestätigung der Ergebnisse querschnittlicher Studien zur Unterrichtsforschung hin, wenn gewisse Hintergrundmerkmale der Schülerinnen und Schüler (soziale Herkunft, Zuwanderungshintergrund, Geschlecht) sowie der Lernumwelt (Schulart, soziale Herkunft, Zuwanderungshintergrund) berücksichtigt werden. Damit stützt der Beitrag aus diesen Studien abgeleitete Schlussfolgerungen zur Struktur und Bedeutung von Unterrichtsbeschreibungen grundsätzlich in ihrer Richtung und in ihren Verhältnissen zur Leistung. Er lässt aber auch vermuten, dass selbst eine sorgfältige querschnittliche Modellierung die absolute Höhe der Beziehungen überschätzt. Demnach ist nicht von einer Verfälschung der gefundenen Zusammenhänge mit Unterrichtsmerkmalen durch querschnittliche Analysen internationaler Large-Scale Assessment Daten – zumindest in Bezug auf die Leistung der Schülerinnen und Schüler – auszugehen. In weiterführenden Arbeiten wird geprüft, inwieweit das auch für motivational-affektive Merkmale gilt (Kuger, Schiepe-Tiska & Klieme, 2017).

Zusammenfassend leisten die vorgelegten Arbeiten einen Beitrag dazu, Merkmale der Unterrichtsqualität als wichtige Indikatoren eines Bildungssystems in internationalen Large-Scale Assessments konzeptionell zu präzisieren, auszuwählen und zu operationalisieren. Sie schlagen zudem einen theoretischen Rahmen vor, der die Untersuchung von Zusammenhängen naturwissenschaftlich-mathematischen Unterrichts mit mehrdimensionalen Bildungszielen in Large-Scale Assessments ermöglicht. Anknüpfend an Forschungsbereich I liefern die hier vorgelegten Ergebnisse wichtige empirisch fundierte Erkenntnisse dazu, wie, eingeordnet in den internationalen Vergleich, aktuell der mathematisch-naturwissenschaftliche

Unterricht in Deutschland am Ende der Sekundarstufe I aus Sicht der Jugendlichen aussieht. Darüber hinaus erweitern und vertiefen die Arbeiten den Stand der Unterrichtsforschung in der Untersuchung domänenübergreifender sowie fachspezifischer Qualitätsmerkmale des Unterrichts und berücksichtigen dabei nicht nur kognitive, sondern auch motivational-affektive Lernergebnisse. Sie machen deutlich, dass für die Weiterentwicklung des mathematisch-naturwissenschaftlichen Unterrichts vor allem Potentiale in einer verbesserten wahrgenommenen Unterstützung durch die Lehrkraft, verstärkter Differenzierung und vermehrter individueller Rückmeldung liegen. Diese fachungebundenen Unterrichtsmerkmale sollten stärker in der professionellen Wahrnehmung der MINT-Lehrkräfte verankert werden. Fachspezifische Entwicklungsmöglichkeiten zeigen sich im Mathematikunterricht im Einsatz außermathematischer Anwendungsaufgaben sowie verstärkter kognitiver Aktivierung und in den Naturwissenschaften in der Förderung epistemischer Lernaktivitäten. Sowohl für die Unterrichtsforschung als auch die Praxis scheint das sorgfältige Arrangement domänenübergreifender und fachspezifischer Unterrichtsmerkmale unter der Berücksichtigung des entsprechenden Lernmilieus gerade für das Erreichen mehrdimensionaler Bildungsziele eine wichtige Rolle zu spielen.

In Bezug auf methodenspezifische Forschungsaspekte liefern die Arbeiten sowohl für die Unterrichtsforschung als auch für das Bildungsmonitoring Evidenz, dass auch mit einer als Querschnitt angelegten Large-Scale Assessment Studie unter der Voraussetzung sorgfältiger Modellierungen verlässliche Aussagen über die Qualität von Unterricht und dem Erreichen multipler Lernergebnisse getroffen werden können, die über Schularten hinweg vergleichbar sind.

Ein wichtiges Ziel für weitere Forschung bleibt, die Art und Weise des Zusammenspiels domänenübergreifender und fachspezifischer Unterrichtsmerkmale und deren Wirkung auf mehrdimensionale Bildungsziele genauer zu untersuchen. Dorfner, Förtsch und Neuhaus (2018) haben zum Beispiel zeigen können, dass an Gynasien für die Entwicklung des Interesses im Fach Biologie der Einfluss von Klassenführung und Unterstützung durch kognitive Aktivierung mediiert wurde. Auch die Frage, ob ein häufigeres Vorkommen von Tiefenstrukturen im Unterricht generell mit besseren Leistungen und motivational-affektiven Orientierungen einhergehen

oder es dabei eher auf ein gewisses Maß ankommt, ist unklar. Teig, Scherer und Nilsen (2018) fanden zum Beispiel einen kurvlinearen Zusammenhang zwischen kognitiver Aktivierung und Mathematikleistung. Andere nicht-kognitive Lernergebnisse wurden dabei bisher nicht in den Blick genommen. Nach wie vor bleibt die Frage, wie das Erreichen mehrdimensionaler Bildungsziele in Schule und Unterricht gefördert werden kann, theoretisch spannend und praktisch bedeutsam.

3. Eigenleistungen im Kontext der Fachpublikationen

Die Entstehungszusammenhänge der eingebrachten Fachpublikationen und die jeweiligen Eigenleistungen werden im Folgenden dargelegt. Die Beiträge 2, 3 und 7 teilen den Kontext des nationalen PISA 2015-Berichtsbandes, der von mir mitherausgegeben wurde (Reiss et al., 2016).

Aufsatz 1. Der Beitrag entstand im Rahmen der internationalen Item-Entwicklung für PISA 2015 unter der Leitung des Leibnitz-Institut für Bildungsforschung und Bildungsinformation (DIPF). Er wurde von mir als Erstautorin federführend konzipiert und verfasst. Zudem war ich für die Entwicklung des Indikatorensystems sowie der Erhebungsinstrumente verantwortlich sowie für die Einarbeitung von Anregungen und Kritik seitens der Co-Autoren, der Herausgeber sowie der anonymen Gutachter.

Aufsatz 2. Der Beitrag entstand im Kontext der nationalen PISA 2015 Berichterstattung. Er wurde von mir als Erstautorin initiiert, geplant und zu großen Teilen verfasst. Die Datenauswertung erfolgte in Kooperation zwischen mir und den Co-Autorinnen. Für die Einarbeitung von Anregungen und Kritik seitens der Co-Autorinnen und der Herausgeber war ich verantwortlich.

Aufsatz 3. Der Beitrag entstand im Kontext der nationalen PISA 2012 Berichterstattung und wurde von mir als Erstautorin entworfen und verfasst. In Kooperation mit der Co-Autorin wurden die Datenanalysen durchgeführt und die Ergebnisse von mir interpretiert und diskutiert. Zudem war ich für die Überarbeitung des Beitrags nach der Begutachtung durch die Herausgeber verantwortlich.

Aufsatz 4. Der Beitrag entstand im Rahmen der nationalen PISA 2015 Berichterstattung. Er wurde von mir in Zusammenarbeit mit den Co-Autoren konzipiert und federführend verfasst. Die Datenauswertung wurde von mir durchgeführt und die Ergebnisse gemeinsam mit den Co-Autoren dargestellt und diskutiert. Ich war zudem verantwortlich für die Einarbeitung von Anregungen und Kritik seitens der Co-Autoren und der Herausgeber.

Aufsatz 5. Der Beitrag entstand im Rahmen der Erweiterungsstudie PISA Plus 2012/2013. Er wurde von mir geplant und zu großen Teilen verfasst. Zudem habe ich die Datenanalysen durchgeführt und Anregungen sowie Kritik seitens der Co-Autoren und externen Gutachter eingearbeitet.

Aufsatz 6. Der Beitrag entstand im Rahmen der internationalen Item-Entwicklung für PISA 2015 unter der Leitung des DIPF. Mein Beitrag als Co-Autorin bestand in der substantiellen Mitwirkung der Planung des Beitrags, der Auswahl und Konstruktion von Skalen sowie der Überarbeitung des Textes.

Aufsatz 7. Der Beitrag entstand im Kontext der nationalen PISA 2015 Berichterstattung. Als Erstautorin wurde er von mir in Kooperation mit den Co-Autoren konzipiert und in weiten Teilen verfasst. Die Datenauswertung erfolgte in Kooperation mit den Co-Autoren Heine und Schmidtner. Darüber hinaus war ich für die Einarbeitung der Anregungen und Kritik seitens der Co-Autoren und Herausgeber verantwortlich.

Aufsatz 8. Der Beitrag entstand im Rahmen der nationalen PISA 2012 Berichterstattung. Er wurde von mir gemeinsam mit den Co-Autoren geplant und mit Ausnahme des theoretischen Hintergrunds verfasst. Die Datenauswertung erfolgte in Kooperation mit dem Co-Autor Heine. Anregungen und Kritik seitens der Co-Autoren und Herausgeber wurden hauptverantwortlich von mir eingearbeitet.

Aufsatz 9. Der Beitrag wurde von mir konzipiert und verfasst. Des Weiteren habe ich die Datenanalysen durchgeführt und Anregungen sowie Kritik seitens der Co-Autoren und anonymen Gutachter eingearbeitet.

Aufsatz 10. Der Beitrag wurde von mir geplant, verfasst, eingereicht und überarbeitet.

Aufsatz 11. Der Beitrag entstand im Kontext der Erweiterungsstudie PISA Plus 2012/2013. Als Co-Autorin habe ich intensiv an der Konzeption, Aus- und Überarbeitung des Beitrags sowie an der Überarbeitung nach der Rückmeldung durch die externen Gutachter mitgewirkt.

4. Literaturverzeichnis

- Aldrup, K., Klusmann, U., Lüdtke, O., Göllner, R. & Trautwein, U. (2018). Social support and classroom management are related to secondary students' general school adjustment. A multilevel structural equation model using student and teacher ratings. *Journal of Educational Psychology*. <https://doi.org/10.1037/edu0000256>
- Alfieri, L., Brooks, P. J., Aldrich, N. J. & Tenenbaum, H. R. (2011). Does discovery-based instruction enhance learning? *Journal of Educational Psychology*, 103, 1–18. <https://doi.org/10.1037/a0021017>
- Almlund, M., Duckworth, A. L., Heckman, J. & Kautz, T. (2011). *Personality psychology and economics*. Cambridge, MA: National Bureau of Economic Research. <https://doi.org/10.3386/w16822>
- Baumert, J. (1997). Kommentar - Zielkonflikte in der Grundschule. In F. E. Weinert & A. Helmke (Hrsg.), *Entwicklung im Grundschulalter* (S. 317–321). Weinheim: Beltz.
- Baumert, J., Stanat, P. & Demmrich, A. (2001). PISA 2000: Untersuchungsgegenstand, theoretische Grundlagen und Durchführung der Studie. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider et al. (Hrsg.), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 15–68). Opladen: Leske + Budrich.
- Baumert, J., Stanat, P. & Watermann, R. (2006). Schulstruktur und die Entstehung differenzieller Lern- und Entwicklungsmilieus. In J. Baumert (Hrsg.), *Herkunftsbedingte Disparitäten im Bildungswesen. Differenzielle Bildungsprozesse und Probleme der Verteilungsgerechtigkeit: Vertiefende Analysen im Rahmen von PISA 2000* (1. Aufl., S. 95–188). Wiesbaden: Verlag für Sozialwissenschaften.
- Baumert, J., Lehmann, R.H., Lehrke, M., Schmitz, B., Clausen, M., Hosenfeld, I. et al. (1997). *TIMSS. Mathematisch-naturwissenschaftlicher Unterricht im internationalen Vergleich. Deskriptive Befunde*. Opladen: Leske & Budrich.
- Berliner, D. C. (2005). The Near Impossibility of Testing for Teacher Quality. *Journal of Teacher Education*, 56, 205–213. <https://doi.org/10.1177/0022487105275904>
- Betts, J. R. (2011). The economics of tracking in education. In E. A. Hanushek, S. Machin & L. Woessmann (Eds.), *Handbook of the economics of education* (pp. 341–381). Amsterdam: North-Holland.

- Blossfeld, H.-P., Bos, W., Daniel, H.-D., Hannover, B., Köller, O., Lenzen, D. et al. (2015). *Bildung. Mehr als Fachlichkeit*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Börlin, J. (2012). *Das Experiment als Lerngelegenheit. Vom interkulturellen Vergleich des Physikunterrichts zu Merkmalen seiner Qualität*. Berlin: Logos-Verlag.
- Bruner, J. S. (1961). The act of discovery. *Harvard Educational Review*, 31, 21–32.
- Danner, D., Lechner, C. M. & Spengler, M. (Hrsg.). (2019). *Do we need socio-emotional skills?* (Research Topic). *Frontiers in Psychology*. Zugriff am 04.06.2019.
Verfügbar unter <https://www.frontiersin.org/research-topics/8713/do-we-need-socio-emotional-skills>
- Doll, J. & Prenzel, M. (2003). Einleitung in das Beiheft. In J. Doll & M. Prenzel (Hrsg.), *Bildungsqualität von Schule. Schulische und außerschulische Bedingungen mathematischer, naturwissenschaftlicher und einfacher Kompetenzen* (S. 9–30). Weinheim: Beltz.
- Dorfner, T., Förtsch, C. & Neuhaus, B. J. (2018). Effects of three basic dimensions of instructional quality on students' situational interest in sixth-grade biology instruction. *Learning and Instruction*, 56, 42–53.
<https://doi.org/10.1016/j.learninstruc.2018.03.001>
- Doroganova, A. (2019, Februar). *How to motivate students for a career in science? The role of teaching and learning conditions in heterogeneous classrooms*, Vortrag auf der Nachwuchstagung der 7. Tagung der Gesellschaft für Empirische Bildungsforschung, Köln, Deutschland.
- Dumont, H., Neumann, M., Maaz, K. & Trautwein, U. (2013). Die Zusammensetzung der Schülerschaft als Einflussfaktor für Schulleistungen. Internationale und nationale Befunde. *Psychologie in Erziehung und Unterricht*, 60, 163–183.
<https://doi.org/10.2378/peu2013.art14d>
- Fauth, B., Decristan, J., Rieser, S., Klieme, E. & Büttner, G. (2014). Student ratings of teaching quality in primary school: Dimensions and prediction of student outcomes. *Learning and Instruction*, 29, 1–9.
<https://doi.org/10.1016/j.learninstruc.2013.07.001>
- Fend, H. (2008). *Neue Theorie der Schule. Einführung in das Verstehen von Bildungssystemen* (2. Aufl.). Wiesbaden: Verlag für Sozialwissenschaften.

- Forbes, C. T., Neumann, K. & Schiepe-Tiska, A. (eingereicht). Science teaching and learning: Analysis of PISA 2015 data from the United States and Germany. *International Journal of Science Education*.
- Furtak, E. M., Seidel, T., Iverson, H. & Briggs, D. C. (2012). Experimental and Quasi-Experimental Studies of Inquiry-Based Science Teaching: A Meta-Analysis. *Review of Educational Research*, 82, 300–329.
<https://doi.org/10.3102/0034654312457206>
- Geiser, C. (2010). *Datenanalyse mit Mplus. Eine anwendungsorientierte Einführung*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Gruehn, S. (2000). *Unterricht und schulisches Lernen. Schüler als Quellen der Unterrichtsbeschreibung* (Pädagogische Psychologie und Entwicklungspsychologie, Bd. 12). Zugl.: Berlin, Freie Univ., Diss., 1998. Münster: Waxmann.
- Hattie, J. (2009). *Visible Learning. A synthesis of over 800 meta-analyses relating to achievement*. New York: Routledge.
- Heine, J.-H., Mang, J., Borchert, L., Gomolka, J., Kröhne, U., Goldhammer, F. & Sälzer, C. (2016). Kompetenzmessung in PISA 2015. In K. Reiss, C. Sälzer, A. Schiepe-Tiska, E. Klieme & O. Köller (Hrsg.), *PISA 2015. Eine Studie zwischen Kontinuität und Innovation* (S. 383–430). Münster: Waxmann.
- Hörner, W. (1997). "Europa" als Herausforderung für die Vergleichende Erziehungswissenschaft - Reflexionen über die politische Funktion einer pädagogischen Disziplin. In C. Kodron, B. v. Kopp, U. Lauterbach, U. Schäfer & G. Schmidt (Hrsg.), *Vergleichende Erziehungswissenschaft. Herausforderung - Vermittlung - Praxis; Festschrift für Wolfgang Mitter zum 70. Geburtstag* (S. 65–80). Köln: Böhlau.
- Humboldt, W. v. (1903). *Ideen zu einem Versuch, die Grenzen der Wirksamkeit des Staats zu bestimmen, gesammelte Schriften* (Nr. 1). Berlin/Leipzig.
- Ivanov, S. (2011). Naturwissenschaftliche Kompetenz und fachbezogene Einstellungen. In U. Vieluf, S. Ivanov & R. Nikolova (Hrsg.), *KESS 10/11. Kompetenzen und Einstellungen von Schülerinnen und Schülern an Hamburger Schulen am Ende der Sekundarstufe I und zu Beginn der gymnasialen Oberstufe* (HANSE - Hamburger Schriften zur Qualität im Bildungswesen, Bd. 10, S. 183–214). Münster: Waxmann.

- Ivanov, S. & Nikolova, R. (2010). Naturwissenschaftliche Kompetenz. In W. Bos & C. Gröhlich (Hrsg.), *KESS 8. Kompetenzen und Einstellungen von Schülerinnen und Schülern am Ende der Jahrgangsstufe 8* (HANSE - Hamburger Schriften zur Qualität im Bildungswesen, Bd. 6, S. 79–118). Münster: Waxmann.
- Jansen, M., Scherer, R. & Schroeders, U. (2015). Students' self-concept and self-efficacy in the sciences. Differential relations to antecedents and educational outcomes. *Contemporary Educational Psychology*, 41, 13–24.
<https://doi.org/10.1016/j.cedpsych.2014.11.002>
- Jude, N. (2016). The assessment of learning contexts in PISA. In S. Kuger, E. Klieme, N. Jude & D. Kaplan (Hrsg.), *Assessing contexts of learning. An international perspective* (S. 39–52). Berlin: Springer.
- Kirschner, P. A., Sweller, J. & Clark, R. E. (2006). Why Minimal Guidance During Instruction Does Not Work. An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching. *Educational Psychologist*, 41, 75–86. https://doi.org/10.1207/s15326985ep4102_1
- Kjærnsli, M. & Lie, S. (2011). Students' Preference for Science Careers: International comparisons based on PISA 2006. *International Journal of Science Education*, 33, 121–144. <https://doi.org/10.1080/09500693.2010.518642>
- 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. Forschungsbefunde, Reforminitiativen, Praxisberichte und Video-Dokumente* (S. 43–57). Berlin: BMBF.
- KMK (2003). *Beschlüsse der Kultusministerkonferenz - Bildungsstandards im Fach Mathematik für den Mittleren Schulabschluss (Beschluss vom 04. Dezember 2003)*.
- KMK (2005a). *Beschlüsse der Kultusministerkonferenz - Bildungsstandards im Fach Biologie für den Mittleren Schulabschluss (Beschluss vom 16. Dezember 2004)*. München: Wolters Kluwer.
- KMK (2005b). *Beschlüsse der Kultusministerkonferenz - Bildungsstandards im Fach Chemie für den Mittleren Schulabschluss (Beschluss vom 16. Dezember 2004)*. München: Wolters Kluwer.

- KMK (2005c). *Beschlüsse der Kultusministerkonferenz - Bildungsstandards im Fach Physik für den Mittleren Schulabschluss (Beschluss vom 16. Dezember 2004)*. München: Wolters Kluwer.
- Kobarg, M., Prenzel, M., Seidel, T., Walker, M., McCrae, B., Cresswell et al. (2011). *An international comparison of science teaching and learning. Further results from PISA 2006*. Münster: Waxmann.
- Köller, O. & Baumert, J. (2002). Entwicklungen schulischer Leistungen. In R. Oerter & L. Montada (Hrsg.), *Entwicklungspsychologie* (5. Aufl., S. 756–786). Weinheim: Beltz.
- Kuger, S. & Klieme, E. (2016). Dimensions of context assessment. In S. Kuger, E. Klieme, N. Jude & D. Kaplan (Hrsg.), *Assessing contexts of learning. An international perspective* (S. 3–38). Berlin: Springer.
- Kuger, S., Klieme, E., Jude, N. & Kaplan, D. (Hrsg.). (2016). *Assessing contexts of learning. An international perspective*. Berlin: Springer.
- Kuger, S., Schiepe-Tiska, A. & Klieme, E. (2017, März). *Leistungs- und Interessensentwicklung im Mathematikunterricht der Sekundarstufe*. Vortrag auf der 5. Konferenz der Gesellschaft für Empirische Bildungsforschung (GEBF), Heidelberg, Deutschland.
- Kuger, S., Klieme, E., Lüdtke, O., Schiepe-Tiska, A. & Reiss, K. (2017). Mathematikunterricht und Schülerleistung in der Sekundarstufe. Zur Validität von Schülerbefragungen in Schulleistungsstudien. *Zeitschrift für Erziehungswissenschaft*, 20 (S2), 61–98. <https://doi.org/10.1007/s11618-017-0750-6>
- Kunter, M. & Ewald, S. (2016). Bedingungen und Effekte von Unterricht: Aktuelle Forschungsperspektiven aus der pädagogischen Psychologie. In N. McElvany, W. Bos & H. G. Holtappels (Hrsg.), *Bedingungen und Effekte guten Unterrichts* (Dortmunder Symposium der Empirischen Bildungsforschung, S. 9–32). Münster: Waxmann.
- Kunter, M. & Voss, T. (2011). Das Modell der Unterrichtsqualität in COACTIV: Eine multikriteriale Analyse. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 115–132). Münster: Waxmann.

- Kunter, M. (2005). *Multiple Ziele im Mathematikunterricht* (Pädagogische Psychologie und Entwicklungspsychologie, Bd. 51). 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., Brunner, M., Baumert, J., Klusmann, U., Krauss, S., Blum, W. et al. (2005). Der Mathematikunterricht der PISA-Schülerinnen und -Schüler. *Zeitschrift für Erziehungswissenschaft*, 8, 502–520. <https://doi.org/10.1007/s11618-005-0156-8>
- Lavonen, J. & Laaksonen, S. (2009). Context of teaching and learning school science in Finland. Reflections on PISA 2006 results. *Journal of Research in Science Teaching*, 46, 922–944. <https://doi.org/10.1002/tea.20339>
- Lazonder, A. W. & Harmsen, R. (2016). Meta-Analysis of Inquiry-Based Learning. *Review of Educational Research*, 86, 681–718. <https://doi.org/10.3102/0034654315627366>
- Marksteiner, T., Kuger, S. & Klieme, E. (2018). The potential of anchoring vignettes to increase intercultural comparability of non-cognitive factors. *Assessment in Education: Principles, Policy & Practice*, 8, 1–21. <https://doi.org/10.1080/0969594X.2018.1514367>
- Möller, K. (2016). Bedingungen und Effekte qualitätvollen Unterrichts - ein Beitrag aus fachdidaktischer Perspektive. In N. McElvany, W. Bos & H. G. Holtappels (Hrsg.), *Bedingungen und Effekte guten Unterrichts* (Dortmunder Symposium der Empirischen Bildungsforschung, S. 43–64). Münster: Waxmann.
- Müller, F., Schiepe-Tiska, A. & Strohmaier, A. R. (2017). *EyeQuest – Cross-cultural comparison of eye movements and self-reports*. Vortrag auf dem jährlichen Treffen des National Council on Measurement in Education (NCME), San Antonio, TX, USA. Zugriff am 04.06.2019. Verfügbar unter http://www.ma.edu.tum.de/fileadmin/tueds11/www/poster/NCME_2017.pdf
- Müller, K., Prenzel, M., Seidel, T., Schiepe-Tiska, A. & Kjærnsli, M. (2016). Science teaching and learning in schools: Theoretical and empirical foundations for investigating classroom-level processes. In S. Kuger, E. Klieme, N. Jude & D. Kaplan (Hrsg.), *Assessing contexts of learning. An international perspective* (S. 423–446). Berlin: Springer.

- Murdock, T. B. & Miller, A. (2003). Teachers as sources of middle school students' motivational identity. Variable-centered and person-centered analytic approaches. *The Elementary School Journal*, 103, 383–399.
<https://doi.org/10.1086/499732>
- Murphy, N. A. & Hall, J. A. (2011). Intelligence and interpersonal sensitivity. A meta-analysis. *Intelligence*, 39 (1), 54–63. <https://doi.org/10.1016/j.intell.2010.10.001>
- National Center for Education Statistics. (1995). *Two years later: Cognitive gains and school transitions of NELS:88 eighth graders*. Washington, DC: U.S. Department of Education.
- OECD (2013a). *PISA 2012 assessment and analytical framework: Mathematics, reading, science, problem solving and financial literacy*. Paris: OECD Publishing.
- OECD (2013b). *PISA 2012 Results: Ready to Learn: Students' Engagement, Drive and Self-Beliefs (Volume III)*. Paris: OECD Publishing.
- OECD (2014). *PISA 2012 Technical Report*. Paris: OECD Publishing.
- OECD (2016). *PISA 2015 Results (Volume I): Excellence and Equity in Education*. Paris: OECD Publishing.
- OECD (2017). *PISA 2015 assessment and analytical framework: Mathematics, reading, science, problem solving and financial literacy*. Paris: OECD Publishing.
- OECD (2018). *The future of education and skills. Education 2030*. Accessed 04.06.2019. Retrieved from [https://www.oecd.org/education/2030/E2030%20Position%20Paper%20\(05.04.2018\).pdf](https://www.oecd.org/education/2030/E2030%20Position%20Paper%20(05.04.2018).pdf)
- Oser, F. & Baeriswyl, F. J. (2001). Choreographies of teaching: Bridging instruction to learning. In V. Richardson (Ed.), *Handbook of research on teaching* (4th ed., pp. 1031–1065). Washington, D.C.: American Educational Research Association.
- Pant, H. A., Stanat, P., Schroeders, U., Roppelt, A., Siegle, T. & Pöhlmann, C. (Hrsg.). (2013). *IQB-Ländervergleich 2012. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I*. Münster: Waxmann.
- Pinquart, M. & Sörensen, S. (2000). Influences of socioeconomic status, social network, and competence on subjective well-being in later life. A meta-analysis. *Psychology and Aging*, 15, 187–224. <https://doi.org/10.1037/0882-7974.15.2.187>

- Praetorius, A.-K., Klieme, E., Herbert, B. & Pinger, P. (2018). Generic dimensions of teaching quality. The German framework of Three Basic Dimensions. *ZDM Mathematics Education*, 50, 407–426. <https://doi.org/10.1007/s11858-018-0918-4>
- Prenzel, M. (2012). Empirische Bildungsforschung morgen: Reichen unsere bisherigen Forschungsansätze aus? In M. Gläser-Zikuda, T. Seidel, C. Rohlfs, A. Gröschner & S. Ziegelbauer (Hrsg.), *Mixed methods in der empirischen Bildungsforschung* (S. 273–286). Münster: Waxmann.
- Prenzel, M., Carstensen, C., Frey, A., Drechsel, B. & Rönnebeck, S. (2007). PISA 2006 - Eine Einführung in die Studie. In M. Prenzel, C. Artelt, J. Baumert, W. Blum, M. Hammann, E. Klieme et al. (Hrsg.), *PISA 2006. Die Ergebnisse der dritten internationalen Vergleichsstudie* (S. 31–60). Münster: Waxmann.
- Purves, A. C. (1987). The Evolution of the IEA: A Memoir. *Comparative Education Review*, 31 (1), 10–28. <https://doi.org/10.1086/446653>
- Rakoczy, K. (2008). *Motivationsunterstützung im Mathematikunterricht: Unterricht aus der Perspektive von Lernenden und Beobachtern* (Bd. 65). Münster: Waxmann.
- Reiss, K., Klieme, E., Köller, O. & Stanat, P. (Hrsg.). (2017). *PISA Plus 2012 – 2013. Kompetenzentwicklung im Verlauf eines Schuljahres* (Zeitschrift für Erziehungswissenschaft Sonderheft). Wiesbaden: Springer VS.
- Reiss, K., Sälzer, C., Schiepe-Tiska, A., Klieme, E. & Köller, O. (Hrsg.). (2016). *PISA 2015. Eine Studie zwischen Kontinuität und Innovation*. Münster: Waxmann.
- Robbins, S. B., Lauver, K., Le, H., Davis, D., Langley, R. & Carlstrom, A. (2004). Do psychosocial and study skill factors predict college outcomes? A meta-analysis. *Psychological Bulletin*, 130, 261–288. <https://doi.org/10.1037/0033-2909.130.2.261>
- Robitzsch, A., Lüdtke, O., Köller, O., Kröhne, U., Goldhammer, F. & Heine, J.-H. (2017). Herausforderungen bei der Schätzung von Trends in Schulleistungsstudien. *Diagnostica*, 63, 148–165. <https://doi.org/10.1026/0012-1924/a000177>
- Schiepe-Tiska, A., Heine, J.-H., Lüdtke, O., Seidel, T. & Prenzel, M. (2016). Mehrdimensionale Bildungsziele im Mathematikunterricht und ihr Zusammenhang mit den Basisdimensionen der Unterrichtsqualität. *Unterrichtswissenschaft*, 44, 211–225. <https://doi.org/10.3262/UW1603211>

- Schiepe-Tiska, A., Reiss, K., Obersteiner, A., Heine, J.-H., Seidel, T. & Prenzel, M. (2013). Mathematikunterricht in Deutschland: Befunde aus PISA 2012 [PISA 2012. Advances and challenges in Germany]. In M. Prenzel, C. Sälzer, E. Klieme & O. Köller (Hrsg.), *PISA 2012. Fortschritte und Herausforderungen in Deutschland* (S. 123–154). Münster: Waxmann.
- Schiepe-Tiska, A., Roczen, N., Müller, K., Prenzel, M. & Osborne, J. (2016). Science-related outcomes: Attitudes, motivation, value beliefs, strategies. In S. Kuger, E. Klieme, N. Jude & D. Kaplan (Hrsg.), *Assessing contexts of learning. An international perspective*. Berlin: Springer.
- Schiepe-Tiska, A. & Schmidtner, S. (2013). Mathematikbezogene emotionale und motivationale Orientierungen, Einstellungen und Verhaltensweisen von Jugendlichen in PISA 2012. In M. Prenzel, C. Sälzer, E. Klieme & O. Köller (Hrsg.), *PISA 2012. Fortschritte und Herausforderungen in Deutschland* (S. 99–122). Münster: Waxmann.
- Schiepe-Tiska, A., Schmidtner, S., Müller, K., Heine, J.-H., Neumann, K. & Lüdtke, O. (2016). Naturwissenschaftlicher Unterricht in Deutschland in PISA 2015 im internationalen Vergleich. In K. Reiss, C. Sälzer, A. Schiepe-Tiska, E. Klieme & O. Köller (Hrsg.), *PISA 2015. Eine Studie zwischen Kontinuität und Innovation* (S. 133–176). Münster: Waxmann.
- Schiepe-Tiska, A., Simm, I. & Schmidtner, S. (2016). Motivationale Orientierungen, Selbstbilder und Berufserwartungen in den Naturwissenschaften in PISA 2015. In K. Reiss, C. Sälzer, A. Schiepe-Tiska, E. Klieme & O. Köller (Hrsg.), *PISA 2015. Eine Studie zwischen Kontinuität und Innovation* (S. 99–132). Münster: Waxmann.
- Schiepe-Tiska, A. (2019). School Tracks as Differential Learning Environments Moderate the Relationship between Teaching Quality and Multidimensional Learning Goals in Mathematics. *Frontiers in Education*, 4, 4.
<https://doi.org/10.3389/feduc.2019.00004>

- Schiepe-Tiska, A., Rönnebeck, S., Schöps, K., Neumann, K., Schmidtner, S., Parchmann, I. et al. (2016). Naturwissenschaftliche Kompetenz in PISA 2015 – Ergebnisse des internationalen Vergleichs mit einem modifizierten Testansatz. In K. Reiss, C. Sälzer, A. Schiepe-Tiska, E. Klieme & O. Köller (Hrsg.), *PISA 2015. Eine Studie zwischen Kontinuität und Innovation* (S. 45–98). Münster: Waxmann.
- Schiepe-Tiska, A., Rönnebeck, S., Heitmann, P., Schöps, K., Prenzel, M. & Nagy, G. (2017). Die Veränderung der naturwissenschaftlichen Kompetenz von der 9. zur 10. Klasse bei PISA und den Bildungsstandards unter Berücksichtigung geschlechts- und schulartspezifischer Unterschiede sowie der Zusammensetzung der Schülerschaft. *Zeitschrift für Erziehungswissenschaft*, 20, 151–176.
<https://doi.org/10.1007/s11618-017-0754-2>
- Schroeders, U., Penk, C., Jansen, M. & Pant, H. A. (2013). Geschlechtsbezogene Disparitäten. In H. A. Pant, P. Stanat, U. Schroeders, A. Roppelt, T. Siegle & C. Pöhlmann (Hrsg.), *IQB-Ländervergleich 2012. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I* (S. 249–274). Münster: Waxmann.
- Seel, N. M. & Hanke, U. (Hrsg.). (2015). *Erziehungswissenschaft. Lehrbuch für Bachelor-, Master- und Lehramtsstudierende*. Berlin, Heidelberg: Springer.
- Seidel, T. & Prenzel, M. (2008). Assessment in large-scale studies. In J. Hartig, E. Klieme & D. Leutner (Eds.), *Assessment of Competencies in Educational Contexts* (pp. 279–304). Göttingen: Hogrefe.
- Seidel, T. & Reiss, K. (2014). Lerngelegenheiten im Unterricht. In T. Seidel & A. Krapp (Hrsg.), *Pädagogische Psychologie* (6., vollständig überarbeitete Aufl., S. 253–276). Weinheim: Beltz.
- Teig, N., Scherer, R. & Nilsen, T. (2018). More isn't always better. The curvilinear relationship between inquiry-based teaching and student achievement in science. *Learning and Instruction*, 56, 20–29.
<https://doi.org/10.1016/j.learninstruc.2018.02.006>
- Treiber, B. & Weinert, F. E. (1982). Gibt es theoretische Fortschritte in der Lehr-Lern-Forschung? In B. Treiber & F. E. Weinert (Hrsg.), *Lehr-Lern-Forschung* (S. 242–290). München: Urban & Schwarzenberg.

- Tupac-Yupanqui, A., Schiepe-Tiska, A., Weis, M., Heine, J.-H. & Reiss, K. (2019, Februar). *Der Einfluss von unterschiedlichen Antwortformaten auf Zusammenhänge zwischen mathematikbezogenen Lernstrategien und Mathematikkompetenz in PISA 2012*, Poster auf der Nachwuchstagung der 7. Tagung der Gesellschaft für Empirische Bildungsforschung, Köln, Deutschland.
- Van de Vijver, F. J. R. & He, J. (2016). Bias assessment and prevention in noncognitive outcome measures in context assessments. In S. Kuger, E. Klieme, N. Jude & D. Kaplan (Hrsg.), *Assessing contexts of learning. An international perspective* (S. 229–254). Berlin: Springer.
- Walter, O., Senkbeil, M., Rost, J., Carstensen, C. & Prenzel, M. (2006). Die Entwicklung der naturwissenschaftlichen Kompetenz von der neunten zur zehnten Klassenstufe: Deskriptive Befunde. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand et al. (Hrsg.), *PISA 2003. Untersuchungen zur Kompetenzentwicklung im Verlauf eines Schuljahres* (S. 88–118). Münster: Waxmann.
- Weinert, F. E. (2002). Vergleichende Leistungsmessung in Schulen – eine umstrittene Selbstverständlichkeit. In F. E. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (2., unveränd. Aufl., S. 17–31). Weinheim: Beltz.
- Wenger, M., Lüdtke, O. & Brunner, M. (2018). Übereinstimmung, Variabilität und Reliabilität von Schülerurteilen zur Unterrichtsqualität auf Schulebene. *Zeitschrift für Erziehungswissenschaft*, 83, 261. <https://doi.org/10.1007/s11618-018-0813-3>