

Prozessbezogene Kompetenzen in zentralen Abschlussarbeiten im Fach Mathematik? – Eine Analyse aller ZP10-Aufgaben Nordrhein-Westfalens von 2007 bis 2019

Bruno Scheja · Anna Schreck · Benjamin Rott 

Eingegangen: 6. April 2022 / Überarbeitet: 21. Mai 2023 / Angenommen: 4. Juli 2023 / Online publiziert: 22. September 2023
© The Author(s) 2023

Zusammenfassung Die Aktivitäten „Problemlösen“, „Modellieren“ und „Argumentieren“ stellen breit anerkannte im Mathematikunterricht zu vermittelnde Kompetenzen dar. Welche Rolle sie im Mathematikunterricht spielen, hängt aber auch davon ab, ob und in welchem Maße sie in zentralen Abschlussprüfungen abgeprüft werden. Um den Stellenwert der drei Kompetenzen in der Zentralen Abschlussprüfung nach Jahrgang 10 (ZP10) des Bundeslands Nordrhein-Westfalen zu erfassen, wurde eine Dokumentenanalyse aller Aufgaben der ZP10 des Bundeslands Nordrhein-Westfalen seit ihrer Einführung 2007 bis einschließlich 2019 durchgeführt: Insgesamt wurden 1886 Aufgaben aus 68 Prüfungsdokumenten analysiert, wobei jede Kompetenz auf einer vierstufigen Skala kodiert wurde. Die Ergebnisse zeigen, dass kaum mehr als ein minimaler Anspruch in Bezug auf die genannten Kompetenzen gefordert wird, der zudem über den gesamten Zeitraum nur geringe Veränderungen erfährt.

Schlüsselwörter Aufgabenanalyse · Mathematische Kompetenzen · Zentrale Prüfung · Kognitive Prozesse

✉ Dr. Dr. Bruno Scheja · Anna Schreck · Prof. Dr. Benjamin Rott
Institut für Mathematikdidaktik, Universität zu Köln, Gronewaldstr. 2, 50931 Köln, Deutschland
E-Mail: schejab@uni-koeln.de

Anna Schreck
E-Mail: aschrec1@uni-koeln.de

Prof. Dr. Benjamin Rott
E-Mail: brott@uni-koeln.de

Process-related competencies of central examinations in Mathematics—An analysis of all ZP10 tasks in North Rhine-Westphalia from 2007 to 2019

Abstract The activities problem solving, modelling, and reasoning are identified as central educational goals. However, the role they play in class teaching also depends on whether they are tested in central examinations. To assess the significance of these three competencies for mathematics classes, an analysis of all tasks (document analysis) of the central final examination after year 10 (ZP10) of the state of North Rhine-Westphalia since its introduction in 2007 up to 2019 was conducted: A total of 1886 tasks from 68 examination documents were analyzed, with each competency being coded on a four-level scale. The results show that hardly more than a minimum requirement is demanded in terms of the competences mentioned.

Keywords Task analysis · Mathematical skills · Central examination · Cognitive processes

1 Einleitung

Nicht zuletzt die Veröffentlichung der PISA-Ergebnisse 2001 zeigte, dass der deutsche Unterricht Defizite hinsichtlich der Verwendung und der Vernetzung von vorstellungs- und kompetenzorientierten Aufgaben aufweist (Baumert et al. 1997; Neubrand 2002). Die in der Folgezeit eingeleiteten Reformen waren vielschichtig. Ein zentrales Element dieser war die Übernahme des kompetenzbasierten Bildungsverständnisses in der mathematischen (Schul-)Bildung (vgl. Neubrand 2015; Turner et al. 2015). Eine Orientierung des mathematischen Unterrichts an Kompetenzen bedeutete insbesondere eine Umorientierung weg vom mathematischen Gegenstand hin zur Ausföhrung mathematischer Aktivitäten und Prozesse (Reiss et al. 2008). Innerhalb der Mathematikdidaktik haben vorrangig das KOM- (Niss 2003) und das RAND-Projekt (RAND Mathematics Study Panel 2003) diesen kompetenzorientierten Bildungsansatz dargelegt und ausgebaut. Niss und Højgaard versuchen mit ihrem Kompetenzmodell eine in der Folgezeit mehrfach angepasste und breit konsentzierte Antwort darauf zu geben, welche Teilkompetenzen mathematische Bildung umfasst (vgl. Niss 2003; Niss und Højgaard 2011, 2019). Dies passt zur Neuausrichtung der Unterrichtstheorie und Schulpraxis, die sich nicht länger an der Vermittlung von „trägem Wissen“, das in außerschulischen Kontexten nur selten angewendet werden kann, ausrichtet, sondern am verständnisvollen Umgang mit Mathematik (Klieme et al. 2003). In Deutschland manifestierte sich diese Fokusänderung z. B. in der Einführung der Bildungsstandards und der ländereigenen Kernlehrpläne (KLPe). Sie können als *Zuschnitte des Kompetenzmodells* von Niss betrachtet werden, die jeweils ein Konzept von mathematischer Bildung operationalisieren. In Nordrhein-Westfalen sind diese Kompetenzmodelle zum Teil schulformspezifisch und damit auch innerhalb des Bundeslands keineswegs einheitlich (Blum et al. 2005, 2019; Neubrand und Neubrand 2010). Unabhängig der Ebene (bundes-, landes- oder schulformebe- ne) finden die mathematischen Kompetenzen „Problemlösen“, „Argumentieren“ und

„Modellieren“ (dort als *prozessbezogene Kompetenzen* bezeichnet) stets Berücksichtigung. Diese drei Kompetenzfacetten bilden zentrale Prozesse und Aktivitäten ab und sind somit zugleich sowohl international als auch national breit konsentiert (vgl. KMK 2003b, 2004; MSW 2011; Stacey und Turner 2015).

Infolge der Kompetenzorientierung und des begleitenden Steuerungswandels des Bildungssystems sollten, so die bildungspolitische Annahme, diese curricularen Vorgaben zugleich zur inhaltlich-methodischen Grundlage für ein dichtes Netz vergleichender Leistungskontrollen innerhalb der Bundesländer werden. Hierzu zählen bspw. Vergleichsarbeiten (VERA 8), zentrale Prüfungen am Ende der Sekundarstufe I (ZP10) und das Zentralabitur. In den eingeführten zentralen (Abschluss-)Prüfungen weisen das Zentralabitur sowie die ZP10 eine „Individualbewertung“ mit Selektions- und Zertifizierungsfunktion aus. Die Befunde zur Wirkung derartiger Prüfungen deuten dabei auch auf eine Reihe unerwünschter Effekte hin. Hierzu zählt unter anderem, dass ein wesentlicher Teil der Unterrichtszeit für das Einüben prüfungsrelevanter Inhalte und Testformate verwendet wird und die zunehmende Verengung des Curriculums auf Prüfungsinhalte, welche sich zudem auf mehrere Schuljahre erstrecken kann (Konarzewski 2008; Kühn 2010; Madaus et al. 2009). Die beispielhaft genannten Effekte zeigen, dass sich die zentral gestellten Prüfungen im Hinblick auf Steuerung vermutlich nicht nur auf Lernende, sondern auch auf das gesamte Lernarrangement (Eltern, Lehrkräfte, Schule) auswirken.

Dieser Artikel soll einen Beitrag zur allgemein- wie auch zur fachdidaktischen Forschung leisten, indem empirisch überprüft wird, welche Rolle die prozessbezogenen Kompetenzen „Problemlösen“, „Modellieren“ und „Argumentieren“ in den Aufgaben der NRW ZP10 des Zeitraums 2007 bis 2019 spielen. Hierzu wird das jeweils eingeforderte Kompetenzniveau der Aufgaben empirisch erfasst und unter verschiedenen Gesichtspunkten (u. a. in Abhängigkeit von der Schulform, dem Zeitverlauf) vor dem Hintergrund eines fachdidaktisch fundierten, gemeinsamen Referenzrahmens – drei eigens erstellten Kategoriensystemen zu drei Kompetenzen – bewertet. Eine solche Aufgabenanalyse liefert unter normativer Perspektive Erkenntnisse auf die Frage „Was bedeutet es, in Mathematik kompetent zu sein?“ in Bezug auf konstituierende Denkhaltungen und Tätigkeiten.

Die Anzahl groß angelegter empirischer Studien, die das Aufgabenpotential von Testaufgaben auf der Grundlage von Aufgabenmerkmalen untersuchen, ist überschaubar (Drüke-Noe 2014; Jordan et al. 2006; Scheja 2019). In den bisherigen Studien geht es zumeist um das Konzept der kognitiven Aktivierung, für deren Untersuchung größtenteils die Kategorien der COACTIV-Studie (Jordan et al. 2006) verwendet werden oder um den Implementationsstand der Bildungsstandards bzw. der KLPe (Kühn und Drüke-Noe 2013; Neubrand und Neubrand 2010). Indem der vorliegende Beitrag davon abweichend die ZP10-Aufgaben hinsichtlich der drei prozessbezogenen *Kompetenzen* „Problemlösen“, „Modellieren“ und „Argumentieren“ und ihren Anforderungsniveaus analysiert, wird ein Forschungsdesiderat aufgegriffen. Eine Betrachtung der Realisierung der übrigen drei Kompetenzen der Bildungsstandards („Darstellungen verwenden“, „mit Mathematik symbolisch, formal und technisch umgehen“, sowie „Kommunizieren“) wird hier aus methodischen Gründen ausgeklammert; so finden die Kompetenzen „Darstellungen verwenden“ und „mit Mathematik symbolisch, formal und technisch umgehen“ im nordrhein-westfälischen

lischen KLP, der die Grundlage für die Konzeption der Testaufgaben darstellt, keine Berücksichtigung (s. MSW 2004a).¹ Weitere Gründe sind a), dass die ZP10 stets in zwei Teile unterteilt wird (mit und ohne Hilfsmittel), womit die Anforderungen an die technische Performanz nicht immer von Lernenden gezeigt werden sollen und dass die Breite technischer Hilfsmittel, wenn überhaupt, dann auf grafikfähige Taschenrechner beschränkt ist und Hilfsmittel wie beispielsweise GeoGebra, CAS-Systeme etc. gar nicht zulässig sind sowie b) dass Kommunizieren in erster Linie auf das sprachbasierte „schriftliche oder mündliche Darstellen und Präsentieren von Überlegungen“ (Blum et al. 2010, S. 48) in der ZP10 beschränkt ist und somit Kommunizieren weder in seiner Vielseitigkeit abgebildet wird, noch die geeignete Wahl von Kommunikationsmitteln als ein Anforderungsmerkmal/-gegenstand von schriftlich zu bearbeitenden Aufgaben aufzufassen ist. Bezüglich der Kompetenzdefinitionen beziehen wir uns dabei auf die Bildungsstandards – unterstützt durch weitere fachdidaktische Konzeptualisierungen (u. a. von Jordan et al. 2006), da die Bildungsstandards die verbindliche und einheitliche Vorlage für alle KLPe von NRW bilden und somit eine gemeinsame methodische Vergleichsbasis schaffen.

Die bereits erwähnte Steuerungs- bzw. Orientierungsfunktion zentraler Prüfungen, speziell derer, die insb. eine „Individualbewertung“ ausweisen (vgl. Peek 2007), ist maßgebend für die Untersuchung der ZP10-Aufgaben: Sollte sich also herausstellen, dass gemessen an unseren Kategoriensystemen die mathematischen Kompetenzen in den ZP10 nur einen geringen Stellenwert besitzen im Gegensatz zur bloßen Abfrage von Wissen und Routineverfahren, hätte dies vermutlich Folgen für die Unterrichtsgestaltung in der Sek. I (s. 2.2). Inwiefern unsere Analysebefunde zu den ZP10-Aufgaben aufgrund des gewählten und fachdidaktisch aus der Literatur abgeleiteten Kompetenzreferenzrahmens dann praktisch aussagekräftig sind bzw. praktische Implikationen liefern können, wird abschließend diskutiert.

2 Theorie

An die Arbeiten des KOM-Projekts bezüglich mathematischer Kompetenzen anknüpfend, gehen wir zunächst bildungssystemunabhängig auf die Frage ein, was es bedeuten kann, Mathematik zu beherrschen bzw. in Mathematik kompetent zu sein (Abschn. 2.1). Wir zeigen auf, inwiefern diese zentralen Kompetenzen einen konzeptionellen Einfluss auf die curricularen Vorgaben für den Mathematikunterricht in Deutschland und Nordrhein-Westfalen haben (2.2.1) und gehen danach auf die Organisation der ZP10 in NRW ein (2.2.2). Da insbesondere die Aufgaben zentraler Abschlussprüfungen als unterrichtliches Vorbild für Lehr- und Lernanforderungen dienen, sind sie damit auch steuerungswirksam in der Schulausbildung; dieser sogenannte Washbackeffekt wird abschließend betrachtet (2.2.3).

¹ Neubrand und Neubrand (2010, S. 171) sprechen in diesem Zusammenhang von einer *Inkongruenz* des Kernlehrplans zu den Bildungsstandards.

2.1 Mathematische Kompetenzen

Wie in der Einleitung dargelegt, wurde mit Einführung der Bildungsstandards und der daran ausgerichteten NRW KLPe die Kompetenzorientierung zum zentralen Kriterium für die Gestaltung und Bewertung von Lehr-Lernprozessen. Der, der Neuausrichtung zugrundeliegende, Kompetenzbegriff fußt auf lernpsychologischen Konzepten. So definiert Weinert (2001, S. 27) Kompetenzen als „die bei Individuen verfügbaren oder von ihnen erlernbaren kognitiven 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.“ Kompetenzorientierung wird somit als eine Verbindung und Verknüpfung des kognitiven Wissens und Könnens mit motivationalen, volitionalen und sozialen Aspekten und Fähigkeiten verstanden. Abweichend von Weinert wird der Kompetenzbegriff bei anderen Autoren funktional bestimmt, indem bestimmte Klassen von Anforderungen betrachtet werden (u. a. Hartig und Klieme 2006, S. 129; Niss und Højgaard 2019; Stacey und Turner 2015) Im fachbezogenen Kontext, das heißt in bestimmten Kontexten und Situationen, wie sie im Rahmen dieser Arbeit untersucht werden, wird der Kompetenzbegriff nach Hartig und Klieme (2006) a) im kognitiven Sinne und b) in Bezug auf bestimmte Arten von Herausforderungen verwendet. Für Situationen, in denen die Herausforderungen mathematischer Natur sind, kann der Begriff der Kompetenz wie folgt konkretisiert werden: „*Mathematical competence* is someone’s insightful readiness to act appropriately in response to all kinds of *mathematical* challenges pertaining to given situations“ (Niss und Højgaard 2019, S. 12). Derartige Situationen können in inner- wie auch in außermathematische Kontexte eingebettet sein. Da diese Definition jedoch alle Arten mathematischer Herausforderungen einschließt, ist es erforderlich, die Situationen und die damit verbundenen Herausforderungen zu differenzieren, zu deren Bewältigung mathematische Kompetenz befähigen soll. Derartige herausforderungsbezogene Ausprägungen mathematischer Kompetenz (engl. „*mathematical competence*“) werden als mathematische Kompetenzen („*mathematical competency*“) bezeichnet (s. Leuders 2014). Niss und Højgaard (2019) unterscheiden im Anschluss an eine theoretische und erfahrungsbasierte Analyse acht, in *zwei Gruppen* trennbare, mathematische Kompetenzen, die die Gesamtheit mathematischer Aktivität umfassen sollen. Die *erste Gruppe* umfasst die Fähigkeit, mathematische Fragen zu stellen und zu beantworten. Hierzu gehören laut Niss und Højgaard (ebd.) die folgenden vier, nicht disjunkten mathematischen Kompetenzen:

- **mathematisches Denken:** Bei dieser Kompetenz geht es um die Fähigkeit, mathematische Untersuchungen vorzunehmen. Dies beinhaltet z. B., mathematikspezifische Fragen zu stellen und sich auf solche beziehen zu können (Niss und Højgaard 2019, S. 15; vgl. KMK 2003a, S. 8; Schoenfeld 1992, S. 335). Als Beispiele wird u. a. das Stellen von Fragen wie „Existieren ...?“, „Wenn ja, unter welchen Bedingungen [gilt ...]?“, „Ist es möglich, dass ...?“ und „Wenn ein Objekt Eigenschaft A besitzt, besitzt es dann notwendigerweise auch Eigenschaft B?“ genannt. Während die konstituierenden Prozesse und Aktivitäten dieser Kompetenz in der Literatur

bereits umrissen und typisiert wurden (s. Cuoco et al. 1996; Mason et al. 2010), fehlen bisher Arbeiten, die den Aspekt der Anforderungsniveaus dieser Prozesse in den Blick nehmen. Ein praktisch umsetzbarer Versuch, Niveaubstufungen darzulegen, finden sich lediglich im KMK-Entwurf (2003a).

- **mathematisches Problemlösen:** Niss und Højgaard (2019, S. 15 f.) beschreiben diese Kompetenz als das Aufwerfen und Lösen mathematischer Probleme und zählen dazu auch die Analyse und kritische Reflexion eigener und fremder Lösungsversuche sowie die Anwendung von Problemlösestrategien. Mit Problemen bzw. Problemlöseaufgaben meinen sie Aufgaben, die vom Lernenden nicht schematisch abgearbeitet werden können, womit ihr Problembegriff zu der geläufigen Definition in der Psychologie und der Mathematikdidaktik passt (vgl. Dörner 1987; Rott 2013; Schoenfeld 1992). Ob und inwieweit die problemhaltigen Situationen unterschiedlich komplexe Herausforderungen darstellen, kann mittels Anforderungsniveaus abgebildet werden (Blum et al. 2010). So wird im Falle einer reinen Wiedergabe bekannten Wissens die Problemlösesituation als wenig herausfordernd eingestuft; wird hingegen ein Zusammenfügen vorhandener Elemente eingefordert, so dass neues Wissen entsteht, das für die Lösung des Problems erforderlich ist, so wird das Anforderungsniveau der Problemlöseaufgabe als hoch eingestuft (vgl. KMK 2004).
- **mathematisches Modellieren:** Nach Niss und Højgaard (2019, S. 16) bezieht sich diese Kompetenz auf die Nutzung von Modellen zur Analyse und Bewältigung von außermathematischen Situationen. Sie berufen sich hinsichtlich der Teilschritte des Modellierens u. a. auf den Modellierungskreislauf von Blum und Leiß (2005), der als Konsens in großen Teilen der Mathematikdidaktik angesehen werden kann (vgl. Greefrath et al. 2013; Kaiser und Sriraman 2006). Unter anderem in den Arbeiten von Neubrand et al. (2002) oder auch Turner et al. (2015) werden spezifische Aktivitäten identifiziert, die eine anforderungsbezogene Abstufung des Modellierens erlauben. Diese reichen von der Nutzung vertrauter und direkt erkennbarer Standardmodelle über die die Vornahme mehrschrittiger Modellierungen innerhalb weniger unklar formulierter Einschränkungen bis hin zur Bildung von Modellen zu einer komplexen Situation (ebd.). Dabei erfordert die Modellierung von Stufe zu Stufe „in wachsendem Maße komplexe Konstruktionsleistungen.“ (Neubrand et al. 2002, S. 106).
- **mathematisches Argumentieren:** Während in der disziplinübergreifend gesichteten Literatur durchaus mehrere Auffassungen vom Argumentieren verwendet werden (Budke et al. 2015), wird in der Mathematikdidaktik mathematisches Argumentieren häufig verstanden als der „im Unterricht stattfindende soziale Prozess, bestehend aus dem Anzeigen eines Begründungsbedarfs und dem Versuch, diesen Begründungsbedarf zu befriedigen“ (Schwarzkopf 2000, S. 240). Niss und Højgaard (2019, S. 16) betonen ausschärfend die produktive und rezeptive Facette dieser Kompetenz, indem sie Argumentieren als das Produzieren und Analysieren von Argumenten auffassen; mit Argumenten sind hier vor allem Ketten von Aussagen gemeint, die durch Schlussfolgerungen miteinander verbunden sind. Neubrand et al. (2002) erfassen Argumentieren dichotom und weisen anhand von PISA-Aufgaben nach, dass das Einfordern dieser Aktivität schwierigkeitsgenerierend wirkt. Die Bandbreite der Anforderungen differenzierend wird die Argu-

mentationsintensität beispielsweise in den Bildungsstandards oder im Rahmen der COACTIV-Studie auf drei bzw. vier Anforderungsbereichen bzw. Niveaus erfasst (Jordan et al. 2006; KMK 2003b). Diese bestimmen sich im Wesentlichen dadurch, inwieweit und in welcher Länge und Komplexität Argumentationsketten selbst entwickelt bzw. nachvollzogen werden müssen (Jordan et al. 2006; KMK 2003b).

Nach Niss und Højgaard (2019) ist von der *ersten* eine *zweite Gruppe* von Kompetenzen zu unterscheiden, welche die Fähigkeiten des Umgangs mit der Sprache, den Konstrukten und Werkzeugen der Mathematik umfasst (s. Einleitung, warum hier nicht näher darauf eingegangen wird). Hierzu zählen: Der Umgang mit Werkzeugen, Das Kommunizieren, Der Umgang mit mathematischen Symbolen sowie Der Umgang mit Darstellungen mathematischer Einheiten. Sie werden im Rahmen dieser Studie nicht untersucht (s. Kap. 1).

Aus den Beschreibungen der ersten Kompetenz-Gruppe wird deutlich, dass mathematische Kompetenzen im obigen Sinn den Charakter spezifischer fachbezogener Konstrukte haben, die mit Hilfe von Messinstrumenten operationalisiert werden können. Für die im Fokus dieser Arbeit stehenden Kompetenzen Problemlösen, Modellieren und Argumentieren kann weitergehend gefolgert werden, dass diese Graduierungen der spezifischen Fähigkeiten in Kompetenz-Niveaus erlauben. Die Nutzung der Kompetenz-Niveaus als *Analysemittel* im Rahmen von empirischen Untersuchungen (PISA oder in Implementationsstudien (z. B. bzgl. der Bildungsstandards)) ermöglicht es festzustellen, inwieweit Aufgaben unterschiedliche Anforderungen bezüglich der einzelnen Kompetenzen stellen und im Fall der Aufgabenbearbeitung, inwiefern jemand in Bezug auf die zentralen mathematischen Aktivitäten kompetent ist (Blum et al. 2019; Hartig und Klieme 2006; Kühn und Drüke-Noe 2013).

2.2 Lehrpläne und Zentrale Prüfungen

In diesem Abschnitt gehen wir zunächst auf den Stellenwert der mathematischen Kompetenzen in den deutschen Lehrplänen ein, bevor wir die ZP10 in NRW und etwas allgemeiner die Steuerungsfunktion solcher Prüfungsformen thematisieren.

2.2.1 Lehrpläne und Kompetenzen

Die länderübergreifenden Bildungsstandards für das Fach Mathematik, vereinbart als Reaktion auf den sogenannten PISA-Schock, greifen die Kompetenzorientierung explizit auf und stellen „prozessbezogene Kompetenzen“ in den Vordergrund. Sie werden bundesland-konkret adaptiert und von den dort gültigen Curricula (je nach Bundesland Kernlehrpläne oder Kerncurricula) aufgegriffen. Dies gilt folglich ebenso für das im Rahmen dieser Studie untersuchte Schulsystem Nordrhein-Westfalens. Hier bilden die Bildungsstandards im Fach Mathematik beispielsweise die inhaltlich-konzeptionelle Grundlage für die kompetenzorientierten Mathematik-KLPe der Sek. I der fünf Schulformen – Hauptschule (MSW 2011), (Abend-)Realschule (MSW 2004a), Gesamtschule, Sekundarschule (MSW 2004b) und Gymnasium (MSW 2019a) – mit dem Bildungsziel des Mittlerer Schulabschlusses, weshalb die

KLPe im Fach Mathematik über all diese Schulformen hinweg „große Schnittmengen aufweisen“ (Büchter und Pallack 2012, S. 63).

Die konkrete Ausbuchstabierung der Kompetenzen ist zwischen dem Bund und den einzelnen Bundesländern sowie in verschiedenen Jahren des Inkrafttretens nicht immer konsistent: In einer Anhörungsfassung (KMK 2003a) waren die mathematischen Kompetenzen noch direkt am KOM-Projekt von Niss orientiert, indem alle als bei Niss (2003) „allgemein“ bezeichneten Kompetenzen aufgeführt wurden; die offiziell veröffentlichte Version (KMK 2004)² sieht sechs „prozessbezogene“ Kompetenzen vor: (K1) mathematisch argumentieren, (K2) Probleme mathematisch lösen, (K3) mathematisch modellieren, (K4) mathematische Darstellungen verwenden, (K5) mit symbolischen, formalen und technischen Elementen der Mathematik umgehen und (K6) kommunizieren. In NRW kursieren verschiedene Zuschnitte dieser Liste, in der Regel mit vier (aus den sechs der KMK zusammengelegten) Kompetenzen (z. B. MSW 2004a). Das an den Bildungsstandards prozessual und inhaltlich orientierte Kompetenzmodell der KLPe ist eine Kompetenzmatrix, die mittels zwei „Dimensionen“ aufgespannt wird:

Prozessbezogene Kompetenzen Sie werden in der Auseinandersetzung mit mathematischen Inhalten erworben und umfassen die Tätigkeiten, die in ihrer Breite das Spektrum des mathematischen Arbeitens ausmachen (Wittmann und Müller 2007). Hierzu zählen zumeist: Argumentieren/Kommunizieren, Problemlösen, Modellieren und Werkzeuge (MSW 2004a, 2004b, 2011, 2019a).

Inhaltsbezogene Kompetenzen Mathematische Inhalte, anhand derer die prozessbezogenen Kompetenzen konkretisiert werden, sind jeweils inhaltsbezogenen Kompetenzen zugeordnet. Diese durchziehen das mathematische Curriculum spiralförmig (als Kompetenzprofile nach der 6., 8. und 10. Klasse) und begünstigen so, sachgebietsübergreifendes und vernetztes Denken sowie ein Verständnis grundlegender mathematischer Begriffe zu erreichen. Hierzu zählen Inhaltsbereiche *Arithmetik/Algebra*, *Geometrie*, *Funktionen* und *Stochastik* (MSW 2004a, 2004b, 2011, 2019a).

Die Bildungsstandards definieren noch eine dritte Kompetenzdimension, die nicht von den nordrheinwestfälischen KLPe berücksichtigt wird (vgl. MSW 2004a, 2004b, 2011):

Anforderungsbereiche Kompetenzen, die im Lösungsprozess eines mathematischen Problems eingefordert werden, können auf unterschiedlichen kognitiven Anspruchsniveaus angesiedelt sein. Die Dimension Anforderungsbereiche operationalisiert niveauspezifisch drei Stufen kognitiven Anspruchs: *Reproduzieren (Stufe I)*, *Zusammenhänge herstellen (Stufe II)* sowie *verallgemeinern und reflektieren (Stufe III)* (ebd.).

Bei der Konkretisierung des zweidimensionalen Kompetenzmodells in den KLPe wird seitens der Bildungspolitik der Anspruch erhoben, die Bildungsstandards im Fach Mathematik für den Mittleren Schulabschluss schulformübergreifend als ein

² Die sieben „prozessbezogenen Kompetenzen“ der aktuellen Fassung der Bildungsstandards (KMK 2022) sind nicht Gegenstand der vorliegenden Studie.

einheitliches Mindestniveau aller fünf Schulformen bzw. ihrer Bildungsgänge zu verankern, die diesen Abschluss verleihen. Über dieses Mindestniveau hinausgehend, werden hier zudem in Form eines curricularen Additums schulformspezifische Profile verankert (MSW 2011).

Die bisherigen Ausführungen zeigen, dass die Bildungsstandards (und damit auch das Kompetenzmodell von Niss und Højgaard) tiefgehende Orientierungen bezüglich zu erlernender bzw. zu vermittelnder mathematischer Prozesse und Aktivitäten bieten als die KLPe, die die Kompetenzen verkürzt darstellen. So führen die nordrhein-westfälischen KLPe zumeist nur vier prozessbezogene Kompetenzen auf; eine „Anspruchsdimension“ (Blum et al. 2010, S. 19), die es erlaubt, unterschiedliche Kompetenzniveaus zu unterscheiden und zu bewerten (s. 2.1), bleibt unberücksichtigt. Damit hat die Referenz auf die Bildungsstandards bei einer Untersuchung der Kompetenzanforderungen von Mathematikaufgaben erhöhte Aussagekraft.

Die NRW-KLPe sind ihrerseits ein zentraler Bestandteil eines landesinternen Gesamtsystems zur Steuerung standardorientierter Unterrichtsentwicklung. Tragende Elemente dieses Systems sind u.a. Parallelarbeiten, zentrale Abschlussprüfungen (ZP10), Schulprogrammarbeit sowie interne und externe Evaluationen. Bezogen auf die standardorientierte Leistungsmessung mittels der ZP10 zeigt sich diese Wirkung darin, dass man im Rahmen der Aufgabenentwicklung beansprucht, das jeweilige Kompetenz-Konstrukt der NRW KLPe und somit auch der zugrundeliegenden Bildungsstandards angemessen zu operationalisieren (Büchter und Pallack 2012).

2.2.2 ZP10 in NRW

Die ZP10 bilden seit ihrer Einführung im Schuljahr 2006/07 das schriftliche Prüfungsverfahren für den Abschluss der Sekundarstufe I (Klasse 5–10) an weiterführenden Schulformen in NRW und werden zentral vom Ministerium für Schule und Weiterbildung gestellt. Sie sind damit tragende Elemente eines landesinternen Systems standardorientierter Unterrichtsentwicklung. Die fünf Schulformen, an denen die ZP10 geschrieben wird, sind in Tab. 1 mit den jeweiligen Voraussetzungen sowie den – im Sinne der Durchlässigkeit des Schulsystems (s. MSW 2019b, § 12–16) – Möglichkeiten des Abschlusserwerbs zusammengefasst.

Die ZP10 ist die einzige Prüfung, die in allen fünf weiterführenden Schulformen Nordrhein-Westfalens als verpflichtend deklariert ist bzw. war: für Gymnasien galt dies nur bis zum Schuljahr 2010/2011; seitdem haben diese infolge der Schulzeitverkürzung durch G8 nicht mehr an der ZP10 teilgenommen.³ Die in der ZP10 erreichte Note geht mit 50% in die Abschlussnote des jeweiligen Faches ein. Die Prüfung umfasst, abhängig von der angestrebten Abschlussart (Tab. 1), im Fach Mathematik 90 (HSA) bzw. 120 min (FOR) Bearbeitungszeit. Die Prüfung ist zweigeteilt: Im ersten Prüfungsteil – mit maximal sieben Aufgaben, teilweise in mehrere Teilaufgaben untergliedert – werden sogenannte „Basiskompetenzen“ (MSW 2006, S. 4) in nicht aufeinander bezogenen Aufgaben abgeprüft (Bearbeitungsdauer: 30 min). Im zweiten Prüfungsteil – mit maximal vier Aufgaben, die teilweise mehrere Teilaufgaben in außer- und innermathematischen Kontexten beinhalten – gibt es Aufgaben, die

³ Im Zuge der Umstellung des Gymnasiums auf G9 wird dort ab 2024 auch die ZP10 wieder eingeführt.

Tab. 1 Nordrhein-westfälische Schulformen der Sekundarstufe I, die dort vergebenen Abschlüsse und ihre Voraussetzungen (MSW 2019b)

Abschlussart	Schulformen der Sekundarstufe I				
	Hauptschule	Realschule	Gesamtschule	Sekundarsch	Gymnasium
<i>Hauptschulabschluss nach der Klasse 10 (HSA)</i>	Ja (Typ A: Hauptschul- zweig)	Nein	Ja (falls versetzt)	Ja (falls ver- setzt)	Nein
<i>Realschulabschluss bzw. Fachoberschulreife (FOR)</i>	Ja (Typ B: Realschul- zweig)	Ja (falls versetzt)	Ja (mind. zwei Fächer mit Erweiterung)	Ja (Bildungsgang Realschule)	Ja (falls ver- setzt)

Inhalte aus dem Jahrgangsstufen 9 und 10 voraussetzen (Bearbeitungsdauer: 60 bzw. 90 min).

2.2.3 Washbackeffekte durch zentrale Prüfungen

Einflüsse zentraler Prüfungen auf unterrichtliches Handeln werden als Washback- oder Rückkoppelungseffekte bezeichnet (Prodromou 1995) und können nach Watanabe (2004) in ihrer Wirkungsausrichtung positiv, neutral oder negativ sein. Die zumeist in der internationalen Erziehungswissenschaft und vor allem in der Linguistik angesiedelten Untersuchungen zeigen, dass sich die von Kühn (2010) als *positiv* eingestufteten *Washbackeffekte* durch zentrale Prüfungen in der Berücksichtigung neuartiger Inhalte, Aufgabenformate und Methoden zeigen. Im Hinblick auf Mathematikunterricht können zentrale Prüfungen einen Beitrag zur Professionsentwicklung der Lehrkräfte leisten, welcher sich nach Richter et al. (2014) beispielsweise im positiven Unterstützungsverhalten (bspw. in Form von Leistungsdifferenzierung) gegenüber den Lernenden oder auch einer verstärkten Kompetenzorientierung des erteilten Unterrichts zeigt. Richter et al. (Richter et al. 2014) stellten zudem fest, dass Lehrende vermehrt Unterrichtszeit nutzen, um innovative Elemente zentraler Prüfungen in die Lernprozesse zu integrieren, womit „Orchideen-Themen“ einzelner Lehrkräfte reduziert werden und der Unterricht eine gesteigerte Konformität mit dem der Prüfung zugrunde liegenden Curriculum aufweist (vgl. Kühn 2010).

Eine als *negativ* wahrgenommene *Wirkungsrichtung* der Washbackeffekte wird zumeist in Verbindung mit High-Stakes-Testing thematisiert. Unter High-Stake-Tests sind Prüfungen zu subsumieren, deren Ergebnisse für die Bildungsbeteiligten (Lernende, Lehrende, Schulleitungen) zukünftige Konsequenzen besitzen. Hinsichtlich der Zukunftsbedeutung von Tests liefern empirische Studien Hinweise darauf, dass der „Grad der stakes (low, moderate oder high stakes)“ (Kühn 2010, S. 72) neben der Wirkungsrichtung (positiv vs. negativ) zugleich auch den Wirkungsgrad (niedrig vs. hoch) beeinflusst (Alderson und Wall 1993; Cheng 2005). Zentrale Tests wie die ZP10 werden als High-Stakes-Tests charakterisiert, da sie auf der individuellen Ebene Qualifikations- und Selektionsfunktion aufweisen. Zu den negativen Washbackeffekten in Fächern mit High-Stakes-Testing (z.B. in Mathematik und Englisch) zählt in erster Linie, dass ein wesentlicher Teil der Unterrichtszeit für das Einüben prüfungsrelevanter Inhalte und Testformate (bspw. Operatorenverständnis)

verwendet wird (Kühn 2010; Madaus et al. 2009). Für Mathematik berichten u. a. Richter et al. (2014) in Bezug auf den Low-Stakes-Test VERA von einer teilweisen Verengung des Curriculums auf die im Test enthaltenen Themen. Einem vergleichbaren Forschungsansatz folgend, wird dies von Konarzewski (2008) ebenfalls für die polnische Mittelschulprüfung (High-Stakes-Test) bestätigt. Studien zeigen weitergehend, dass eine derartige Einflussnahme sich durchaus auf mehrere Schuljahrgänge erstrecken kann (Bachman und Palmer 1996), sodass davon ausgegangen werden kann, dass sich der Unterrichtsstoff längerfristig und in beträchtlichem Maße am Prüfungsstoff und damit nur eingeschränkt an den verbindlichen (Kern-)Lehrplänen bzw. allgemeinen Lernzielen des Faches orientiert.

Insgesamt scheint der Wirkungsgrad der ZP10 als ein typischer High-Stakes-Test hoch zu sein. Schlussfolgerungen über die Wirkungsausrichtung sind hingegen weniger eindeutig, wenngleich die nicht intendierte Verengung des Curriculums hinsichtlich der Qualität und Intensität als schwerwiegend einzuschätzen ist (vgl. Maier 2010).

3 Forschungsfragen

Angesichts der besonderen Bedeutung der ZP10 (2.2.2) und ihrer Wirkung auf die Unterrichtsgestaltung (2.2.3), erscheint eine Analyse und Evaluation der in 2.1 identifizierten Prozesse und Aktivitäten besonders angebracht. Aus diesen beiden funktionellen Strängen (Orientierung der Lernprozesse an Kompetenzen und Steuerungswirksamkeit der ZP10) ergeben sich folgende Forschungsfragen:

1. Inwiefern charakterisieren die Kompetenzen „Problemlösen“, „Modellieren“ und „Argumentieren“ die Aufgaben der ZP10 im Zeitraum von 2007 bis 2019?

Konkret geht es um die Frage, welche der genannten Kompetenzen in den ZP10-Aufgaben auf welchem Kompetenzniveau eingefordert werden. Bei diesen Analysen geht es nicht nur um die Erfassung bestimmter Aufgabenmerkmale, sondern auch um die Feststellung von Schwerpunktsetzungen auf bestimmte bzw. innerhalb bestimmter Kompetenzen. Um sicherzugehen, dass die im Mittel identifizierten Kompetenzanforderungen nicht in verschiedenen Jahren starken Schwankungen unterliegen, folgt direkt:

2. Welche Unterschiede sind in den Aufgaben der ZP10 im untersuchten Zeitraum bezüglich der Kompetenzniveaus der Kompetenzen „Problemlösen“, „Modellieren“ und „Argumentieren“ feststellbar?

Frage (2) stellt sich auch mit Blick auf die Gleichwertigkeit schulischer Ausbildung und die zeitliche Vergleichbarkeit der hier untersuchten Abschlüsse. Anders formuliert geht es darum, ob Trends bezüglich einzelner Kompetenzen erkennbar sind – wohlwissend, dass der Stellenwert prozessbezogener Kompetenzen in zentralen Prüfungen einem Wandel unterworfen sein kann, wie es beispielsweise die Diskussion um das Hamburger Zentralabitur (Kaiser und Busse 2014) oder Ände-

rungen in den curricularen Vorgaben in Polen (Scheja 2019) zeigen. In NRW sind diesbezüglich keine signifikanten Änderungen zu erwarten (s. 2.2).

3. Inwiefern unterscheiden sich die bei der Bearbeitung der ZP10-Aufgaben erforderlichen Kompetenzniveaus der Kompetenzen „Problemlösen“, „Modellieren“ und „Argumentieren“ in Bezug auf die Abschlussarten HSA und MSA?

In NRW können nach Jahrgang 10 zwei verschiedene Abschlussarten erworben werden (Tab. 1). Der Fokus der Untersuchung wird dabei auf Unterschiede zwischen den Prüfungsvarianten zum Hauptschulabschluss und zur Fachoberschulreife gelegt. Mithilfe der Analyse der ZP10-Aufgaben vor dem Hintergrund des gewählten Kompetenzmodells als gemeinsamen Referenzrahmen soll untersucht werden, ob bzw. inwiefern die ZP10 diese äußere Differenzierung abbilden. Passen Unterschiede in den Kompetenzanforderungen der Testaufgaben zur angenommenen Ordinalität der Leistungsniveaus der Schulformen nach „Wertigkeit“ des zu erlangenden Abschlusses, also Fachoberschulreife > Hauptschulabschluss (2.2.2)? Eine naheliegende Hypothese wäre, dass die untersuchten Testaufgaben für die Fachoberschulreife aufgrund der Vorbereitung für die gymnasiale Oberstufe im Mittel höhere Kompetenzniveaus sowohl in den einzelnen Kompetenzen als auch in der Gesamtschau haben sollten als für den Hauptschulabschluss. Es fehlen umfassende, empirische Analysen zu den genannten Aspekten in (zentralen) Testaufgaben; diese Forschungslücke soll hier geschlossen werden.

4 Methode

4.1 Datenbasis

Datengrundlage der vorliegenden Studie sind die Aufgaben des Mathematikteils der NRW ZP10 der Jahre 2007 bis 2019.⁴ Als „Aufgabe“ verstehen wir in dieser Studie gemäß Neubrand (2002, S. 16) „eine Aufforderung zur gezielten Bearbeitung eines eingegrenzten mathematischen Themas. Aufgaben sind immer Auseinandersetzung mit einem Beispiel eines Sachverhalts.“ Überträgt man dieses Begriffsverständnis auf die in dieser Studie untersuchten Aufgaben der ZP10, führt dies zu (a) Aufgaben, die auf dem Testbogen nicht weiter unterteilt werden (z.B. Aufgabe 1, wenn es keine Teilaufgaben dazu gibt) bzw. (b) Teilaufgaben, falls eine entsprechende Unterteilung der Aufforderung vorliegt (bspw. Aufgabe 4b) als einzelne Analysegegenstände. Diese Auffassung wird durch die Auswertungsanleitung für Lehrkräfte gestützt, in der in diesen beiden Fällen den Aufgaben Rohpunkte zugewiesen werden. Im Sinn einer einheitlichen Begrifflichkeit in Bezug auf Aufgaben und Teilaufgaben bezeichnen wir unsere Analysegegenstände im Folgenden als *Items*. In Summe

⁴ Die Jahrgänge 2020, 2021 und 2022 werden hier nicht berücksichtigt. Aufgrund der Corona-Pandemie wurden 2020 die ZP10-Aufgaben durch die Lehrkräfte konzipiert (intern); 2021 und 2022 wurden die Aufgaben zwar zentral konzipiert, die tatsächlich eingesetzten Aufgabensätze wurden jedoch ausnahmsweise durch die Lehrkräfte am Vortag ausgewählt.

Tab. 2 Verteilung der Items der nordrhein-westfälischen ZPI10 pro Bildungsgang und Abschlussjahr

Jahr/ Schulform	Gesamtschule (G-Kurs)	Hauptschule (Typ A)	Hauptschul- abschluss (HSA)	Mittlerer Schulab- schluss (MSA)	Abendreal- schule	Gesamtschule (E-Kurs)	Hauptschule (Typ B)	Mittlerer Schulab- schluss (Abendre- alschule)	Real- schule	Gymnasium	Summe
2007	28	28	-	-	-	27	27	-	27	24	161
2008	26	26	-	-	-	26	26	-	26	22	152
2009	25	25	-	-	28	29	29	-	29	28	193
2010	29	29	-	-	32	31	31	-	31	28	211
2011	25	25	-	-	27	27	27	-	27	-	158
2012	25	25	-	-	24	26	26	-	26	-	152
2013	29	29	-	-	33	33	33	-	33	-	190
2014	27	27	-	-	31	31	31	-	31	-	178
2015	25	25	-	-	30	30	30	-	30	-	170
2016	-	-	27	29	-	-	-	29	-	-	85
2017	-	-	24	27	-	-	-	27	-	-	78
2018	-	-	26	27	-	-	-	27	-	-	80
2019	-	-	24	27	-	-	-	27	-	-	78
GES	HS-Abschluss: 579										
	FOR-Abschluss: 1205 (RS) + 102 (GYM)										
	1886										

In den Spaltenüberschriften finden sich die offiziellen Zuweisungen zu den Schultypen aus dem jeweiligen Kopf der Prüfungsdokumente. In der untersten Zeile sieht man, wie diese Daten zu zwei Bildungsabschlüssen zusammengefasst werden: Hauptschulabschluss und Fachoberschulreife, Jg. 10

aller Schulformen und Jahrgänge gibt es 68 ZP10-Prüfungsdokumente (Tab. 2) mit insgesamt 1886 Items, die für diese Studie analysiert wurden.

Die in Tab. 2 aufgeführte Auflistung bedeutet allerdings nicht, dass es fast 1900 verschiedene Items gibt. Im Gegenteil, in den Jahren (zeilenweise) sind die Items, die zum HS-Abschluss führen, jeweils identisch. Die Items, die zum FOR-Abschluss führen (ohne Gymnasien), sind jeweils identisch mit Ausnahme der Abendrealschulen: Neben veränderten Anreden (Siezen statt Duzen) in allen Dokumenten gibt es hier in den Jahren 2009, 2010 und 2012 jeweils zwei bis vier veränderte Items im Vergleich zu anderen Testversionen. Die Items für Gymnasien stimmen zu guten Teilen mit den Items für Realschulen überein. In allen vier Jahrgängen gibt es aber veränderte Teilaufgaben (=Items) und teilweise komplett ausgetauschte Aufgaben (mit mehreren Teilaufgaben). Da die hier vorgestellte Auswertung im Sinne einer Dokumentenanalyse alle vorhandenen Dokumente beinhalten soll, erfolgt kein Versuch, Aufgabenähnlichkeit zu operationalisieren. In Bezug auf die Forschungsfragen (1) und (2) werden daher alle Items ausgewertet. In Bezug auf Forschungsfrage (3) wird für jeden Abschluss und jedes Jahr ein Itemsatz ausgewählt: „Hauptschule (Typ A)“ bzw. „Hauptschulabschluss (HSA)“ sowie „Realschule“ bzw. „Mittlerer Schulabschluss (MSA)“, insgesamt 812 Items.

4.2 Analyse der Kompetenzen

Die ZP10-Aufgaben wurden forschungsmethodisch einer Dokumentenanalyse, genauer einer Aufgabenanalyse, unterzogen. Die hier verwendete rationale Aufgabenanalyse stellt theoretisch die „ideale Performanz“ eines Aufgabenlösenden fest (Resnick und Ford 1984, S. 84). Den theoretischen Rahmen für die rationale Aufgabenanalyse der ZP10 lieferten die Analyseschemata für mathematische Aufgaben der COACTIV-Studie (vgl. Jordan et al. 2006), die Aufgabendimensionen nach Maier et al. (2010) sowie die von der KMK formulierten Anforderungsbereiche für unterrichtliche Mathematikaufgaben „Reproduzieren“, „Zusammenhänge herstellen“, „Verallgemeinern und Reflektieren“ (2004, S. 11 ff., 2003b, S. 13 ff.; s. auch 2.2.1). Zur Erstellung eigener an die Untersuchungszwecke angepasster Kategoriensysteme (s. unten) wurden zunächst die Items mittels der genannten Aufgabentaxonomien deduktiv, theoriegeleitet analysiert (2.1), wobei erste Oberflächenmerkmale von Aufgaben bzw. Differenzierungsmöglichkeiten festgestellt werden konnten. Unabhängig davon wurden konkrete ZP10-Prüfungs- sowie Schulbuchaufgaben hinsichtlich der erforderlichen mathematischen Aktivitäten ausgewertet.

Das beschriebene Analyseverfahren führte dazu, dass (neue) ordinalskalierte Kategoriensysteme – aufgefasst als Kompetenzniveaus – zu den Kompetenzen „Problemlösen“, „Modellieren“ und „Argumentieren“ erstellt, Analysekatoren definiert, ausdifferenziert und voneinander abgegrenzt werden konnten. In „Rücklaufschleifen“ (Mayring 2010) wurden die aufgestellten Kategorien u. a. an den Items weiter erprobt und revidiert. Die Kategoriensysteme operationalisieren die unterschiedlichen Kompetenzniveaus der betrachteten Kompetenzen in je vier Komplexitätsstufen.

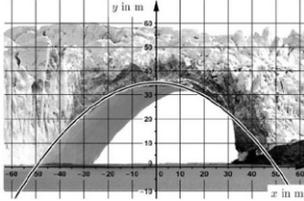
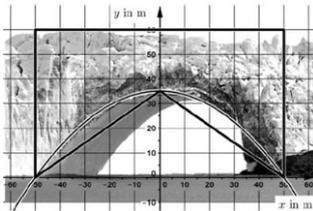
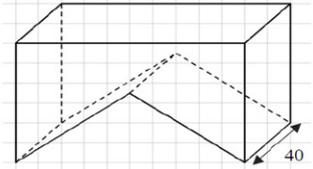
Drei Beispielitems (ZP10, 2017, MSA, A. 3 a-c)	Kodierung
<p>Gletschereis-Brücke Am Moreno-Gletscher in Argentinien gab es eine Brücke aus Eis. Sie entstand, weil Wasser den Gletscher unterhöhlt hat. Am 10.03.2016 ist die riesige Eis-Brücke eingestürzt (siehe Fotostrecke).</p>  <p>Der Brückenbogen konnte annähernd mit einer Parabel beschrieben werden (Abbildung 1).</p>  <p style="text-align: center;">Abbildung 1: Eis-Brücke am Moreno-Gletscher, Bogen durch eine Parabel angenähert.</p>	
<p>a) Entnimm der Abbildung 1 die Höhe h über dem Wasserspiegel und die Spannweite des parabelförmigen Brückenbogens.</p>	<p>P0 M1 A0</p>
<p>b) Bestimme die Funktionsgleichung der Parabel, die den Brückenbogen beschreibt, in der Form: $f(x) = ax^2 + c$.</p> <p>Rico möchte schätzen, wie viele Kubikmeter Eis bei dem Einsturz der Brücke ins Wasser fielen. Er kann aber Flächen, die durch eine Parabel begrenzt werden, nicht berechnen. Deshalb zeichnet er Hilfslinien ein (Abbildung 2) und fertigt eine Skizze an (Abbildung 3).</p>  <p style="text-align: center;">Abbildung 2: Hilfslinien zur Idee von Rico</p>  <p style="text-align: center;">Abbildung 3: Skizze zur Berechnung</p>	<p>P1 M1 A0</p>
<p>c) Wird mit Ricos Idee die eingestürzte Eismenge zu groß oder zu klein geschätzt? Begründe deine Entscheidung.</p>	<p>P0 M1 A1</p>

Abb. 1 Drei Beispielitems mit Kodierungen in Bezug auf alle drei Kategoriensysteme

Im Folgenden werden die verwendeten Kategoriensysteme vorgestellt und an Beispielen verdeutlicht (Abb. 1). Weitere, ausführlich kommentierte, Beispiele zu den einzelnen Kategorien befinden sich im Anhang.

4.2.1 Problemlösen

Das Kategoriensystem zum Problemlösen (Tab. 4) basiert u. a. auf den von Maier et al. (2010) beschriebenen Dimensionen „Kognitive Prozesse“ und „Repräsentationsformen des Wissens“. Hierbei wird der Umgang mit Repräsentationsformen im Sinne einer heuristischen Strategie integriert. Die Operationalisierung orientiert sich an der Tätigkeit des „innermathematischen Modellierens“ aus der COACTIV-Studie (Jordan et al. 2006) sowie auf den Anforderungsbereichen der KMK-Vorgaben (2003b) zum „Problemlösen“. Das Kompetenzniveau des Problemlösens nimmt hier mit der Neuartigkeit, der Vielzahl und Vielfalt der auszuführenden Lösungsschritte sowie mit dem erforderlichen flexiblen Umgang mit Wissen (bspw. beim Transformieren von Darstellungsformen) zu. Die Kategorie „P0 – kein Problemlösen erforderlich“ bezieht sich auf Routineaufgaben; damit sind Aufgaben gemeint, die in den Schulbüchern der Sekundarstufe I standardmäßig und wiederholend vorkommen, um eine Routine im Umgang mit „grundlegenden Begriffen, Sätzen und Verfahren in einem abgegrenzten Gebiet“ (KMK 2003b, S. 13) zu vermitteln (s. Anhang). Kategorie „P1 – Naher Transfer/niedrige Komplexität“ grenzt Aufgaben von der Kategorie „P0“ ab, die mehrere bekannte Rechenverfahren kombinieren und oder deren Lösung aus gut bekannten Aufgabenlösungen abgeleitet werden kann. Aufgaben der Kategorie „P2 – Weiter Transfer/mittlere bis hohe Komplexität“ erfordern weitere problemlösespezifische Tätigkeiten, wie die die Verknüpfung von „Kenntnisse[n], Fertigkeiten und Fähigkeiten [...], die in der Auseinandersetzung mit Mathematik auf verschiedenen Gebieten erworben wurden“ (KMK 2003b, S. 13), was bedeutet, dass der Lösungsweg häufig nicht direkt ersichtlich ist, sondern zunächst durch z. B. Skizzen, systematisches Probieren oder Vorüberlegungen erschlossen wird. Die Kategorie „P3“ unseres Systems, „echtes Problemlösen“, bezieht sich auf nicht zuvor geübte Lösungsideen und -verfahren sowie auf die Notwendigkeit, heuristisch zu arbeiten (Tab. 3).

Tab. 3 Die Kompetenz Problemlösen und ihre Kompetenzniveaus

Kat.	Kategorienname	Operationalisierung
P0	Nicht notwendig/ einfache Reproduktion	Abfragen von Basiswissen. Items, die durch Anwenden einfacher Standardverfahren (z. B. Dreisatz, Flächen- oder Volumenberechnung mit bekannten Formeln etc.), die zuvor mehrfach (mit anderen Werten) bearbeitet wurden, gelöst werden können
P1	Naher Transfer/niedrige Komplexität	Die Lösung des Items ist mehrschrittig und geht über Reproduktion hinaus; in der Aufgabenstellung wird jedoch deutlich, welches Wissen im Lösungsprozess zur Anwendung kommen soll
P2	Weiter Transfer/mittlere bis hohe Komplexität	Die Problemsituation, in der das Wissen angewendet werden soll, ist bisher unbekannt. Der mathematische Ansatz (inkl. erforderlicher Größen) ist aus dem Aufgabentext evtl. nicht ersichtlich, das vorhandene Wissen muss auf passende Strukturen/Prinzipien erst durchsucht werden. Mehrschrittiges bzw. komplexes kritisches Denken, d. h. von mehreren Denkansätzen ausgehend
P3	Echtes Problemlösen	Items, bei denen Lösungsideen und Verfahren in zuvor nicht geübter Weise (hier bezogen auf den KLP) entwickelt und kombiniert werden müssen. In der Regel sind zum Finden eines Ansatzes Heuristiken nötig

4.2.2 Mathematisches Modellieren

Grundlage des Kategoriensystems zum Modellieren (Tab. 5) bilden die Dimension „Lebensweltbezug“ von Maier et al. (2010), das COACTIV-Kategoriensystem „Außermathematische Modellieren“ (Jordan et al. 2006) und die Anforderungsbereiche der KMK (2003b). Das Kompetenzniveau von Modellierungsaufgaben nimmt hier mit der Vielzahl der durchzuführenden Modellierungsschritte, mit der Anzahl der zu treffenden Schätzparameter und der Offenheit des Lösungswegs und der Ausgangssituation zu. Relevant sind für diese Kompetenz insbesondere die Modellierungsschritte „Mathematisieren“ und „Validieren“ (s. Maaß 2007; MSW 2004a, 2011; Niss und Højgaard 2019).

Anders als in den Bildungsstandards, in denen sich Anforderungsbereich I der Kompetenz „mathematisch modellieren“ auf die „Erfahrungswelt“ bezieht (vgl. KMK 2003b), gibt es sowohl bei Maier et al. (2010) als auch bei Jordan et al. (2006) eine Kategorie für Aufgaben „ohne Lebensweltbezug“ bzw. für deren Bearbeitung „außermathematisches Modellieren nicht notwendig“ ist. In Anlehnung an diese beiden Kategoriensysteme haben wir eine Kategorie „M0“ gewählt, in die alle innermathematischen Aufgaben fallen (s. Anhang). Sowohl in den Bildungsstandards als auch im Kategoriensystem von Jordan et al. gibt es in Anforderungsbereich I bzw. in der zweiten Kategorie einen Verweis auf „vertraute und direkt erkennbare Modelle“ bzw. auf ein „explizit gegebenes Modell“. Dies noch etwas verschärfend, beziehen wir – wie Maier et al. – die Kategorie „M1“ insbesondere auf eingekleidete Aufgaben, also Aufgaben, bei denen der Realitätsbezug konstruiert wirkt und für die Bearbeitung keine Modellannahmen getroffen oder Daten recherchiert bzw. geschätzt werden müssen.

Die jeweils nächste Kategorie in den Bildungsstandards und bei Jordan et al. (2006) bezieht sich auf die Komplexität der zu nutzenden Modelle, auf „mehrere Schritte“ und die Notwendigkeit zu „interpretieren“. Maier et al. (2010) machen

Tab. 4 Die Kompetenz Modellieren und ihre Kompetenzniveaus

Kat.	Kategorienname	Operationalisierung
M0	Modellieren nicht notwendig	Keinerlei Realweltbezug, Umwandeln von Größen(einheiten) zählt nicht als Realitätsbezug. Ablesen von Werten aus einem Diagramm mit Kontext hingegen schon
M1	Modellieren auf niedrigem Niveau notwendig	Kontextbezug vorhanden, aber nur als <i>Einkleidung</i> , d. h. es müssen keine Modellannahmen getroffen werden. Übersetzungen, die unmittelbar ausgeführt werden können, da das Modell explizit gegeben ist oder unmittelbar naheliegt und die Interpretation direkt möglich ist
M2	Modellieren auf mittlerem Niveau notwendig	Überschaubare Übersetzungen, die nicht unmittelbar ausführbar sind (z. B. müssen verschiedene Gegenstände miteinander in Beziehung gesetzt oder beim Mathematisieren <i>mehrere</i> Schritte durchgeführt werden). Es müssen einfache Modellannahmen selbst getroffen werden und bei Schätzaufgaben muss mit geschätzten Zwischenergebnissen weitergerechnet werden
M3	Modellieren auf hohem Niveau notwendig	Verwendete mathematische Modelle (wie Formeln, Gleichungen, Darstellungen von Zuordnungen, Zeichnungen, Ablaufpläne) vergleichen, reflektieren, kritisch beurteilen; Modell-Ergebnisse validieren; komplexe mathematische Modelle entwickeln. Es müssen komplizierte Modellannahmen selbst getroffen werden

ihre dritte Kategorie an einem „konstruierten, aber authentisch wirkenden Lebensweltbezug“ fest. In Kategorie „M2“ des hier vorgestellten Kategoriensystems geht es passend zu den angeführten Vorarbeiten um mehrschrittige Mathematisierungen und eigenständiges Treffen einfacher Modellannahmen und/oder Schätzen von Daten.

In der jeweils höchsten Kategorie geht es in den Bildungsstandards und bei Jordan et al. (2006) um die Arbeit mit komplexen mathematischen Modellen, die kritisch beurteilt und reflektiert werden müssen. Maier et al. operationalisieren ihre vierte Kategorie insb. mit einem „realen Lebensweltbezug“. In der Kategorie „M3“ unseres Kategoriensystems geht es dazu passend um komplexe mathematische Modelle, die reflektiert und kritisch beurteilt werden müssen. Ergänzend zu den Systemen aus der Literatur schlagen wir Operationalisierungen für Schätz- und Fermiaufgaben vor (Tab. 4).

4.2.3 Mathematisches Argumentieren

Das Kompetenzverständnis von mathematischem Argumentieren, das für die Erstellung des Kategoriensystems (Tab. 6) genutzt wurde, wird vornehmlich im Anhang erörtert. Unseren Operationalisierungen liegen die Anforderungsstufen von Jordan et al. (2006) zugrunde, in denen – genauso wie in den Anforderungsbereichen der KMK-Vorgaben (2003b) – *Routineargumentation* bzw. *Standardargumente*, *mehrschrittige Argumentation* und *komplexe Argumentation* unterschieden werden. Mit Standardargumenten sind aus dem Unterricht gut bekannte Argumente (bspw. Eigenschaften mathematischer Objekte, wiederkehrende Begründungen wie die Nullteilerfreiheit) gemeint. Mehrschrittige Argumentationen umfassen zwei und mehr Argumente und setzen diese begründet zueinander in Beziehung. Komplexe Argumentationen umfassen viele Argumente und neue Zusammenhänge werden begründet, die in der vorliegenden Form im Unterricht noch nicht behandelt wurden, mit den erlernten Mitteln aber verstanden bzw. hergeleitet werden können.

Im Unterschied zu den Anforderungsbereichen der KMK-Vorgaben (2003b) gibt es bei Jordan et al. (2006) eine Kategorie für Aufgaben, die kein mathematisches Argumentieren einfordern. Jordan et al. (ebd.) beschreiben jedoch nicht, wie sich Argumentationsaufgaben dieser Kategorie „Argumentation nicht notwendig“ von Aufgaben der Kategorie „Argumentation auf niedrigerem Niveau notwendig“ unterscheiden. Unsere Untersuchung der ZP10-Aufgaben hat gezeigt, dass Aufgaben ohne Argumentationsanforderungen z.B. das Eintragen einer Größer-Kleiner-Relation oder das bloße Angeben eines Rechenergebnisses sein können. Die Frage, die sich bei der Aufgabeneinstufung in die Kategorie „A0 – keine Argumentation notwendig“ – stellt, ist: „was kann als Argumentieren gelten?“ Mindestanforderung ans Argumentieren ist nach den Anforderungsbereichen der KMK-Vorgaben (2003b, Anforderungsbereich I: *Reproduzieren*) die Reproduktion von Sätzen, Verfahren, Rechnungen und Herleitungen als Routineargumente. Wir verstehen darunter die Wiedergabe von Sätzen, Verfahren etc., die Schulbüchern der Sekundarstufe I zufolge gut eingeübt wurden, zum Zweck einer Begründung. Im Unterschied zu A0 muss also z.B. eine Rechenaufgabe der Kategorie „A1“ vordergründig argumentative Fertigkeiten einfordern und nicht bloß auf die technische bzw. algorithmische Performanz des Aufgabenlösenden ausgerichtet sein. Jordan et al. (2006) ordnen der

Kategorie „Argumentation auf niedrigem Niveau notwendig“ einschrittige oder rein rechnerische Argumente zu. Dieser Formulierung ist zu entnehmen, dass eine Rechnung nur dann in die Kategorie „Argumentation auf niedrigem Niveau notwendig“ fällt, wenn sie zur Begründung eines Sachverhalts genutzt wird. Soll die in der Aufgabe verlangte Rechnung lediglich die Beherrschung eingeübter Lösungsverfahren demonstrieren, ist keine kognitive Verarbeitungstiefe hinsichtlich der Begründung bzw. der Gültigkeit eines Sachverhalts gegeben. Argumentieren vom Niveau „A1“ ist für uns also der erste Schritt des Reflektierens bzw. Begründens eines mathematischen Sachverhalts. Bei Aufgaben der Kategorie „A1“ müssen zwischen der Rechnung und der mathematischen Situation explizit Bezüge hergestellt werden. Es muss zumindest durchdacht und begründet werden, ob und warum beispielsweise eine Rechnung in diesem (Aufgaben-)Zusammenhang zielführend ist.

In Anlehnung an das Kategoriensystem von Jordan et al. (2006) und an den Anforderungsbereich II „Zusammenhänge herstellen“ der KMK-Vorgaben (2003b) für „mathematisches Argumentieren“ bezieht sich „A2“ unseres Kategoriensystems auf die Angabe mehrerer Argumente zur Begründung, auf das Herstellen von Zusammenhängen zwischen wenigen, aber unterschiedlichen Einzelargumenten bzw. auf das Erläutern gegebener Zusammenhänge zwischen Einzelargumenten.

Unsere Kategorie „A3“ greift die Beschreibung der Kategorie „Argumentation auf hohem Niveau notwendig“ von Jordan et al. (2006) sowie den Anforderungsbereich III „Verallgemeinern und Reflektieren“ der KMK-Vorgaben (2003b) für die Kompetenz „Argumentieren“ auf, die einerseits eine (schon relativ abstrakte und beispielungebundene) Beweistätigkeit, also das Entwickeln einer komplexen Argumentation, und andererseits das Abwägen, Bewerten und Erläutern gegebener komplexer Argumentationen und Aussagen verlangen (Tab. 5).

Tab. 5 Die Kompetenz „Argumentieren“ und ihre Kompetenzniveaus

Kat.	Kategorienname	Operationalisierung
A0	Nicht notwendig	Verlangt wird, den Rechenweg oder das Ergebnis zu notieren oder Definitionen zu nennen. Auch das Ablesen von Werten aus Diagrammen, Graphen etc. bzw. das Ergänzen von Werten in Tabellen etc. gehören in diese Niveaustufe
A1	Argumentation auf niedrigem Niveau notwendig	Gefragt ist die bloße Wiedergabe von Standardargumenten (beispielsweise den Term einer linearen Funktion im Sachzusammenhang zu begründen). Argumentationen, für die Alltagswissen genügt, werden abgefragt. Erforderliche Begründungen basieren meist auf einschrittigen Rechnungen. Alternativ soll der Lösungsweg oder die Vorgehensweise beschrieben werden
A2	Argumentation auf mittlerem Niveau notwendig	Überschaubare mehrschrittige, auch begrifflich geprägte mathematische Argumente sollen entwickelt und schriftlich dargelegt werden (beispielsweise soll begründet werden, warum eine quadratische Funktion einen gegebenen Graphen beschreibt und eine andere Funktion nicht)
A3	Argumentation auf hohem Niveau notwendig	Komplexe mathematische Argumente (Begründungen, Beweise, Verfahren, Verallgemeinerungen) werden entwickelt und schriftlich darlegt oder gegebenenfalls werden solche nachvollzogen. Verschiedene Arten von mathematischen Argumentationen werden z. B. hinsichtlich der Effizienz verglichen und bewertet

4.2.4 Wissenschaftliche Güte

In der Vorbereitung der Kodierung wurden 100 zufällig ausgewählte Items von allen Autor:innen zur Einarbeitung kodiert und konsensuell validiert. Die *Objektivität* des Instrumenteneinsatzes wurde überprüft, indem die Zweitautorin weitere 200 zufällig ausgewählte Items kodiert hat; diese Kodierungen wurden mit je 100 Kodierungen der beiden anderen Autoren verglichen. Für die 100 unabhängig voneinander kodierten Items konnte eine Interraterübereinstimmung von insgesamt Cohens Kappa=0,85 ermittelt werden. Da die Übereinstimmungswerte sehr gut waren (Mislevy 2006; Wirtz und Caspar 2002), wurden die Kategoriensysteme zur weiteren Kodierung der Items verwendet. Alle Items wurden unabhängig voneinander in Bezug auf alle vier Kompetenzen auf Niveaus von 0 bis 3 kodiert.

Eine Überprüfung der *Reliabilität* wurde hier als Re-Test-Reliabilität realisiert: Die Zweitautorin hat im Rahmen einer Überarbeitung des vorliegenden Manuskripts nach mehr als einem Jahr Abstand 200 zufällig ausgewählte Aufgaben erneut kodiert, mit einer Intraraterübereinstimmung von Cohens Kappa=0,95.

Zur Einschätzung der *Validität* der für diese Studie entwickelten Kodierung wurden in anderen Studien genutzte Beispielaufgaben zu unterschiedlichen Kompetenzniveaus mit unseren Schwierigkeitseinstufungen derselben Aufgaben verglichen. Vermeintlich naheliegend wäre im Hinblick auf die Validitätsmessung zunächst eine Korrelation zwischen empirischen Aufgabenschwierigkeiten und unseren Aufgabekodierungen gewesen. Der Vergleich der Kodierungen mit nach empirischer Schwierigkeit gestuften Aufgaben (z. B. aus Blum et al. 2019) ist allerdings nur in eine Richtung sinnvoll: Mit steigendem Anspruchsniveau ist eine Abnahme der Lösungshäufigkeit und -richtigkeit zu erwarten; anders herum lassen sich allerdings sehr einfach Aufgaben ohne oder mit nur sehr geringen Bezügen zu den mathematischen Kompetenzen finden, die dennoch schwierig oder fehlerträchtig sind. Dies lässt keine inhaltlich sinnvoll interpretierbaren Korrelationen zwischen kodiertem Anspruchsniveau und empirischer Schwierigkeit erwarten. Daher haben wir uns stattdessen für einen Vergleich unserer Aufgabekodierungen mit Kompetenzabstufungen von Beispielaufgaben u. a. aus der COACTIV-Studie (Jordan et al. 2006) entschieden, der ohne den Rückgriff auf empirische Daten auskommt; konkret auf die beiden Kategorien „außermathematisches Modellieren“ und „Argumentation“, die sehr eindeutig den hier verwendeten Kategorien „Mathematisches Modellieren“ bzw. „Mathematisches Argumentieren“ zugeordnet werden können. Die Anwendung unserer Kodierschemata auf die veröffentlichten COACTIV-Aufgaben zeigt, dass (a) die ganze Bandbreite unserer Codes (Kode 0 bis Kode 4) in den von unserer Studie unabhängig angeführten Aufgaben vorkommt und dass (b) signifikante (rang-)korrelative Zusammenhänge zwischen den als ähnlich konzipiert eingeschätzten Kategorien vorherrschen (COACTIV: „außermathematisches Modellieren“ verglichen mit „Mathematisches Modellieren“ aus dieser Studie: $r_s=0,81$; COACTIV: „Argumentation“ verglichen mit „Mathematisches Argumentieren“ aus dieser Studie: $r_s=0,67$; siehe Anhang, S. XV ff.). Die Aufgabenbeispiele der COACTIV-Studie zu den Kompetenzen „innermathematisches Modellieren“ und „mathematische Darstellungen“ sind vom Vergleich mit unseren Kompetenzabstufungen ausgenommen (kein Korrelationswert berechnet), da zwar durchaus Schnittmengen zwischen den

Kompetenzen mit unserer Kompetenzdefinition von „Problemlösen“ existieren, aber nicht von einer Konstruktleichheit ausgegangen werden kann. Ein solcher Vergleich der Kompetenzabstufungen wäre bezüglich der Validitätseinschätzung unserer Kompetenzkategorien daher wenig aussagekräftig.

Deshalb haben wir zusätzlich Aufgaben ($N=6$) aus den KMK-Bildungsstandards (2003b) für das „mathematische Problemlösen“ mit unserem „Problemlösen“-Kategoriensystem kodiert; hier zeigt sich, dass wir diese Aufgaben hinsichtlich der Problemlöseanforderungsniveaus ähnlich einschätzen. Für die Kompetenz „mathematisches Problemlösen“ beträgt der Korrelationswert mit den Beispielaufgaben $r_s=0,95$.

Ergänzend haben wir auch für die anderen beiden Kompetenzen „Argumentieren“ und „Modellieren“ die Rangkorrelationen zwischen den Einschätzungen aufgrund unseres Kategoriensystems und den Einordnungen in die Anforderungsbereiche durch die KMK-Vorgaben (2003b) bestimmt. Die Korrelationen zwischen den Einstufungen sind signifikant mit $r_s=0,84$ für die Kompetenz „Argumentieren“ ($N=7$) und $r_s=0,98$ für die Kompetenz „Modellieren“ ($N=6$).

5 Ergebnisse

5.1 Gesamtübersicht Kompetenzanforderungen

Zur Beantwortung der ersten Forschungsfrage, welche Kompetenzanforderungen die ZP10-Aufgaben insgesamt charakterisieren, gibt Tab. 6 eine Übersicht. Die Anzahlen der jeweiligen Kodierungen in der Tabelle zeigen, dass ein Großteil der Items nur ein sehr geringes Kompetenzniveau besitzt. In Bezug auf die Kompetenzen „Problemlösen“ und „Argumentieren“ sind jeweils über 80 % der Items (und damit der Median) in Kategorie „0“ verortet worden. Beim „Modellieren“ wurden vergleichbar viele Items in Kategorie „1“ kodiert, was minimalen Realitätsbezügen bzw. einer Einkleidung ohne Modellierungstätigkeit entspricht – komplett innermathematisch (Kategorie „0“) sind knapp 24 % der Items. Kategorie „3“, maximales Anforderungsniveau, wurde in der gesamten Stichprobe bei keinem Item in Bezug auf keine der vier Kompetenzen kodiert.

Auffällig ist in den Daten das Ausbleiben der Maximalwerte. Hierzu stellen wir zunächst fest, dass dies ein zentrales Ergebnis unserer Untersuchung der NRW ZP10 darstellt, da es durchaus Aufgaben mit höheren Anforderungsniveaus aus vergleichbaren Kontexten gibt (s. auch Anhang); Abb. 2 zeigt ein Beispiel für die Kategorie „P3“ (*Problemlösen*) aus einer Abschlussprüfung in Bayern.

Tab. 6 Übersicht über die Kodierungen aller ZP10-Aufgaben NRW (2007–2019)

	Kompetenzen	N	Anz. Kat. „0“	Anz. Kat. „1“	Anz. Kat. „2“	Anz. Kat. „3“
Alle Ab- schlüsse	<i>Problemlösen</i>	1886	1541 (81,7%)	341 (18,1%)	4 (0,2%)	0 (0,0%)
	<i>Modellieren</i>	1886	451 (23,9%)	1433 (76,0%)	2 (0,1%)	0 (0,0%)
	<i>Argumentieren</i>	1886	1518 (80,5%)	229 (12,1%)	139 (7,4%)	0 (0,0%)

Beispielitem (ZP10 in Bayern, 2018, Mathematik II, Nachtermin A. 2.4)	Kodierung
<p>Die Zeichnung zeigt das Trapez ABCD. Der Punkt F liegt auf der Strecke [AB], der Punkt E liegt auf der Strecke [CD] und die Diagonale [AC] schneidet die Strecke [EF] im Punkt M. Ein Kreis um M berührt die Strecke [CE] im Punkt S und schneidet die Strecke [MC] im Punkt G sowie die Strecke [ME] im Punkt H.</p> <p>Es gilt: $[AB] = 7 \text{ cm}$; $BC = 6 \text{ cm}$; $CD = 10 \text{ cm}$; $\sphericalangle CBA = 90^\circ$; $\sphericalangle DCB = 90^\circ$; $[AF] = [DE] = 3 \text{ cm}$.</p> <p>Berechnen Sie die Länge b des Kreisbogens GH.</p>	<p>P3 M0 A0</p>

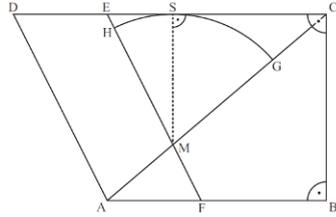


Abb. 2 Beispielitem

Zur weiteren Einordnung: Eine Analyse der ZP10-Aufgaben 2008 (Neubrand und Neubrand 2010) zeigt, dass keine der Aufgaben aus den Prüfungsvarianten „MSA“ (S. 88) und „HSA“ (S. 86) dem von den KMK-Vorgaben (2003b, 2004) definierten Anforderungsbereich III zugeordnet wurde. Mit Blick auf die COACTIV-Tätigkeiten hat Drücke-Noe (2014) die hessischen ZP-Testaufgaben (2007–2009) bewertet und festgestellt, dass die geforderten Tätigkeiten nur punktuell über Standardaktivitäten, also Reproduktion hinausgehen. Zu einem ähnlichen Schluss kommen auch Kühn und Drücke-Noe (2013), nämlich dass die Testaufgaben zentraler Abschlussprüfungen der Bundesländer (2007–2011) in erster Linie einfache Routinetätigkeiten erfordern. Diese Befunde bestätigen unsere Ergebnislage und unterstreichen die Validität unserer Kategoriensysteme.

5.2 Zeitliche Unterschiede der drei Kompetenzen

Forschungsfrage (2) adressiert Unterschiede bezüglich der Kompetenzanforderungen im untersuchten Zeitraum. Mediane sind hierfür nicht aussagekräftig, da sie ausnahmslos – in allen Jahren und für alle Schulformen – beim „Problemlösen“ und „Argumentieren“ Kategorie „0“ und beim „Modellieren“ Kategorie „1“ annehmen. Stattdessen wird der prozentuale Anteil der Kategorie „0“ für die vier mathematischen Kompetenzen pro Jahr betrachtet. Kategorie „0“ ist insofern interessant, als dass in Bezug auf die Kompetenzen keinerlei Anforderungen gestellt werden – beispielsweise können Items ausschließlich mit Routineverfahren gelöst werden, es werden keine Begründungen irgendwelcher Art eingefordert oder es gibt keinen Realitätsbezug. Die restlichen Werte entfallen dann jeweils fast alle auf Kategorie „1“ (s. Tab. 6). Abb. 3 zeigt den zeitlichen Verlauf der prozentualen Häufigkeit der Kategorie „0“ in Bezug auf die vier Kompetenzen nach Jahren.

Es zeigt sich, dass im Hinblick auf das „Problemlösen“ mit Ausnahme von 2009 ein fast konstanter Anteil (80–90%) der Items ausschließlich unterrichtsnahe Reproduktionen abtestet. „Modellieren“ wird zu Beginn des betrachteten Zeitraums in knapp unter 20% der Items nicht verlangt, d. h. nur wenige Aufgaben sind rein innermathematisch formuliert. Diesbezüglich ist zudem kein eindeutiger Trend erkennbar: In den Jahren seit 2014 ist der Anteil an Items komplett ohne Realitätsbezüge allerdings auf gut 40% angestiegen. Modellieren jenseits von Kategorie „1“ („reine

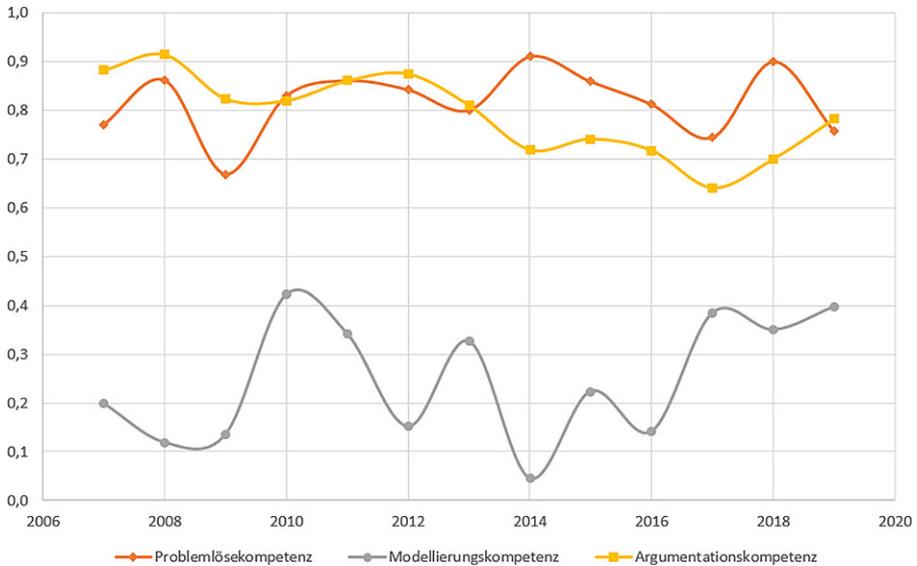


Abb. 3 Zeitliche Entwicklung der Kategorie „0“

Einkleidung“) kommt mit Ausnahme von zwei Items gar nicht vor. Das „Argumentieren“ wurde bei Einführung der ZP10 in knapp 90 % der Items nicht eingefordert. Dieser Anteil sinkt über die Jahre sukzessiv, so dass schließlich 2017 im Schnitt etwa in 4 von 10 Aufgaben Begründungen verlangt werden. In den beiden Folgejahren steigt der Anteil an Items, die kein Argumentieren erfordern, wieder, wobei hier im Vergleich zu 2007–2013 dennoch häufiger Begründungen eingefordert werden.⁵

5.3 Vergleich der Bildungsgänge

In den Vergleich der Bildungsgänge geht nur ein Teil der Items ein (s. Abschn. 4.1). „Die Prüfungsaufgaben orientieren sich jeweils an den entsprechenden angestrebten Abschlüssen“ (NRW o.J.). Da die gymnasialen Items hinsichtlich der Stichprobengröße und des betrachteten Aufgabenzeitraums nicht mit den Prüfungsvarianten HSA und MSA vergleichbar sind, adressiert Forschungsfrage (3) den Vergleich der Kompetenzniveaus von Hauptschulabschluss und der Fachoberschulreife. Die Kodierungen der gymnasialen Items sind für einen informellen Vergleich auch in Tab. 7 aufgeführt, werden aber nicht auf signifikante Unterschiede zu den Kodierungen der Prüfungsvarianten MSA und HSA geprüft.

Mann-Whitney-U-Tests zeigen (auch nach einer Bonferroni-Korrektur: $p = 0,05/4 = 0,0125$) signifikante Unterschiede zwischen den beiden Abschlüssen auf: „Problemlösen“ ($U = 56.296,5$; $Z = 2,42$; $p_{(einseitig)} = 0,008$), „Modellieren“ ($U = 69.170$; $Z = -2,3$; $p_{(einseitig)} = 0,011$), „Argumentieren“ ($U = 56.082,5$; $Z = 2,5$; $p_{(1)} = 0,006$). Das

⁵ Gerechnet wurde mit allen 1886 Aufgaben; nimmt man nur die in Abschn. 5.3 betrachteten 812 Aufgaben, ergeben sich nahezu identische Werte bzw. Verläufe der vier Graphen.

Tab. 7 Verteilung der Items innerhalb der drei Bildungsgänge

	Kompetenzen	<i>N</i>	Anz. Kat. „0“	Anz. Kat. „1“	Anz. Kat. „2“	Anz. Kat. „3“
<i>HSA</i>	<i>Problemlösen</i>	340	303 (89,1 %)	37 (10,9 %)	0 (0,0 %)	0 (0,0 %)
	<i>Modellieren</i>	340	60 (17,6 %)	280 (82,4 %)	0 (0,0 %)	0 (0,0 %)
	<i>Argumentieren</i>	340	293 (86,2 %)	35 (10,3 %)	12 (3,5 %)	0 (0,0 %)
<i>FOR</i>	<i>Problemlösen</i>	370	291 (78,6 %)	78 (21,1 %)	1 (0,3 %)	0 (0,0 %)
	<i>Modellieren</i>	370	103 (27,8 %)	266 (71,9 %)	1 (0,3 %)	0 (0,0 %)
	<i>Argumentieren</i>	370	281 (75,9 %)	50 (13,5 %)	39 (10,5 %)	0 (0,0 %)
<i>GYM</i>	<i>Problemlösen</i>	102	71 (69,6 %)	30 (29,4 %)	1 (1,0 %)	0 (0,0 %)
	<i>Modellieren</i>	102	24 (23,5 %)	78 (76,5 %)	0 (0,0 %)	0 (0,0 %)
	<i>Argumentieren</i>	102	88 (86,3 %)	10 (9,8 %)	4 (3,9 %)	0 (0,0 %)

Verteilung der Items innerhalb der drei Bildungsgänge auf die Kategorien „0“, „1“, „2“ und „3“ im Zeitraum zwischen 2007–2019 (Hauptschulabschluss (*HSA*) und Fachoberschulreife (*FOR*)) bzw. 2007–2010 (Gymnasium (*GYM*))

bedeutet, dass die Prüfungsvariante zur Fachoberschulreife im Vergleich zum Hauptschulabschluss signifikant höhere Anforderungen auf dem Niveau „1“ hinsichtlich der Kompetenzen „Problemlösen“ und „Argumentieren“ stellt und weniger eingekleidete Aufgaben (Items mit „Modellierungscharakter“ auf dem Anforderungsniveau „1“) enthält.

6 Diskussion

Der vorliegende Beitrag stellt die Ergebnisse einer Analyse der NRW ZP10-Aufgaben des Zeitraums 2007–2019 aller dort abschlussverleihenden Schulformen (Hauptschule, Realschule, Abendrealschule, Sekundarschule, Gesamtschule und Gymnasium) vor. Der Fokus dieser Analyse waren die drei prozessbezogenen Kompetenzen „Problemlösen“, „Modellieren“ und „Argumentieren“ (Abschn. 2.1).⁶ Hierbei wurde – auf der Grundlage von vierstufig (0–3) operationalisierten Kategorien – erfasst, welche Kompetenzniveaus die ZP10-Aufgaben jeweils aufweisen (Abschn. 4.2). Unter der Annahme, dass Zentrale Prüfungen am Ende der Sekundarstufe I eine gewisse Steuerungs- und Orientierungsfunktion hinsichtlich der Unterrichtsinhalte des Mathematikunterrichts haben, sollte unsere Studie u. a. den Stellenwert der genannten Kompetenzen in der ZP10 ermitteln.

Die rationale Aufgabenanalyse zeigt, dass diesen drei Kompetenzen gemessen an unseren fachdidaktisch theoretisch fundierten Kategoriensystemen eine zumeist sehr geringe Bedeutung zukommt, da in den Aufgaben höhere Kompetenzniveaus kaum bis gar nicht eingefordert werden. Konkret bedeutet dies für das „Argumentieren“ und das „Problemlösen“, dass hier im Schnitt mehr als 8 von 10 Testaufgaben nicht

⁶ Dass bestimmte Kompetenzniveaus und auch bestimmte Kompetenzen (siehe Einleitung) nicht abgefragt werden, liegt sicherlich auch daran, dass sie im vorliegenden Prüfungsformat (begrenzte Zeit, keine Kollaborationen etc.) nur schlecht oder teilweise auch gar nicht abgefragt werden können. Die Ergebnisse unserer Analyse sollen daher nicht so verstanden werden, dass wir die Arbeit der Autorinnen und Autoren der ZP10-Aufgaben kritisieren wollen. Wir zeigen lediglich auf, was mit den derzeitigen Prüfungsaufgaben (im genutzten Format) NICHT abgetestet wird.

über das Kompetenzniveau „0“ hinausgehen. Damit fordern die Aufgabenlösungen keine Begründungen und kein heuristisches Arbeiten ein. Tätigkeiten mittleren Kompetenzniveaus („2“) mit mehrschrittigen Modellierungen und Argumentationen sowie einem „Transfer“ beim Problemlösen tragen allenfalls punktuell zum Kompetenzaufbau bei; hohes Kompetenzniveau („3“) wird kategorienübergreifend in keinem der analysierten Items verlangt. Dabei kann eine Entwicklung dieses Aufgabenprofils allenfalls in Bezug auf das Argumentieren festgestellt werden, das seit 2014 tendenziell in höherem Maße eingefordert wird als im vorangegangenen Zeitraum. Damit sind die Ergebnisse bezüglich eines derart aufgefassten kognitiven Aufgabenanspruchs des ZP10-Aufgabensatzes anschlussfähig an die wenigen bislang vorliegenden Forschungsergebnisse zu innerdeutschen MSA-Prüfungen (Drüke-Noe 2014; Kühn und Drüke-Noe 2013; Scheja 2019).

Aus theoretischer Sicht verankern die ZP10 im Ergebnis kaum Anforderungen hinsichtlich der drei Kompetenzen, die als anspruchsvoll bezeichnet werden können. Diskutieren ließe sich, ob die Kategorien unseres Kategoriensystems bezüglich geringerer Kompetenzanforderungen der Stufe 0 bis 1, die in den ZP10 vorliegen mögen, unsensibel sind bzw. zu wenig differenzieren. Diesem Einwand kann Folgendes entgegenhalten werden: a) Der Fokus der Studie ist in erster Linie die fachdidaktisch, theoretisch fundierte Analyse der ZP10-Aufgaben bezüglich der Frage, welche Kompetenzniveaus dort zu welchem Anteil abgefragt werden. Und b) eine weitere Differenzierung z. B. des Anforderungsniveaus 0 der drei prozessbezogenen Kompetenzen würde zwar mehr Aufschluss über die Unterschiede der ZP10-Aufgaben auf diesem Niveau bringen, damit würden aber die ZP10-Aufgaben dennoch kein höheres Kompetenzniveau abverlangen. Der Fokus der Studie ginge in Fall b) auch verloren, der nicht untersucht, welche Unterschiede die ZP10-Aufgaben (auf geringem Kompetenzniveau) im Einzelnen überhaupt charakterisieren. Eine solche detaillierte Analyse fokussiert nicht mehr die kompetenzbezogenen Niveauunterschiede der ZP10-Aufgaben, sondern andere kriteriale Aufgabenunterschiede.

Es ließe sich fragen, ob unsere Kategorien 0 zu wenig differenzieren. Dies betrachten wir an folgendem Beispiel (Tab. 8). Es zeigen sich Unterschiede in den Anforderungen, sodass es tatsächlich möglich wäre, die Aufgaben noch weiter zu differenzieren. Ein genauerer Blick zeigt aber (s. Tab. 8), dass diese Abstufung nichts daran ändert, dass keine der beiden Aufgaben (bzw. keine der in Kategorie 0 verorteten Aufgaben) die (aus der Literatur abgeleiteten) Kriterien für Kategorie 1 erfüllt bzw. ein entsprechendes Kompetenzniveau einfordert. Eine feinere Abstufung ist also möglich, dient aber nicht dazu, unsere Forschungsfrage zu beantworten.

Wie in den Abschn. 4.2.1–4.2.3 dargestellt, orientieren sich unsere Kategorieabstufungen an anderen weitläufig rezipierten Anforderungsniveaus aus Fachdidaktik (z. B. Jordan et al. 2006) und Bildungspolitik (KMK 2003b). Unsere Kategorien sind daher aus theoretischer Sicht fundiert und durch unsere Einschätzungsvergleiche (4.3) bezüglich ihres Anforderungsniveaus angemessen belegt. Unser Kategoriensystem wurde nicht unter Berücksichtigung von empirischen Aspekten, wie z. B. durch Auswertung von Schülerlösungen sowie der Lösungshäufigkeiten, erstellt und wir beanspruchen deshalb auch keine abschließenden Aussagen über die Aufgabenschwierigkeit der ZP10-Aufgaben.

Tab. 8 Beispielaufgaben aus der ZP10 2019, Abendrealschule, und 2017, HSA, die mit „A0 – Argumentation nicht notwendig“ kodiert wurden

Kat.	Beispiele	Kommentare
A0	Aufgabe 1 Ordnen Sie die Zahlen der Größe nach. Beginnen Sie mit der kleinsten Zahl. $\frac{6}{10}$, $-0,626$, $-6,26$, $\frac{1}{6}$ (ZP10, 2019, Abendrealschule, Prüfungsteil I, Aufgabe 1)	Diese beiden Aufgaben werden der Kategorie „A0 – Argumentation nicht notwendig“ zugeordnet, da sie nicht primär, das mathematische Argumentieren abfragen: – Aufgabe 1 verlangt das Ordnen der angegebenen Zahlen durch Kleiner-Größer-Relationen ohne Begründung.
A0	Aufgabe 1. a) Eine Kaugummikugel hat einen Durchmesser von 14 mm. Bestätigen Sie durch eine Rechnung, dass das Volumen einer Kaugummikugel ca. $1,44 \text{ cm}^3$ beträgt. (ZP10, 2019, Abendrealschule, Prüfungsteil II, Aufgabe 1.a)	– Aufgabe 1. a) erfordert das rechnerische Bestätigen eines vorgegebenen Ergebnisses durch einen ange deuteten bzw. vorgegebenen Rechenweg

Da die ZP10-Prüfungen einen großen Einfluss auf die Abschlussnote im Fach Mathematik nach der 10. Klasse haben, gibt es für den Unterricht, der Lernende auf diese Prüfungen vorbereitet, dem vorliegenden theoretisch-gestützten Befund vermutlich kaum Anreize, mathematische Kompetenzen im Sinne des Modells von Niss und Højgaard (2019) zu fördern. Aufgaben, die sich am Kompetenzprofil der ZP10 orientieren, sind für Lehrkräfte leichter fokussierbar, können im Unterricht gut geübt werden und sind einfach prüfbar (Büchter und Pallack 2012). Im Sinne einer Vermittlung der Mathematik, wie sie u. a. im Modell von Niss und Højgaard (2019), der Bildungsstandards und der NRW-Kernlehrpläne konzeptualisiert wird, wäre es somit wünschenswert, vermehrt Prüfungsaufgaben zu berücksichtigen, die die drei Kompetenzen einfordern, ohne dabei ausgewählte Kompetenzniveaus im aufgezeigten niedrigen Maße zu fokussieren.

Andererseits zeigt sich aus der Durchsicht empirischer Befunde zu den durchschnittlichen Ergebnissen der ZP10, dass diese im hier betrachteten Zeitraum in Mathematik ausnahmslos im befriedigenden oder ausreichendem Notenbereich lagen.⁷ Damit werden die Aufgaben der ZP10 von den Prüflingen offenbar als schwierig wahrgenommen und eine weitere Erhöhung der kompetenzbezogenen Anforderungen würde vermutlich dazu führen, dass sich der Anteil der Lernenden, die *keinen* Abschluss erreicht, erhöhen würde, was möglichst zu vermeiden ist.⁸ Hier scheint eine gleichzeitige Anbahnung der Anhebung des Kompetenzniveaus aus der Wirkungsrichtung des Mathematikunterrichts tragfähiger (Mathematikunterricht → ZP10), indem man beispielsweise bei den inhaltlichen Vorgaben (z. B. die Spannweite) ansetzt, deren unterrichtliche Umsetzung bisher vorwiegend Aktivitäten und Prozesse mit geringen Anspruchsniveaus fördert (Jordan et al. 2006) und im Hinblick darauf zudem entwicklungsarm zu sein scheint (s. Henschel et al. 2019). Eine Reduktion der inhaltlichen Breite im Sinne von „konzeptionelle Tiefe vor inhaltlicher

⁷ u. a. www.standardsicherung.schulministerium.nrw.de/cms/zentrale-pruefungen-10/ergebnisrueckmeldung/ergebnisberichte/.

⁸ s. <https://de.statista.com/statistik/daten/studie/440740/umfrage/anteil-der-schulabgaenger-innen-ohne-hauptschulabschluss-in-deutschland/>.

Breite“ würde zeitliche Ressourcen für mehr gehaltvolle Frage- und Problemkontexte bereitstellen und damit vermehrt komplexere Prozesse und Tätigkeiten in den Lernfokus stellen.

Die Annahme dieser Studie, dass die ZP10-Aufgaben der untersuchten Bildungsgänge der Sekundarstufe I auch mit einer Abstufung der Anforderungen innerhalb der drei Kompetenzen einhergehen, konnte zwischen den Bildungsgängen mit HSA-Prüfungen und denen mit RS-Prüfungen identifiziert werden. Dies deckt sich mit den Ergebnissen von Neubrand und Neubrand (2010) für das Jahr 2008 und der Vorgabe, dass innerhalb der drei Bildungsgänge der Sekundarstufe I zwei Abschlussarten verliehen werden (Hauptschulabschluss Klasse 10 und die Fachoberschulreife), lässt jedoch zugleich die Vorgaben der KLPe der Bildungsgänge konzeptionell zum Teil außen vor. Vergleicht man die prozentualen Häufigkeiten der Kompetenzniveaus (Tab. 8) in den Prüfungsvarianten, wird deutlich, dass im Bildungsgang Gymnasium nur im Hinblick auf die Kompetenz „Problemlösen“ höhere Anforderungen gestellt werden. Dies überrascht, da gerade das Gymnasium auf die Oberstufe vorbereiten soll und hier neben einer breiten inhaltlichen Basis eine Wissenschaftspropädeutik im Sinne einer „Vertrautheit mit spezifischen mathematischen Arbeitsprozessen und den prozessbezogenen Kompetenzen“ (Neubrand und Neubrand 2010, S. 112) erwartet werden kann. Dieser Befund ist allerdings nur eingeschränkt referenzierbar, da die untersuchten Testaufgaben des Gymnasiums einen deutlich kürzeren Zeitraum abdecken (2007–2010) als die der übrigen Schulformen. Interessant ist, wie sich die aktuell vollziehenden Reformen des Gymnasiums (Verlängerung der Schulzeit von 8 auf 9 Schuljahre (MSW 2019a); Einführung eines neuen KLPs; Wiedereinführung der ZP10 im Jahr 2024) auf die Aufgabekultur auswirken werden. So zeigt eine erste Analyse des neuen KLPs, dass das Kompetenzspektrum in Anlehnung an die Bildungsstandards breiter und differenzierter dargestellt wird, womit eine Einflussnahme auf die Konzeption der Testaufgaben (GYM) erwartet werden kann.

Als Limitation der vorliegenden Studie ist einerseits zu nennen, dass lediglich die Aufgaben, aber nicht die Produkte oder sogar die Bearbeitungsprozesse von Schülerinnen und Schülern ausgewertet wurden. Wir können also nicht mit letzter Sicherheit sagen, dass prozessbezogene Kompetenzen bei der Bearbeitung der ZP10-Aufgaben kaum eine Rolle spielen. Aber unsere Analysen zeigen, dass alleine die Aufgabenkonzeption den in der Fachdidaktik breit konzeptualisierten Kompetenzen keinen großen Stellenwert beimisst. Zudem beruhen die erwarteten Effekte auf die Gestaltung des prüfungsvorbereitenden Unterrichts aufseiten der Lehrkräfte (s. 2.2) auch hauptsächlich auf kursierenden Aufgabensammlungen und weniger auf ausgewerteten Prozessen von Schülerinnen und Schülern.

Unsere Studie erlaubt auch keine Aussagen darüber, welche sonstigen (sicherlich auch positiven) Effekte die Installation und Durchführung der ZP10 haben und welche Fertigkeiten und Kompetenzen eingefordert und angeregt werden.

Schließlich bleiben unsere Ergebnisse auf die ZP10 und NRW beschränkt, Aussagen über andere (zentrale) Prüfungen und andere Bundesländer sind nicht möglich. Die exemplarische Betrachtung einer ZP10-Aufgabe aus Bayern (s. Tab. A2 im Anhang) lässt vermuten, dass eine vergleichbare Analyse in anderen Bundesländern jedoch anders ausfallen könnte. Immerhin haben wir mit NRW das bevölkerungs-

stärkste Bundesland ausgewählt und können damit Aussagen über einen guten Teil der deutschen Schullandschaft treffen.

Zusatzmaterial online Zusätzliche Informationen sind in der Online-Version dieses Artikels (<https://doi.org/10.1007/s11618-023-01191-7>) enthalten.

Funding Open Access funding enabled and organized by Projekt DEAL.

Interessenkonflikt B. Scheja, A. Schreck und B. Rott geben keinerlei Interessenkonflikte an, die für den Inhalt dieses Artikels von Bedeutung sind.

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

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

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

Literatur

- Alderson, J. C., & Wall, D. (1993). Does Washback exist? *Applied Linguistics*, 14(20), 115–129.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: designing and developing useful language tests*. Oxford: Oxford University Press.
- Baumert, J., Lehmann, R., Lehrke, M., Schmitz, B., Clausen, M., Hosenfeld, I., Köller, O., & Neubrand, J. (1997). *TIMSS – Mathematisch-naturwissenschaftlicher Unterricht in internationalen Vergleich*. Opladen: Leske + Budrich.
- Blum, W., & Leiß, D. (2005). “Filling Up”—the Problem of Independence-preserving Teacher Interventions in Lessons with Demanding Modelling Tasks. In *CERME 4—Proceedings of the Fourth Congress of the European Society for Research in Mathematics Education* (S. 1623–1633). <http://erme.site/cerme-conferences/cerme-4/>. Zugegriffen: 10. Aug. 2023.
- Blum, W., Drüke-Noe, C., Leiß, D., Wiegand, B., & Jordan, A. (2005). Zur Rolle von Bildungsstandards für Qualitätsentwicklung im Mathematikunterricht. *ZDM – Mathematics Education*, 37(4), 267–274.
- Blum, W., Drüke-Noe, C., Hartung, R. & Köller, O. (2010). *Bildungsstandards Mathematik: konkret. Sekundarstufe I. Aufgabenbeispiele, Unterrichts Anregungen, Fortbildungsideen*. Berlin: Cornelsen.
- Blum, W., Roppelt, A., & Müller, M. (2019). Kompetenzstufenmodelle für das Fach Mathematik. In P. Stanat, S. Schipolowski, N. Mahler, S. Weirich & S. Henschel (Hrsg.), *IQB-Bildungstrend 2018. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I im zweiten Ländervergleich* (S. 60–71). Münster: Waxmann.
- Büchter, A., & Pallack, A. (2012). Methodische Überlegungen und empirische Analysen zur impliziten Standardsetzung durch zentrale Prüfungen. *Journal für Mathematikdidaktik*, 33, 59–85.
- Budke, A., Kuckuck, M., Meyer, M., Schäbitz, F., Schlüter, K., & Weiss, G. (2015). *Fachlich argumentieren lernen. Didaktische Forschungen zur Argumentation in den Unterrichtsfächern*. Münster: Waxmann.
- Cheng, L. (2005). *Changing language teaching through language testing. A washback study*. Cambridge: Cambridge University Press.
- Cuoco, A., Goldenberg, E. P., & Mark, J. (1996). Habits of mind: an organizing principle for mathematics curricula. *Journal of Mathematical Behaviour*, 15, 375–402.
- Dörner, D. (1987). *Problemlösen als Informationsverarbeitung* (3. Aufl.). Mainz: Kohlhammer.

- Drücke-Noe, C. (2014). *Aufgabenkultur in Klassenarbeiten im Fach Mathematik. Empirische Untersuchungen in neunten und zehnten Klassen*. Wiesbaden: Springer Spektrum.
- Greefrath, G., Kaiser, G., Blum, W., & Borromeo, F.R. (2013). Mathematisches Modellieren – Eine Einführung in theoretische und didaktische Hintergründe. In F.R. Borromeo, G. Greefrath & G. Kaiser (Hrsg.), *Mathematisches Modellieren für Schule und Hochschule. Theoretische und didaktische Hintergründe* (S. 11–31). Wiesbaden: Springer Spektrum.
- Hartig, J., & Klieme, E. (2006). Kompetenz und Kompetenzdiagnostik. In K. Schweizer (Hrsg.), *Leistung und Leistungsdiagnostik* (S. 127–143). Heidelberg: Springer.
- Henschel, S., Rjosk, C., Holtmann, M., & Stanat, P. (2019). Merkmale der Unterrichtsqualität im Fach Mathematik. In P. Stanat, S. Schipolowski, N. Mahler, S. Weirich & S. Henschel (Hrsg.), *IQB-Bildungstrend 2018. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I im zweiten Ländervergleich* (S. 355–383). Münster: Waxmann.
- Jordan, A., Ross, N., Krauss, S., Baumert, J., Blum, W., Neubrand, M., Löwen, K., Brunner, M., & Kunter, M. (2006). *Klassifikationsschema für Mathematikaufgaben: Dokumentation der Aufgabenkategorisierung im COACTIV-Projekt*. Berlin: Max-Planck-Institut für Bildungsforschung.
- Kaiser, G., & Busse, A. (2014). Diskussion II. *Mitteilungen der DMV*, 22(2), 121–122.
- Kaiser, G., & Sriraman, B. (2006). A global survey of international perspectives on modelling in mathematics education. *ZDM – Mathematics Education*, 38(3), 302–310.
- Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruber, H., Prenzel, M., Reiss, K., Riquarts, K., Rost, J., Tenorth, H.-E., & Vollmer, H. J. (Hrsg.). (2003). *Zur Entwicklung nationaler Bildungsstandards. Eine Expertise*. Bonn: BMBF.
- KMK (2003a). *Bildungsstandards im Fach Mathematik für den Mittleren Schulabschluss. Entwurf vom 4.7.2003*. Herausgegeben vom Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. <https://www.kmk.org/fileadmin/Dateien/pdf/PresseUndAktuelles/2003/Mathematik04072003.pdf>
- KMK (2003b). *Bildungsstandards im Fach Mathematik für den Mittleren Schulabschluss*. München: Wolters Kluwer. https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2003/2003_12_04-Bildungsstandards-Mathe-Mittleren-SA.pdf
- KMK (2004). *Bildungsstandards im Fach Mathematik für den Hauptschulabschluss*. München: Wolters Kluwer.
- KMK (2022). *Bildungsstandards im Fach Mathematik Erster Schulabschluss und Mittlerer Schulabschluss*. https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2022/2022_06_23-Bista-ESA-MSA-Mathe.pdf. Zugegriffen: 31. Mai 2023.
- Konarzewski, K. (2008). *Przygotowanie uczniów do egzaminu: pokusa łatwego zysku. Raport badawczy*. Warszawa: ISP.
- Kühn, S.M. (2010). *Steuerung und Innovation durch Abschlussprüfungen?* Wiesbaden: VS.
- Kühn, S.M., & Drücke-Noe, C. (2013). Qualität und Vergleichbarkeit durch Bildungsstandards und zentrale Prüfungen? – Ein bundesweiter Vergleich von Prüfungsanforderungen im Fach Mathematik zum Erwerb des Mittleren Schulabschlusses. *Zeitschrift für Pädagogik*, 6, 912–932.
- Leuders, T. (2014). Modellierungen mathematischer Kompetenzen – Kriterien für eine Validitätsprüfung aus fachdidaktischer Sicht. *Journal für Mathematikdidaktik*, 35(1), 7–49.
- Maaß, K. (2007). *Mathematisches Modellieren. Aufgaben für die Sekundarstufe I*. Berlin: Cornelsen.
- Madaus, G., Russel, M., & Higgins, J. (2009). *The paradoxes of high-stakes testing. How they affect students, their parents, teachers, principals, schools, and society*. Charlotte.
- Maier, U. (2010). Effekte testbasierter Rechenschaftslegung auf Schule und Unterricht. Ist die internationale Befundlage auf Vergleichsarbeiten im deutschsprachigen Raum übertragbar? *Zeitschrift für Pädagogik*, 56(1), 112–128.
- Maier, U., Kleinknecht, M., Metz, K., & Bohl, T. (2010). Ein allgemeindidaktisches Kategoriensystem zur Analyse des kognitiven Potenzials von Aufgaben. *Beiträge zur Lehrerinnen- und Lehrerbildung*, 28(1), 84–96.
- Mason, J., Burton, L., & Stacey, K. (2010). *Thinking mathematically* (2. Aufl.). London: Pearson.
- Mayring, P. (2010). *Qualitative Inhaltsanalyse. Grundlagen und Techniken* (11. Aufl.). Weinheim: Beltz.
- Mislevy, R.J. (2006). Cognitive psychology and educational assessment. In R. L. Brennan (Hrsg.), *Educational measurement* (4. Aufl. S. 257–306). Greenwood: Phoenix.
- MSW (2004a). *Kernlehrplan für die Realschule in Nordrhein-Westfalen* (Mathematik, Heft 3302). Frechen: Ritterbach.
- MSW (2004b). *Kernlehrplan für die Gesamtschule – Sekundarstufe I in Nordrhein-Westfalen. Mathematik*. (Heft 3601). Frechen: Ritterbach.

- MSW (2006). Durchführung der zentralen Prüfungen am Ende der Klasse 10 im Schuljahr 2006/2007 RdErl. des Ministeriums für Schule und Weiterbildung vom 21. April 2006. <http://www.vbe-nrw.de/downloads/PDF%20Dokumente/ZentrPrufgK110.pdf>. Zugegriffen: 31. Mai 2023.
- MSW (2011). *Kernlehrplan für die Hauptschule in Nordrhein-Westfalen* (Heft 3303). Frechen: Ritterbach.
- MSW (2019a). *Kernlehrplan für die Sekundarstufe I Gymnasium in Nordrhein-Westfalen* (Heft 3401). Frechen: Ritterbach Verlag.
- MSW (2019b). Verordnung über die Ausbildung und die Abschlussprüfungen in der Sekundarstufe I (Ausbildungs- und Prüfungsordnung Sekundarstufe I – APO-SI). <https://bass.schul-welt.de/12691.htm#13-21nr1.1p41>. Zugegriffen: 31. Mai 2023.
- Neubrand, J. (2002). *Eine Klassifikation mathematischer Aufgaben zur Analyse von Unterrichtssituationen: Selbsttätiges Arbeiten in Schülerarbeitsphasen in den Stunden der TIMSS-Video-Studie*. Hildesheim: Franzbecker.
- Neubrand, M. (2015). Bildungstheoretische Grundlagen des Mathematikunterrichts. In R. Bruder, L. Hefendehl-Hebeker, B. Schmidt-Thieme & H.-G. Weigand (Hrsg.), *Handbuch der Mathematikdidaktik*. Berlin: Springer Spektrum.
- Neubrand, J., & Neubrand, M. (2010). *Mathematikdidaktische Analysen der zentralen Prüfungen 2008 in Mathematik am Ende der Klasse 10 in Nordrhein-Westfalen. Analysen von Aufgabenstellungen und Aufgabenbearbeitungen, Hinweise zur Aufgabenkonstruktion und zur Fachunterrichtsentwicklung* (Unveröffentlichter Bericht). Vechta.
- Neubrand, M., Klieme, E., Lüdtke, O., & Neubrand, J. (2002). Kompetenzstufen und Schwierigkeitsmodelle für den PISA Test zur mathematischen Grundbildung. *Unterrichtswissenschaft*, 30(2), 100–119.
- Niss, M. A. (2003). Mathematical competencies and the learning of mathematics: the Danish KOM project. In A. Gagatsis & S. Papastavridis (Hrsg.), *3rd Mediterranean Conference on Mathematical Education* (Athens, 3.-5. January 2003, S. 116–124). Hellenic Mathematical Society.
- Niss, M., & Højgaard, T. (2011). *Competencies and Mathematical Learning—Ideas and inspiration for the development of mathematics teaching and learning in Denmark* (Bd. 485, English translation of part I–VI of Niss and Jensen (2002)). Roskilde: MFUFA, Roskilde University.
- Niss, M., & Højgaard, T. (2019). Mathematical competencies revisited. *Educational Studies in Mathematics*, 102, 9–28.
- NRW Zentrale Prüfungen 10: Fragen und Antworten. <https://www.standardsicherung.schulministerium.nrw.de/cms/zentrale-pruefungen-10/fragen-und-antworten/>Zugegriffen: 31. Mai 2023.
- Peek, R. (2007). Wie aussagekräftig sind zentrale Tests? Über den Umgang mit Individualergebnissen aus Schulleistungstudien, Lernstandserhebungen und zentralen Prüfungen. *Forum Schule, Magazin für Lehrerinnen und Lehrer*, 18, 8–11.
- Prodromou, L. (1995). The backwash effect: from testing to teaching. *ELT Journal*, 49(1), 13–25.
- RAND Mathematics Study Panel (2003). *Mathematical proficiency for all students: toward a strategic research and development program in mathematics education*. Santa Monica: RAND Corporation.
- Reiss, K., Heinze, A., & Pekrun, R. (2008). Mathematische Kompetenz und ihre Entwicklung in der Grundschule. In M. Prenzel, I. Gogolin & H. Krüger (Hrsg.), *Kompetenzdiagnostik*. Wiesbaden: VS.
- Resnick, L. B., & Ford, W. W. (1984). *The psychology of mathematics for instruction*. Hillsdale: Lawrence Erlbaum.
- Richter, D., Böhme, K., Becker, M., Pant, H. A., & Stanat, P. (2014). Überzeugungen von Lehrkräften zu den Funktionen von Vergleichsarbeiten: Zusammenhänge zu Veränderungen im Unterricht und den Kompetenzen von Schülerinnen und Schülern. *Zeitschrift für Pädagogik*, 60(2), 225–244.
- Rott, B. (2013). *Mathematisches Problemlösen. Ergebnisse einer empirischen Studie*. Münster: WTM.
- Scheja, B. (2019). *Analysen zum kognitiven Anspruch von Mathematikaufgaben – Befunde aus zentralen Prüfungen und Lehrerfortbildungen* (Dissertation). Universität zu Köln.
- Schoenfeld, A. H. (1992). Learning to think mathematically: problem solving, metacognition, and sense-making in mathematics. In D. A. Grouws (Hrsg.), *Handbook for research on mathematics teaching and learning* (S. 334–370). New York: MacMillan.
- Schwarzkopf, R. (2000). *Argumentationsprozesse im Mathematikunterricht. Theoretische Grundlagen und Fallstudien*. Hildesheim: Franzbecker.
- Stacey, W. & Turner, R. (2015). *Assessing mathematical literacy*. Cham: Springer.
- Turner, R., Blum, W., & Niss, M. (2015). Using competencies to explain mathematical item demand: a work in progress. In K. Stacey & R. Turner (Hrsg.), *Assessing mathematical literacy* (S. 85–116). Cham: Springer.
- Watanabe, Y. (2004). Methodology in Washback studies. In L. Cheng, Y. Watanabe & A. Curtis (Hrsg.), *Washback in language testing. Research contexts and methods* (S. 19–36). New Jersey: Lawrence Erlbaum Associate.

- Weinert, F. E. (2001). Vergleichende Leistungsmessung in Schulen – eine umstrittene Selbstverständlichkeit. In F. E. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (S. 17–31). Weinheim: Beltz.
- Wirtz, M., & Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität*. Göttingen: Hogrefe.
- Wittmann, E. C. & Müller G. N. (2007). Muster und Strukturen als fachliches Grundkonzept. In G. Walther (Hrsg.), *Bildungsstandards für die Grundschule: Mathematik konkret* (S. 42-65). Berlin: Cornelsen Verlag.

Weiterführende Literatur

- Ministerium für Schule Nordrhein-Westfalen und Weiterbildung des Landes Nordrhein-Westfalen (2017). *Zentrale Prüfungen 2017. Anforderungen für den Hauptschulabschluss nach Klasse 10 (HSA)*.
- Ministerium für Schule Nordrhein-Westfalen und Weiterbildung des Landes Nordrhein-Westfalen (2019). *Zentrale Prüfungen 2019. Anforderungen für den Mittleren Schulabschluss (Abendrealschule)*.