Analysen zum kognitiven Anspruch von Mathematikaufgaben – Befunde aus zentralen Prüfungen und Lehrerfortbildungen

I n a u g u r a l - D i s s e r t a t i o n

zur

Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Universität zu Köln

vorgelegt von

Dr. Bruno Scheja

aus Breslau

Berichterstatter (Gutachter)	Prof. Dr. Benjamin Rott
	Prof. Dr. Andreas Büchter

Tag der mündlichen Prüfung: 27.03.2019

<u>Für Julian</u>

Dank

Das Verfassen der kumulativen Dissertation war ein ereignisreicher, mehrjähriger Lebensabschnitt, in dem mich bestimmte Personen begleitet und unterstützt haben. Ich möchte mich an dieser Stelle bei ihnen namentlich bedanken.

Mein größter Dank gilt meinem Doktorvater Prof. Dr. Benjamin Rott. Er hat diese (noch) ungewöhnliche Form der Dissertation sehr positiv aufgenommen und mich stets mit impulsreichen, aber auch kritischen Diskussionen und Anmerkungen unterstützt und ermuntert - herzlichen Dank.

Ebenfalls sehr herzlich möchte ich mich für das Interesse an meiner Arbeit bei dem Zweitgutachter Prof. Dr. Anderas Büchter bedanken.

Meinen ganz besonderer Dank gilt ebenfalls den Personen meines privaten Umfeldes. Hier wären zunächst Freunde und Bekannte, die in diesem Zeitraum oft zurückstecken mussten, zu nennen. Besonders möchte ich mich bei Anja Löber bedanken, die die Entstehung der Arbeit aktiv begleitet hat. Ihre fachlichen Anmerkungen waren ein wichtiger Beitrag zu meiner Arbeit.

Mein abschließendender Dank gilt meiner Frau Britta, die mich während dieser arbeitsreichen Zeit unterstützt hat und auch in Zeiten mit Spannungspotential geduldig und umsichtig blieb.

1 Einleitung

In der vorliegenden Dissertation wird der Umgang mit dem Konstrukt des kognitiven Anspruchs von Mathematikaufgaben aus einer Doppelperspektive untersucht; hierbei sind zum einen zwei zentrale Abschlussprüfungen – die polnische Mittelschulprüfung und die nordrhein-westfälische ZP 10 (*erste Studie*) – und zum anderen entsprechende Professionswissensentwicklungen im Rahmen einer Lehrerfortbildung Gegenstand der Betrachtungen (*zweite Studie*).

Im einleitenden Abschnitt der Arbeit werden zunächst die Merkmale des Unterrichts identifiziert, die in besonderer Weise und in besonderem Maße auf das Lernen und somit auch auf die Unterrichtsqualität Einfluss nehmen (1.1). Mit Blick auf das fachbezogene Lernen wird darauf aufbauend das Merkmal kognitive Aktivierung im Zusammenhang mit ihrem wichtigsten „Träger“, den Mathematikaufgaben, betrachtet (1.2). Um den Zusammenhang zwischen Mathematikaufgaben und kognitiver Aktivierung systematisch zu erfassen und zu analysieren, findet man eine Reihe von Ansätzen, in denen Aufgaben nach Kriterien kategorisiert werden. Nach einem Überblick über ausgewählte Anforderungsmerkmale werden kognitive Aufgabenmerkmale vorgestellt und ihr Zusammenhang zum Konstrukt des kognitiven Anspruchsniveaus dargelegt. Dabei stützt sich der Autor auf eine Auswahl empirisch bewährter Klassifikationskategorien, die im Rahmen des COACTIV-Projekts formuliert und in der Folgezeit weiterentwickelt wurden. Die diskutierten Aufgabenkategorien werden hier in ein Klassifikationsschema integriert (1.3), dessen Anwendung eine schrittweise Erarbeitung der *zwei* erkenntnisleitenden Anliegen der vorliegenden Dissertation ermöglicht (1.4).

1.1 Mathematikunterricht vor dem Hintergrund seiner Qualität

1.1.1 Lernprozesse im Mathematikunterricht

Sowohl der deutsche als auch der polnische Mathematikunterricht ist in den vergangenen 20 Jahren unabhängig der Schulstufe von tiefgehenden, zugleich dennoch in hohem Maße zueinander ähnlichen Veränderungen betroffen (vgl. Blum et al. 2010; Karpinski et al. 2013; Konarzewski 2004; Kunter & Baumert 2011; MEN 2009; Maciejak et al. 2012; Scheja 2007; Zahorska 2002; 2012). Diese umfassen beispielsweise mit der Einführung von Standards, Kerncurricula, externen (Abschluss-)Prüfungen und Instituten zur Entwicklung und Auswertung von Leistungstests einen Paradigmenwechsel bei der Sicherung von Lehr-Lernqualität. Wunder spricht zusammenfassend von der „empirischen Wende der KMK“ (2002, S. 141), die als Übergang von input-orientierten zu output-orientierten Steuerungsverfahren bezeichnet und von einer sich sukzessive wandelnden Auffassung von Lehr-Lernarrangements begleitet wurde; letztere werden dabei als Schwerpunkte im Rückgriff auf didaktische und lerntheoretische Orientierungen im Diskurs erkennbar

(Mrozek 2015; Samborska 2015; Drüke-Noe & Kühn 2013), deren Ausdruck nach Samborska (2015, S. 47 f.) eine spezifische Verschiebung des Fokus vom Lehren zum Lernen darstellt (vgl. Moleda & Piesyk 1993; Schratz et al. 2011; Wittmann 1990).
Im Hinblick auf das Lernen als den Kern unterrichtlicher Settings und Prozesse geht man nun davon aus, dass sich die angelegten Lernprozesse zunächst am Verständnis der Lernenden orientieren sollten. Auf dieser kognitiv-konstruktivistischen Sichtweise fußend, wird Lernen nach Baumert und Köller (2000) als ein aktiver, individueller, kumulativer mentaler Prozess aufgefasst, der durch eine aktiv-eigenständige Wissenskonstruktion charakterisiert ist (vgl. Dylak 2013; Shuell 1993; Mietzel 2003; Neubert et al. 2001; Piaget 1928). In einem hohem Maße an das Vorwissen der Schülerinnen und Schüler anknüpfend, werden damit zentrale Voraussetzungen geschaffen, um ausgewogen, wünschenswerte Wissensarten wie Faktenwissen, konzeptuelles Wissen und prozedurales Wissen zu generieren und fortwährend zu vernetzen (vgl. Gagne et al. 1992; Neubrand et al. 2002). Ein zentrales Qualitätsmerkmal eines derartigen verständnisorientierten Wissens sehen Kunter und Voss (2011, S. 86) darin, dass es als „beste Vorbereitung für spätere, gänzlich eigenständig stattfindende Lern- und Problemlösungsprozesse“ betrachtet werden kann.
In dieser Auffassung vom Lernen wird bereits deutlich, dass der Lernerfolg von der Qualität der Lerngelegenheiten abhängt, die ihrerseits zugleich als eine „Ko-Konstruktion“ von Lehrkräften und Lernenden zu betrachten sind. Vor diesem Hintergrund formulieren Baumert et al. zentrale Anforderungen an die prozessgestaltenden Lehrkräfte. So sollen diese „einen Unterricht planen, inszenieren und aktiv gestalten können, der in einem stabilen Ordnungsrahmen die Teilnahmemotivation von Schülerinnen und Schülern sichert, zu kognitivem Engagement und zu verständnisvollem, sinnstiftenden Lernen und zum Erwerb zentraler schulischer Kompetenzen führt, das Bewusstsein des eigenen Könnens stärkt und im besten Fall dauerhaftes dispositionales Interesse an der Sache erzeugt.“ (2011, S. 8). In diesem Ansatz wird deutlich, dass das Lernen im Unterricht nicht einfach „eingeschaltet“ werden kann, sondern vielmehr in geeigneter Weise anzuregen ist (vgl. Prenzel 1995). Eine hinsichtlich der bedingenden Faktoren differenzierende Schematisierung dieser Perspektive findet man beispielsweise in Helmkes Angebot-Nutzungs-Modell von Lernprozessen (vgl. 2004, S. 42), das den Unterricht als ein Lehr-Lern-Wechselspiel zwischen beteiligten Akteuren auffasst. Der von der Lehrkraft durchgeführte Unterricht repräsentiert in seinen Facetten seinerseits ein *Angebot* von Lerngelegenheiten, die sich mit Formen der *Nutzung* dieser durch die Schülerinnen und Schüler verbinden. Das Lernangebot des Mathematikunterrichts umfasst dabei zum einen die durch das Curriculum festgelegten mathematischen Themen, Ziele und Anforderungen und zum anderen die Art und Weise (Darbietungsform, Aufgaben oder Erfolgskontrolle), wie diese Themen durch die Lehrkraft dargeboten werden. In ihrer Bündelung zu einer Lernumgebung führen diese jedoch nicht notwendigerweise und direkt zu gewünschten Wirkungen, da die nachgeordnete Nutzung ihrerseits von Mediationsprozessen auf Schülerseite abhängig ist (ebd.), zu denen nach Schiepe-Tiska et al. (2013, S. 126)

Motivation, Lernaktivitäten und das Interesse zählen. Die *Wirkung* des Unterrichts, d. h. die Effizienz der Nutzung des bereitgestellten Angebots, hängt folglich von der sozial gerahmten Aktivität der beteiligten Lernenden ab.
Auf der Grundlage von theoretischen und empirischen Ansätzen lassen die bisherigen Ausführungen den Unterricht als „eine Gelegenheitsstruktur für verständnisorientierte Lernprozesse“ (Kunter & Voss 2011, S. 87) erkennen. Im Folgenden wird der Frage nachgegangen, welche Merkmale des Lernangebots in besonderem Maße verständnis-orientierte Lernprozesse initiieren und aufrechterhalten.

1.1.2 Dimensionen der Unterrichtsqualität

Vor dem Hintergrund bisheriger Ausführungen zum Lernen innerhalb einer Angebot-Nutzungs-Struktur erscheint es sinnvoll, dass Kunter und Voss die Qualität des Unterrichts daran festmachen, ob und inwieweit es einer Lehrkraft gelingt, „geeignete Strukturen zu schaffen, die den Schülern die Möglichkeit eröffnen, verständnisvolle Lernprozesse zu beginnen und aufrechtzuerhalten, das heißt also für angemessene Anregung zu sorgen“ (2011, S. 87). Derartige Strukturen, die – so der wissenschaftliche Konsens – die Wirkung des Unterrichts in besonderem Maße beeinflussen, können nach Oser und Baeriswyl (2001) zum einen in den *sight structures* (Sichtstrukturen) und zum anderen in den *deep structures* (Tiefenstrukturen) verortet werden. Die Perspektive der Beschreibung von Unterricht mittels Sichtstrukturen bezieht sich auf dessen übergeordnete Organisationsmerkmale, zu denen nach Holzberger und Kunter (2016)

- Organisationsformen (jahrgangsbezogener oder -übergreifender Unterricht, Kurssystem),
- methodische Unterrichtselemente (methodische Großformen, Einzelmethoden wie z. B. Lehrervortrag oder Unterrichtsgespräch) und
- Sozialformen (Individual-, Tandem-, Gruppenarbeit) zählen.

Die Merkmale der Tiefenstrukturen können davon unabhängig variieren und beziehen sich auf die Prozesse der Interaktion zwischen Lehrkräften und Lernenden, den Lernenden untereinander oder der Lernenden und dem Lernstoff (allg.: direkte Lehr-Lern-Prozesse). Dabei zeigen aktuelle Studien der empirischen Unterrichtsforschung, dass Sichtstrukturen den Tiefenstrukturen gegenüber eine deutlich kleinere Erklärungskraft im Hinblick auf den Lernfortschritt von Lernenden besitzen (vgl. Hattie 2012; Kunter & Voss 2011). Bei der konzeptionellen Systematisierung der Tiefenstrukturen lassen sich im aktuellen Diskurs zunächst mit *Effizienz der Klassenführung*, die sich im rahmengebenden und strukturierenden Lehrerhandeln zeigt, sowie *konstruktive Unterstützung*, die vorwiegend am Lernenden orientierende motivationale und affektive Prozesse beinhaltet, zunächst zwei allgemeindidaktische *Dimensionen der Unterrichtsqualität* unterscheiden (vgl. Helm 2016; Kunter & Voss 2011; Riecke-Baulecke 2017). Aufbauend auf der in 1.1.1 beschriebenen kognitiv-konstruktivistischen Annahme über das Lernen werden diese durch eine dritte Qualitätsdimension, die *kognitive Aktivierung*, ergänzt. Sie umfasst und bündelt die

Merkmale des Unterrichts, die das fachbezogene Lernen verständnisorientiert fördern und ist somit aus der Sicht der Mathematikdidaktik von Bedeutung.

1. ***Effizienz der Klassenführung***: Die Befunde der internationalen Unterrichtsforschung zeigen, dass eine effiziente Klassenführung (oder Classroom Management) wesentlich zum Leistungsfortschritt von Klassen beiträgt (vgl. Hattie 2012; Kunter & Voss 2011; Wang et al. 1993). Dabei sind die theoretischen Grundannahmen über die darunter subsumierten Techniken, Strategien und Konzepte nach Pianta und Hamre (2009) in den Arbeiten zur Lehrerunterstützung des selbstregulierten Lernens zu sehen, da sie in erster Linie darauf abzielen, dass Lernende motiviert werden, „sich möglichst lange und intensiv auf die erforderlichen Aktivitäten zu konzentrieren und – als Voraussetzung dafür – den Unterricht möglichst störungsarm zu gestalten“ (Helmke 2004, S. 78). Damit lässt sich eine effiziente Klassenführung zunächst am Ausmaß der aktiven Lernzeit im Fachunterricht festmachen. Vor dem Hintergrund der Anforderungen an eine Lehrkraft, die auf den Klassenunterricht und das pädagogische Handeln Einfluss nehmen (z. B. Öffentlichkeit, Unvorhersagbarkeit oder auch Simultanität) formulieren Evertson et al. (2006) oder Kounin (2006) in der Folgezeit mehrfach empirisch bestätigte Prinzipien, die dabei wesentlich auf die unterrichtliche Steuerungswirkung der Lehrkräfte Einfluss nehmen. Hierzu zählt in erster Linie Kounins Prinzip der Allgegenwärtigkeit der Lehrkraft (*Withitness*), welche in der Lage sein sollte, die im Klassenraum ablaufenden Prozesse zu registrieren, und im Idealfall – wenngleich nicht immer – präventiv zu reagieren. Der Ansatz von Evertson et al. (2006) enthält elf Kategorien eines erfolgreichen Trainingsprogramms an einer Grundschule, die sich durch einen Wandel des Verständnisses hin zu einer schülerzentrierten Klassenführung kennzeichnen. Vor dem Hintergrund, dass neben dem fachlichen Lernen zunehmend auch soziale und emotionale Fähigkeiten gefördert werden sollen, wird dabei die Rolle von Lehrkräften als Gestalter von sozialen Verhaltenserwartungen aufgefasst. Beide Ansätze eint dabei, dass Lehrkräfte, die sich an den Prinzipien orientieren, sich auf diese Weise „prospektiv-vorausschauend und proaktiv verhalten“ (Helmke 2004, S. 84) und auf diese Weise wesentlich weniger Schwierigkeiten in der Klasse haben und damit Ressourcen in Form von aktiver Lernzeit gewinnen.
2. ***Konstruktive Unterstützung***: Wie bereits in 1.1.1 dargelegt, hat das Lern*angebot* keine unmittelbaren oder linearen Effekte auf den Lernertrag; dieser erklärt sich ausschließlich über individuelle Mediationsprozesse (Kognitionen, Motivation und Emotionen), die eine Reorganisation bestehender Wissensstrukturen bedingen (Stankov et al. 2012; Turner et al. 1998). Obwohl eine klare Fassung des Konstruktes „konstruktive Unterstützung“ innerhalb der Unterrichtsforschung noch nicht erreicht ist, kann empirisch begründet davon ausgegangen werden (vgl. Hattie 2012), dass Lernprozesse dann besonders wirksam verlaufen, wenn die Lehrkräfte

das Lernen im Hinblick auf das Ergebnis und den Prozess „aus der Perspektive der Lernenden auffassen und analysieren können“ (vgl. Riecke-Baulecke 2017, S. 159). Als eine zentrale Voraussetzung für diesen Perspektivwechsel sehen Kunter und Voss (2011, S. 89) die domänenspezifische Diagnose der Struktur potentieller bzw. akuter Schwierigkeiten im Lernprozess. Auf diese Weise kann eine „*Lernseits*-Sichtweise“ eingenommen werden (vgl. Schratz et al. 2011), bei der einzelne Aspekte der Lernprozesse (Erklärungen, Fehlerkultur, Feedback oder auch Lern- und Leistungssituationen) derart angelegt werden, dass Lernende in ihrem Streben nach Autonomie, Kompetenz und sozialer Eingebundenheit unterstützt werden. Auf der Grundlage und Befunden der Selbstbestimmungstheorie (SDT; Deci & Ryan 1993) urteilend, tragen hierzu zunächst Faktoren wie *adaptive multiple Erklärungen, Verständnis und Geduld* bei (vgl. Hattie 2012). So bedingt das fachdidaktische wie auch das fachwissenschaftliche Wissen der Lehrkräfte ein tiefergehendes Verständnis des Unterrichtsstoffes, dass ergänzend zu kooperativen Lernformen zum Tragen kommen sollte. Diese Wissensbasis ermöglicht in der Planung und Durchführung der Lernprozesse eine adäquate Auswahl von Aufgaben (und deren Komposition und ggf. Dekomposition), das verständnis- und zielorientierte Formulieren vielfältiger Fragen und Erklärungen. Dabei auftretende Verständnisschwierigkeiten können zugleich erkannt, nachvollzogen und unter Einsatz unterschiedlicher Strategien in vielfältiger Weise angegangen werden. Die zeitliche Taktung derartig angelegter Lernprozesse wird somit adaptiv, d. h. lernseits, bestimmt, was nach Riecke-Baulecke (2017, S. 161) damit einhergeht, dass Lehrkräfte zugleich geduldig im Umgang mit Lernenden sein können. Ein auf der bisher beschriebenen Fähigkeit der Lehrkräfte fußender lernwirksamer Aspekt der Qualitätsdimension konstruktive Unterstützung ist insbesondere der des Gebens und Entgegennehmens eines *formativen Feedbacks*. Als zentraler Bestandteil des *formative assessment* (vgl. Sadler 1989; Black & William 1998) umfasst es nach Klieme et al. zunächst die Formen von Leistungsbeurteilung, „die Informationen über die Diskrepanz zwischen Lernzielen und aktuellem Lernstand liefern und dadurch den Lehrenden und/oder den Lernenden selbst helfen, den weiteren Lernprozess zu gestalten“ (2010, S. 65). Im Gegensatz zum summativen Feedback (vgl. Hattie & Timperley 2007), das einen Bildungsabschnitt auf der Grundlage von Vorgaben bilanzieren soll, zielt das formative Feedback in erster Linie auf die individuelle Leistungsentwicklung, was in Merkmalen wie Zielbezogenheit, Differenziertheit und Wertschätzung einen Niederschlag findet. Es sollte dabei so gestaltet sein, dass es eine präzise Diagnose des Lernstandes ermöglicht und darauf aufbauend positiv auf den Lernprozess wirken kann. Mit dem Ziel einer ersten Operationalisierung spricht Riecke-Baulecke (2017, S. 160) von einem „Dreischritt: Ziel → aktueller Lernstand → Strategie/nächste Schritte“ .

Von einem derart gestalteten formativen Feedback wird im Rahmen der Cognitive

Evaluation Theory (vgl. Deci et al. 1999) eine positive Wirkung auf Motivation und Leistung erwartet, da es „durch die Information über die individuellen Kompetenzen der Lernenden das grundlegende Bedürfnis nach Kompetenz unterstützt." (Klieme et al. 2010, S. 66)

3. ***Kognitive Aktivierung*:** Auf der Grundlage der getroffenen Annahmen über das Lernen wird ein Unterricht als kognitiv aktivierend beschrieben, wenn die Lernenden zu einer aktiven mentalen Auseinandersetzung mit Lerninhalten auf einem für sie optimalen Niveau angeregt werden (Kunter & Voss 2011, S. 88; vgl. Wygotski 1978). Klieme et al. (2009) engen den Begriff ein und verorten ihn zugleich, indem sie feststellen, dass ein derartiger Unterricht inhaltliches Verstehen unterstützt, indem fachliche Konzepte explizit gemacht werden und die Beziehung der elementaren Einheiten untereinander sowie die grundlegenden Ideen, Gesetze, Einsichten, Verfahren und Repräsentationen verdeutlicht werden. Zusammenfassend sehen Klieme et al. (2010) darin die Förderung eines tiefgehenden Verständnisses der Lernenden für die fachlichen Inhalte, die durch eine elaborierte Auseinandersetzung angeregt wird. Einen zentralen Beitrag leisten hierzu Aufgaben und Problemstellungen, die eine höhere kognitive Komplexität aufweisen und somit höhere kognitive Ansprüche an die Lernenden stellen. Diese Prozesse fordern nach Hammer (2016, S. 12) zielgerichtet Argumentationen, die Entwicklung mehrerer Lösungswege oder deren Reflexion beziehungsweise Beurteilung heraus. Das Maß der kognitiven Aktivierung hängt folglich davon ab, inwieweit die Lernenden zu einer fachlich vertieften, selbstständigen Auseinandersetzung mit Inhalten bewegt werden, d. h. inwieweit die mentalen Auseinandersetzungen eine Veränderung, Erweiterung oder Neubildung bestehender Wissensstrukturen bewirken (Riecke-Baulecke 2017, S. 156). Die theoretischen Annahmen über den Einfluss der kognitiven Aktivierung auf das Lernen sind zum einen auf die kognitiven Lerntheorien (vgl. Holzberger & Kunter 2016; Pianta & Hamre 2009) und zum anderen auf die Kognitionsforschung zurückzuführen (Shuell 1996; Mietzel 1993). Aktuelle empirische Befunde, die diese Annahmen zu bestätigen scheinen (vgl. Kunter & Voss 2011), findet man beispielsweise in den Untersuchungen zum japanischen Mathematikunterricht, der im Wesentlichen über kognitiv aktivierende Merkmale konstituiert wird (vgl. Klieme & Bos 2000). Gleichzeitig erzielen japanische Lernende durchgehend sehr gute Ergebnisse im Fach Mathematik, was beispielsweise internationale Vergleichsstudien wie PISA aufzeigen (OECD 2001; MEN 2013). Ein Beitrag von Leuders und Holzäpfel (2011) zur Auffassung des Begriffes kognitive Aktivierung zeigt dennoch, dass das Konzept innerhalb der Mathematikdidaktik unterschiedlich ausgelegt wird, was folgerichtig dazu führt, dass es aktuell auf drei Erscheinungsebenen operationalisiert werden kann.

So wird kognitive Aktivierung

- im Rahmen der COACTIV-Studie in erster Linie anhand der im Unterricht eingesetzten Aufgaben untersucht und bewertet (vgl. Neubrand et al. 2011),
- bei der Pythagoras-Studie primär durch beobachtete Unterrichtsmerkmale zur Beurteilung dieses Merkmals herangezogen (vgl. Klieme et al. 2009),
- in weiteren Studien mittels Selbstwahrnehmungen von Lehrern und Schülern erfasst und beurteilt (vgl. Kunter et al. 2005).

Auf der Grundlage einer kognitiv-konstruktivistischen Auffassung vom Lernen wurden in diesem Abschnitt als Teil der Tiefenstrukturen des Unterrichts drei Basisdimensionen identifiziert (effiziente Klassenführung, konstruktive Unterstützung und kognitive Aktivierung), die unabhängig voneinander und im hohem Maße auf die Leistungen der Lernenden und somit auch auf die Unterrichtsqualität einwirken. Aufgrund des beschriebenen Bezugs zum Forschungsfeld der Mathematikdidaktik[1] wird im Folgenden auf die Basisdimension kognitive Aktivierung im Zusammenhang mit einem ihrer wichtigsten „Träger“, den Mathematikaufgaben, näher eingegangen.

1.2 Der Beitrag von Mathematikaufgaben zur Unterrichtsqualität

Die im Mathematikunterricht eingesetzten Aufgaben bestimmen den Rahmen für die Genese verständnisorientierter Lernprozesse und beschreiben zugleich den Kern der Anforderungen an die Lernenden. Mit dem Ziel einer Festlegung des Begriffes „Aufgabe“ werden vor diesem Hintergrund zunächst dessen Auffassungen vorgestellt (1.2.1) und anschließend Unterschiede in funktionaler Hinsicht diskutiert (1.2.2).

1.2.1 Aufgaben – eine begriffliche Betrachtung

Lehrkräfte steuern mit den eingesetzten Aufgaben die Aktivitäten innerhalb der Lerngelegenheiten, womit ihnen als „‘meeting place’ between teacher and learner“ (Christiansen & Walther, 1986, S. 244) eine zentrale Bedeutung im Lehr-Lern-Gefüge zugesprochen wird. Die Breite der Auffassungen vom Aufgabenbegriff als ein derartiges Bindeglied wird in der Literatur im Wesentlichen dadurch bestimmt, wie eng, bzw. breit die kennzeichnenden Merkmale gefasst werden. Sich durch eine lange Bearbeitungsdauer (z. B. eine Unterrichtsstunde; Neubrand 2002, S. 18) kennzeichnend und damit als sehr weit fassend kann die Definition von Doyle identifiziert werden: „Academic tasks, in other words, are defined by the answers students are required to produce and the routes that can be used to obtain these answers“ (1983, S. 161). Als konstituierende Merkmale legt er dabei *Produkte* (v. a. Aufgabenlösungen in Form von Antworten), *Handlungen* (Tätigkeiten, wie z. B. eine Dreiecks-Konstruktion, die eine Aufgabenlösung bedingt) und *Ressourcen*

1 Hillje (2012, S. 59) ordnet konstruktive Unterstützung und Effizienz der Klassenführung hingegen dem Forschungsfeld der Erziehungswissenschaften zu.

(Wissen und Können der Lernenden, die Aufgabenbearbeitungen eingrenzen) fest, was Hammer (2016, S. 38) mit Zielgerichtetheit, inhaltlicher Eingrenzung und der Vorgabe (implizit oder explizit) des Lösungsweges charakterisiert. Eine bezüglich dieser Merkmale verwandte Definition findet man ebenfalls bei Leuders, in der ergänzend der Aspekt der „Anregung“ (2015, S. 435) zum Denk- bzw. Lernhandeln aufgeführt wird.
Vor allem hinsichtlich der konstituierenden Merkmale auf den Vorarbeiten von Doyle aufbauend, wird die Auffassung des Begriffes Aufgabe von Stein et al. wie folgt weiterentwickelt: „a mathematical task is defined as a classroom activity, the purpose of which is to focus students' attention on a particular mathematical idea. An activity is not classified as a different or new task unless the underlying mathematical idea toward which the activity is oriented changes“ (1996, S. 460). Ein wesentlicher Unterschied zwischen den beiden Auffassungen ist folglich darin zu sehen, dass der Aufgabenumfang bzw. die Bearbeitungsdauer einer Aufgabe deutlich voneinander abweichen, ohne dass Stein et al. eine Aufgabe bereits im Sinne eines „single mathematical problem“ (ebd., S. 485) oder auch von Renkls „kleinsten Interventionen“ (1991, S. 89) bezüglich Sinneinheiten verstehen. Da der Lernprozess im Zuge einer Unterrichtsstunde typischerweise mehrere solcher mathematischer Ideen aufbauend bzw. vernetzend vereint, besteht eine Unterrichtsstunde dieser Auffassung nach aus zwei, drei oder vier Aufgaben, die – im Vergleich zu Doyles Verständnis – zum Teil erheblich kürzere Bearbeitungszeiten benötigen.
Die in dieser Arbeit verwendete verwendete, empirisch bewährte Auffassung vom Aufgabenbegriff (Drüke-Noe 2014; Jordan et al. 2006; Hammer 2016), geht auf die Arbeiten von Neubrand zurück. Sie fasst Aufgaben als *„eine Aufforderung zur gezielten Bearbeitung eines eingegrenzten mathematischen Themas (auf). Aufgaben sind immer Auseinandersetzung mit einem Beispiel eines Sachverhalts.“* (2002, S. 18). Aufgaben werden hier folglich als ein Instrument im Umgang mit Beispielen eines mathematischen Sachverhalts charakterisiert, wobei die zu bearbeitenden Beispiele sowohl konkrete (konkrete Zahlen, konkrete Kontexte, konkrete Größen) als auch allgemeine (variable Größen, Beziehungen zwischen den Ausgangsdaten in Antworten) Beziehungen betrachten können (ebd.). Damit lässt sich der Aufgabenbegriff von Neubrand hinsichtlich der begrifflichen Weite zwischen die Auffassungen von Doyle bzw. Leuders und Renkl bzw. Stein et al. einordnen; vor dem Hintergrund des Ziels der vorliegenden Arbeit erscheint sie damit zugleich hinreichend geeignet; da hier der Umgang mit Mathematikaufgaben auf dem ISCED 2 Level[2] untersucht werden soll, auf dem *erwartet* wird (vgl. MEN 2009; KMK 2003), dass Aufgaben, je nach Anspruchsniveau, sowohl konkrete als auch allgemeine Beziehungen untersuchen.

2 International Standard Classification of Education, lower Secondary (vgl. UNESCO 2011).

1.2.2 Didaktische Grundfunktionen von Aufgaben

Vor dem Hintergrund der in 1.1.1 beschriebenen konstruktivistischen Auffassung vom Lernen, erscheint es sinnvoll, zwischen Aufgaben zu unterscheiden, die einerseits in Lehr-Lern-Prozessen und anderseits im Feld der nachgeordneten Wirkungserfassung verortet werden. Eine solche, hinsichtlich des Verwendungszwecks getroffene Unterscheidung findet man ebenfalls in der Literatur wieder (Büchter & Leuders 2005; Terhart et al. 2009; Luthiger 2014). Diesen Typisierungsansatz verfolgend, formulieren Oelkers und Reusser Anforderungen an die Aufgabenfunktion in einem kompetenzorientierten Unterricht. Bezogen auf Lernangebot materialisieren die dort eingesetzten Aufgaben zunächst „jene Wissens- und Könnenskomponenten, lösen jene Denk- und Arbeitsprozesse aus und aktivieren jene analytischen und synthetischen Figuren des Problemlösens, Argumentierens, Betrachtens und Deutens“ (2008, S. 408), die vorzugsweise in den drei Grunderfahrungen von Winter einen Niederschlag finden (vgl. 1995). Aufgaben, die auf diese Weise eine Schnittstelle der Aktivität und Kommunikation zwischen Lehrkräften und Lernenden bilden, werden als *Lernaufgaben*[3] bezeichnet. In diesem Prozess sind sie nach Neubrand et al. (2011, S. 117) Träger mathematischer Inhalte und bestimmen diese weitgehend, sodass sie als das zentrale Steuerungsmedium für Kompetenzerwerb zu betrachten sind. Die von den Lernenden verlangten bzw. erlebten Anforderungen können folglich an den im Unterricht eingesetzten Aufgaben identifiziert werden. Insofern „leiten“ Lernaufgaben die Lernenden zugleich durch die Grundstruktur von Lerngelegenheiten, die hinsichtlich des Lerncharakters in Experimentier-, Übungs- oder auch Anwendungssituationen unterschieden werden. Die Auswahl, der dabei eingesetzten Lernaufgaben wird zum einen durch curriculare Vorgaben und zum anderen durch das fachdidaktische Wissen der Lehrkräfte maßgeblich beeinflusst (vgl. Baumert & Kunter 2011, S. 37).

Zielt man hingegen nicht auf Entwicklung, sondern auf eine möglichst angemessene Diagnose und Erfassung der Unterrichtswirkungen in Form von Schülerleistungen, so spricht man in funktionaler Hinsicht von *Test-* oder auch *Leistungsaufgaben*. Die Einschätzung des aktuellen Leistungsstands erfolgt dabei in internen oder externen Leistungsevaluationen, die nach Terhart et al. (2009, S. 24) die Funktion haben, zu zertifizieren, welches Können im Lernprozess erworben wurde. Hierbei kann unterschieden werden, auf welcher Ebene bzw. auf welchen Ebenen die Testaufgaben wirksam sein sollen (Systemmonitoring, Unterrichtsentwicklung oder Individualbewertung). So werden Testaufgaben zentraler Prüfungen (eingesetzt in: Zentralen Prüfungen, Vergleichsarbeiten und internationalen Schulleistungsstudien) extern konzipiert, wobei abhängig ihrer Funktion bei der Aufgabenentwicklung beansprucht wird, die in den Vorgaben verankerten Anforderungen in unterschiedlicher Weise zu operationalisieren (vgl. Büchter & Pallack 2012, S. 62 f.). In diesen Aufgaben verankerte kognitive Anforderungsmerkmale können – so die verbreitete Annahme – als ein in den Curricula erwarteter Lernprozessertrag gedeutet

3 Dazu zählen nach Luthiger (2014) zum einen Erarbeitungsaufgaben und zum anderen Übungs- und Anwendungsaufgaben.

werden. Zu den Testaufgaben dezentraler Prüfungen zählen in erster Linie klassen- und stufeninterne Klassenarbeiten bzw. informelle Tests (vgl. ebd.). Ein wesentliches Merkmal der Testaufgaben, die dabei zum Einsatz kommen, ist, dass diese stets durch die Lehrperson auf der Grundlage des erteilten Unterrichts ausgewählt und somit dezentral konzipiert werden. Damit erlauben sie valide Rückschlüsse auf vorgenommene Schwerpunkte und den mathematischen Anspruch des durchgeführten Unterrichts und zeigen darüber hinaus Handlungshinweise in Bezug auf den Einzelschüler bzw. die Lerngruppe auf (vgl. ebd.).
Unabhängig dieser dichotomen funktionalen Typisierung von Mathematikaufgaben liefern vertiefte Struktur- und Anforderungsanalysen dennoch deutliche Hinweise dafür, dass Unterrichts- wie auch Testaufgaben „weitgehend identische kognitive Aufgabenmerkmale aufweisen“ (Drüke-Noe 2014, S. 9). Unterschiede sind hier lediglich im Umgang mit den Lösungen und deren Bewertung feststellbar. So lassen Lernaufgaben Brüche und Divergenzen im Zuge des Kompetenzaufbaus ausdrücklich zu, während bei Testaufgaben in erster Linie Ergebnisorientierung und damit Konvergenz im Vordergrund stehen (vgl. Büchter & Leuders 2005).
Zusammenfassend lässt sich damit festhalten, dass Test- wie auch Lernaufgaben, wie sie im Rahmen dieser Arbeit untersucht werden sollen, in gleicher Weise als Träger kognitiver Aktivitäten und Anforderungen aufgefasst und untersucht werden können.

1.2.3 Merkmale von Mathematikaufgaben und kognitive Aktivierung

Wie in 1.2.1. und 1.2.2 dargelegt wurde, sind Aufgaben ein zentrales Instrument, mit dem Lehrkräfte sowohl den Unterrichtsverlauf als auch die mentalen Aktivitäten der Lernenden innerhalb des Lehr-Lern-Gefüges steuern. Dabei hängt die Initiierung und Aufrechterhaltung solcher Lernprozesse nach Neubrand et al. (2011) davon ab, ob und in welchem Maße Lernende höhere kognitive Operationen ausführen, um die Aufgabe erfolgreich zu lösen. Einen frühen Hinweis auf kognitive Prozesse im Zusammenhang mit Aufgaben findet man bereits bei Prabhus Definition einer Lernaufgabe, die er als „an activity which required learners to arrive to an outcome from given information throug some process of thought, and which allowed teachers to controll and regulate this proces“ (1987, S. 24) auffasst. Derartige kognitiven Prozesse, die beim Lösen einer Aufgabe verlangt werden, können dabei in „Facetten kognitiver Prozesse“ (Neubrand et al. 2011, S. 118) unterschieden werden, deren Orchestrierung zugleich in qualitativer Hinsicht zu unterschiedlichen Lernergebnissen führen kann (vgl. Stein & Lane 1996; Shayer & Adhami 2007). So untersucht eine Studie von Hiebert und Wearne (1993), die in sechsten Klassen eines ländlichen Raumes durchgeführt wurde, den Einfluss von Aufgaben auf die Leistungsentwicklung. Während des zwölfwöchigen Beobachtungszeitraums wurden die Wirkungen in zwei Klassen(-Gruppen) miteinander verglichen, in denen Aufgaben zum einen nach einer konventionellen Lehrbuchmethode im Unterricht eingesetzt wurden. Zum anderen wurde der Stoff mittels Aufgaben mit themengleichen, aktivierend-vernetzenden Zugängen vermittelt (Zugänge durch: Darstellung mit Symbolen, Bildern und Diagrammen, Material, einer Geschichte, einer Geschichte und Aufgabenlösung mit Material). In den abschließenden Tests wurde deutlich, dass unterschiedliche Arten/Zugänge der Aufgaben, in unterschiedlichem Maße effektive Denkprozesse fördern. So zeigten die Mittelwertvergleiche der abschließenden Tests, dass die Lernenden, die den Stoff mittels aktivierend-vernetzender Aufgaben erarbeiteten, zum Teil deutlich größere Leistungszuwächse erreichten (das Vorwissen wurde dabei berücksichtigt); darüber hinaus konnten sie umfassendere Antworten bei alternativen Lösungsansätzen angeben. Annähernd zeitgleich führten Stein und Lane (vgl. 1996) eine dreijährige Studie (QUASAR Project) an vier Projektschulen (6. – 8. Klassen) durch, in der der Zusammenhang zwischen kognitiven Aufgabenmerkmalen (categories of cognitive demands), der Aufgabenimplementation und Leistungsentwicklung (Erfassung mittels schriftlicher Tests und Videoanalysen) untersucht wurden. Dabei wurden 144 Aufgaben über einen Gesamtzeitraum von 3 Jahren eingesetzt, die veränderte Anspruchsmerkmale (hinsichtlich Beziehungshaltigkeit, Lösungswegen und Darstellungsweisen) aufwiesen. Die Autoren stellen dabei zunächst fest: „If we want students to develop the capacity to think, reason, and problem solve then we need to start with highlevel, cognitively complex tasks.“ (ebd., S. 75) Hinsichtlich der kognitiven Aufgabenmerkmale zeigte sich im Einzelnen, dass Lernende dann die größten Leistungszuwächse aufwiesen, wenn sie mit den verwendeten Aufgaben zu hohen kognitiven Aktivitäten (z. B. vernetztes Denken und Argumentieren fördernd) und anspruchsvollen gegenseitigen Austausch angeregt wurden. Demgegenüber zeigten Lernende geringe Lerneffekte, wenn sie im Vorfeld vornehmlich algorithmisch-technische Aufgaben bearbeitet haben (klare, einfache oder routineorientierte Aufgabenstellungen, Standardverfahren, geringer Kommunikationsbedarf).

Kunter und Voss untersuchten im Rahmen des COACTIV-Projekts (Gesamtzeitraum: 2002 – 2006) unter anderem den Einfluss der Qualitätsdimension kognitive Aktivierung auf die Leistungsentwicklung von Neunt- und Zehntklässlern (vgl. 2011). Hierzu fassen sie die Aufgabenauswahl der teilnehmenden Lehrkräfte als das Potential ihrer Lerngelegenheiten zur kognitiven Aktivierung auf. Die insgesamt etwa 44.000 Lern- wie auch Testaufgaben sind dabei so klassifiziert worden, dass darin ihr Beitrag zur kognitiven Aktivierung erkennbar wurde (vgl. Jordan et al. 2006). Das hierzu entwickelte Klassifikationsschema umfasst inhaltliche, curriculare, kognitive und stoffdidaktische Dimensionen, innerhalb derer Aufgaben nach inhalts- sowie tätigkeitsbezogenen Merkmalen analysiert werden, um auf diese Weise die Breite des mathematischen Denkens bei der Aufgabenbearbeitung zu erfassen. Zu diesen vier Dimensionen zählen:

1. ***Inhaltlicher Rahmen:*** Mit den zugehörigen Kategorien wird eine Aufgabe nach inhaltlichen und curricularen Merkmalen klassifiziert, indem sie einerseits einem Stoffgebiet und andererseits einer curricularen Wissensstufe zugewiesen wird (vgl. Jordan et al. 2006; Jordan et al. 2008; Neubrand et al. 2011).
2. ***Kognitiver Rahmen:*** In dieser Dimension werden Typen mathematischen Arbeitens klassifiziert (begriffliche Modellierungsaufgabe, rechnerische Modellierungs-

aufgabe, technische Aufgabe), indem die im Lösungsprozess einer Aufgabe verlangten Denk- und Arbeitsweisen betrachtet werden (vgl. Jordan et al. 2008).

3. ***Kognitive Elemente des Modellierungskreislaufs*****:** Von grundlegender Bedeutung für diese Dimension ist folgende modellhafte Annahme: Der problemhaltige Aufgabenkontext – unabhängig davon, ob inner- oder außermathematisch – wird in ein strukturgleiches Verarbeitungsmodell überführt, in dem Übersetzungs-, Strukturierungs-, Verarbeitungs- und Interpretationsprozesse vorkommen können. Diese Teilprozesse werden zumeist mittels mathematischer Tätigkeiten erfasst, deren Prozessmerkmale auf vier kognitiven Anspruchsniveaus operationalisiert sind (vgl. 1.3.1). Dieses Konstrukt erlaubt somit Aussagen über den Grad der kognitiven Komplexität einer mathematischen Tätigkeit wie auch das kognitive Anspruchsniveau über alle Tätigkeiten und ist zugleich für die vorliegende Arbeit von zentraler Bedeutung (vgl. Neubrand et al. 2002; Jordan et al. 2006).
4. ***Lösungsraum:*** Die zugehörigen Kategorien weisen aus, wie weit eine Aufgabe den Lösungsraum absteckt. Hierzu wird zum einen erfasst, ob die Aufgabenstellung ihrer Bearbeitungsrichtung entspricht oder ob sogenannte „Rückwärtsstrategien“ erforderlich sind. Zum anderen wird die Anzahl der Lösungswege erfasst, die in der Ausgabenstellung explizit verlangt werden (vgl. ebd.).

Die Ergebnisse der COACTIV-Studie zeigen zunächst schulformübergreifend für die 10. Jahrgangsstufe, dass die drei Qualitätsdimensionen in unterschiedlichem Maße auf die Teilaspekte des Lernens wirken (vgl. Kunter & Voss 2011). Dabei beeinflusst der unterrichtliche Einsatz von Aufgaben, die sich durch ein hohes kognitives Anspruchsniveau kennzeichnen, die Leistungsentwicklung der Lernenden mit der höchsten Effektstärke. Solche „qualitätsvollen“ (Biermann et al. 2003, S. 32) Aufgaben regen die Lernenden zu niveauvollen mathematischen Tätigkeiten wie beispielsweise dem Modellieren, Reflektieren, Beurteilen oder dem Argumentieren an und können somit als Indikatoren für einen lernwirksamen Unterricht betrachtet werden.

Zusammenfassend wird an den empirischen Befunden der drei Studien deutlich, dass positive Effekte zwischen kognitiver Aktivierung mittels Lernaufgaben und der Leistungsentwicklung im Mathematikunterricht bestehen. Es zeigt sich dabei insbesondere, dass der Einsatz von Aufgaben, deren Lösungsprozess kognitiv niveauvolle Tätigkeiten verlangt, das fachbezogene Lernen der Schülerinnen und Schüler angemessen initiiert und aufrecht erhält, d. h. verständnisorientiert anregt.

1.3 Kategorialer Umgang mit Mathematikaufgaben

Vor dem Hintergrund des positiven Zusammenhangs zwischen der kognitiven Aktivierung durch Mathematikaufgaben und der Leistungsentwicklung der Lernenden gewinnen kognitive Aufgabenmerkmale und der qualitätsfördernde Umgang mit ihnen an Bedeutung. Um diesen darzulegen, werden zunächst die Entwicklungen ausgewählter

Kategoriensysteme, die einen derartigen Umgang ermöglichen, beschrieben (1.3.1), wonach der Fokus auf die Befunde gerichtet wird, die eine Umsetzung in Abhängigkeit von der Funktion der Aufgaben beleuchten (1.3.2).

1.3.1 Entwicklungen von Kategoriensystemen zur Anspruchsbeschreibung von Mathematikaufgaben

Um das Aufgabenpotential d. h. „eine in der Aufgabe angelegte, aber noch nicht realisierte Nutzungsmöglichkeit für verständnisvolle Lernprozesse“ (Hammer 2016, S. 49) zu erfassen, lassen sich nach Blömeke et al. (2006, S. 33 f.) zunächst sechs elaborierte Kategoriensysteme unterschiedlichen Schwerpunkts identifizieren (vgl. Renkl 1991; Stein et al. 1996, Neubrand 2002, Knoll 2003; Neubrand et al. 2002; Mullis et al., 2003).[4] Als empirisch hinreichend bewährt und aus einschlägiger didaktischer und pädagogisch-psychologischer Literatur entstammend, kann hierzu ebenfalls das zeitlich versetzt entwickelte Kategoriensystem der COACTV-Studie inklusive seiner kategorialen Ergänzungen gezählt werden (vgl. Jordan et al. 2008; Drüke-Noe 2014). Bezüglich der Erfassungsmethoden können diese Studien in Untersuchungen mit differenzierten Verfahren unterteilt werden, in denen die Aufgabenstellung und Aufgabenimplementation im Zusammenhang mit Wirkungen auf die Leistungsentwicklung erfasst werden. Davon sind Studien zu unterscheiden, die zur Analyse und Bewertung des inhärenten Aufgabenpotentials ausschließlich objektiv vorliegende Aufgabenmerkmale heranziehen. Da empirische Befunde darauf hindeuten, dass das instruktionale Handeln der Lehrkräfte einzelne Aufgabeneigenschaften und damit auch Aufgabencharaktere verändern können (vgl. Baumert et al. 1997; Hillje 2012), wird vor dem Hintergrund dieser Arbeit im Folgenden ausschließlich auf Kategoriensysteme eingegangen, die dem zweiten Ansatz zuzuordnen sind. Diese Auswahl wird darüber hinaus dadurch legitimiert, dass die kognitive Komplexität dabei betrachteter Kategorien, d. h. die Kategorieintensität, *nachweislich* zur Schwierigkeit der Aufgaben beträgt. So erfassen beispielsweise Kategorien wie rechnerische (Modellierungs-)Aufgabe, begriffliche (Modellierungs-)Aufgabe und technische Aufgabe nicht das Niveau der kognitiven Prozesse, sondern die Typen mathematischen Arbeitens. Wie Drüke-Noe in ihrer Operationalisierung der Tätigkeit technisches Arbeiten gezeigt hat, kann dieses in allen drei Aufgabenklassen auf einem hohen kognitiven Niveau verlangt werden, womit technischen Aufgaben wie bei Stein et al. (1996) kein niedriges kognitives Anspruchsniveau „an sich“ zugewiesen werden kann (vgl. Neubrand et al. 2002; Jordan et al. 2006).

Ein zentraler Schwerpunkt der Studie von Neubrand et al. (2002) ist zunächst darin verortet festzustellen, aufgrund welcher Anforderungsmerkmale eine Aufgabe schwieriger wird, um aus kognitionstheoretischer Perspektive die inhaltliche Bedeutung von mathematischer Kompetenz zu erhellen. Hierzu wurden insgesamt 117 Aufgaben der nationalen (86) und

4 Eine Übersicht und einen Vergleich dieser Kategoriensysteme findet man bei Neubrand (2002) und Hammer (2016).

internationalen (31) PISA-2000-Untersuchung analysiert. Auf der Grundlage der theoretisch begründeten Annahme, dass das Modellieren den „Kern des mathematischen Denkens“ (ebd. S. 118) ausmacht (vgl. Jordan et al. 2006; Neubrand 2002), konnten dabei als Bestandteil des Kategorienschemas folgende schwierigkeitsgenerierende Aufgabenmerkmale und deren Ausprägungen unterschieden werden:

- *Curriculare Wissensstufe:* Diese Kategorie erfasst, in welchem curricularen Zusammenhang der geforderte Stoff zu verorten ist. Als Ausprägungen werden dabei „Grundkenntnisse“, „einfaches Wissen der Sekundarstufe I“ (vgl. Abb. 2) und „anspruchsvolles Wissen der Sekundarstufe I“ (vgl. 2.1, Abb. 1) unterschieden.
- *Komplexität der Modellierung:* Die Ausprägungen des Modellierens „als Ganzes“ werden durch die Komplexität und den Umfang des Lösungsprozesses von Aufgaben bestimmt. Hierbei lassen sich die Anforderungsstufen „Reproduktion“, „Verknüpfung“ und „Verallgemeinerung“ unterschieden (vgl. Neubrand et al. 2002, S. 111 f.).
- *Art des Kontextes*: Die problemhaltige Situation einer Modellierungsaufgabe kann in einem „innermathematischen“ oder „außermathematischen Kontext“ eingebettet sein (vgl. Abb. 2). Technische Aufgaben, deren Lösungsprozesse keiner Modellierung bedürfen, wird die Ausprägung „ohne Kontext“ zugewiesen (vgl. 2.1, Abb. 1).
- *Umfang der Verarbeitung:* Der Komplexität der Lösungsprozesse hängt oft davon ab, „ob und in welchem Umfang neue Größen und Zwischenergebnisse, die nicht in der Aufgabenstellung selbst vorgegeben sind“ (Neubrand et al. 2002, S. 107) im Lösungsprozess erarbeitet werden müssen. Vor diesem Hintergrund wird dichotom in „hoch“ und „niedrig“ unterschieden, wobei eine Zuordnung zu „hoch“ bei mehr als zwei impliziten Größen erfolgt (vgl. 2.3, Tab. 3, Aufgaben 2 und 4).
- *Offenheit des Modellierungsprozesses:* In dieser Kategorie werden die Aufgaben dichotom danach klassifiziert, ob ihre Aufgabenstellung „einen Ansatz“ oder „mehrere Ansätze“ der Mathematisierung zulassen. Im zweiten Fall wird damit die Anwendung mehrerer Modelle provoziert, womit der „Suchraum“ (Neubrand et al. 2011, S. 123) weiter gesteckt wird.
- *mathematisches Argumentieren:* Abschließend konnte festgestellt werden, dass die explizit in der Aufgabenstellung eingeforderten Argumentationen ebenfalls anspruchsgenerierend wirken. Dieses Aufgabenmerkmal wird mit „ja“ (vgl. 2.1, Abb. 2, „Klassensprecher-Aufgabe“) bzw. „nein“ (vgl. Abb. 2) ebenfalls zweistufig erfasst.

Für die beschriebenen Kategorien konnte in der Studie nachgewiesen werden, dass sich die kognitive Komplexität von Aufgaben mit einer zunehmenden Anzahl von verankerten Denkvorgängen erhöht. Diese werden schrittweise begangen und sind von den Lernenden bewusst zu steuern (vgl. Cohors-Fresenborg et al. 2004).

Ein zentrales Ziel des COACTIV-Projekt war es, die kognitive Aktivierung von

Lerngelegenheiten mittels darin eingesetzter Lern- und Testaufgaben (Einstiegs-, Haus- und Klassenarbeitsaufgaben) zu erfassen und abzubilden (vgl. Jordan et al. 2008; Kunter & Voss 2011; Neubrand et al. 2011).[5] Die oben dargestellten Vorarbeiten von Neubrand et al. (2002) erweiternd, wurde hierzu ein Klassifikationsschema entwickelt, das über eine inhaltliche Analyse hinausgeht und Kategorien berücksichtigt, die zugleich die Dynamik innerhalb der Aufgabenbearbeitung erfassen. Dabei liegt der Fokus der vier Dimensionen (vgl. 1.2.3) und ihrer Beurteilungskategorien primär auf „den Facetten kognitiver Prozesse beim Bearbeiten von Aufgaben“ (Neubrand et al. 2011, S. 118). Die Kategorien werden weiter im Sinne einer Indikatorisierung zumeist vier Ausprägungsstufen ausdifferenziert (vgl. Tab. 1), die im Folgenden dargelegt werden. Weiterführendes zum COACTIV-Klassifikationsschema ist in Jordan et al. (2006) nachzulesen.

Tabelle 1: Zentrale Analysekategorien der COACTIV-Studie (vgl. Jordan et al. 2008)

Dimension	**Kategorie**	**Ausprägung**
Inhaltlicher Rahmen	Stoffgebiet	1- Arithmetik, 2- Algebra, 3- Geometrie, 4- Stochastik
	Curriculare Wissenstufe	1- Grundkenntnisse (Stufe I, in der Grundschule erworben), 2- Einfaches Wissen der Sek. I, 3- Anspruchsvolles Wissen der Sek. I
Kognitiver Rahmen	Typ mathematischen Arbeitens	1- technische Aufgabe, 2- rechnerische Aufgabe, 3- begriffliche Aufgabe
Kognitive Elemente des Modellierungs-kreislaufs	Außermathematisches Modellieren	0- nicht benötigt, 1- Standardmodellierungen, 2- Mehrschrittige Modellierungen, 3- Modellreflexion, -validierung oder -eigenentwicklung
	Innermathematisches Modellieren	0- nicht benötigt, 1- Standardmodellierungen, 2- Mehrschrittige Modellierungen, 3- Modellreflexion, -validierung oder -eigenentwicklung
	Mathematisches Argumentieren	0- nicht benötigt, 1- Standardbegründungen, 2- Mehrschrittige Argumentation, 3- Entwicklung komplexer Argumentationen oder Beurteilen von Argumenten
	Gebrauch mathematischer Darstellungen	0- nicht benötigt, 1- Standarddarstellungen, 2- Wechsel zwischen Darstellungen, 3- Beurteilen von Darstellungen
	Umgehen mit mathematikhaltigen Texten	0- nicht bzw. kaum benötigt, 1- Unmittelbares Textverstehen, 2- Textverstehen mit Umorganisation, 3- Verstehen logisch komplexer Texte
	Grundvorstellungen	0- nicht benötigt, 1- Eine elementare Grundvorstellung oder (triviale) Kombination verwandter elementarer Grundvorstellungen, 2- Eine erweiterte Grundvorstellung oder eine Kombination nichttrivialer elementarer Grundvorstellungen, 3- Mehr als dies
Lösungsraum	Bearbeitungsrichtung	1- vorwärts, 2- rückwärts
	Anzahl der eingef. Lösungswege	0- kein eingeforderter Lösungsweg, 1- ein Lösungsweg, 2- mehrere Lösungswege

Für die vorliegende Arbeit sind die Kategorien der Dimension *Kognitive Elemente des Modellierungskreislaufs* von denen abzugrenzen, die curriculare, inhaltliche (Dimension *Inhaltlicher Rahmen*) und auf die mathematische Arbeitsweise (Dimension *Kognitiver Rahmen*) bezogene Aufgabenmerkmale erfassen. Ihre Bedeutung leitet sich daraus ab, dass damit ausschließlich kognitiv anspruchsgenerierende Tätigkeiten während der Aufgabenbearbeitung abgebildet werden. Dazu werden sie jeweils auf vier *kognitiven*

5 Eine begleitende Unterrichtsbeobachtung fand dabei nicht statt. Neubrand et al. sprechen vor diesem Hintergrund von Aufgaben als dem „Substrat der im Unterricht geschaffenen Lerngelegenheiten.“ (2011, S. 118)

Anspruchsniveaus operationalisiert und drücken damit aus, ob eine Tätigkeit nicht („0“), auf niedrigem („1“), auf mittlerem („2“) oder auf hohem Niveau („3“) im Lösungsprozess verlangt wird. Mit dieser Zuordnung erhält man eine Auskunft über den Grad der kognitiven Komplexität einer Tätigkeit.

Von zentraler Bedeutung ist hierbei – wie bereits in der Studie von Neubrand et al. (2002) –, dass Modellierungsaufgaben einen modellhaften Lösungsprozess aufweisen, dem ein *inner-* oder ein *außermathematischer* Kontext zugrundeliegen kann und unabhängig der beiden Kontexte in „strukturgleichen Zyklen“ (Neubrand et al. 2011, S. 120) durchlaufen wird (vgl. Abb. 1). Erfordert eine Aufgabe lediglich einen kontextlosen Abruf von Fertigkeiten oder Faktenwissen auf bereits vorgegebene Voraussetzungen (vgl. Jordan et al. 2006, S. 34 ff.), so werden verabeitende Leistungen wie das Modellieren oder Strukturieren im Lösungsprozess nicht abgerufen (Niveau: 0). Neubrand et al. (2002, S. 101) sprechen in diesem Fall von einer „technischen Aufgabe”.

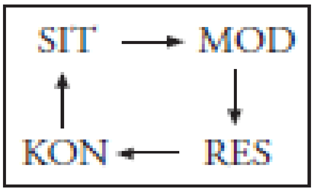


Abbildung 1: Vereinfachter Modellierungskreislauf (Jordan et al. 2006, S. 31)

Der Modellierungs-Kreislauf idealisiert den mehrschrittigen Prozess des Lösens von Modellierungsaufgaben und dabei insbesondere das auf das Lesen und Verstehen der Aufgabenstellung (Kategorie: *Umgang mit math. Texten;* vgl. Cohors-Fresenborg et al. 2004) folgende „Überführen einer problemhaltigen Situation in einen verarbeitbaren Ansatz und den Rückbezug des Ergebnisses auf die Ausgangssituation” (Jordan et al. 2006, S. 31). Neben dem kognitiven Anspruch an die kontextuellen Übersetzungsprozesse wird das beim Durchlaufen des Modellierungskreislaufs aktivierte Argumentationsniveau erfasst. Unter dem *mathematischen Argumentieren* fassen Jordan et al. die Fähigkeit, „eine geschlossene Argumentationskette zu präsentieren oder verschiedene Formen von mathematischen Argumentationen zu verstehen bzw. zu bewerten.“ (ebd., S. 40) Dazu zählen alle Arten explizit eingeforderter Begründungen, so z. B. auch die Reflexion über ein mathematisches Modell. Analog zum Prozess des Modellierens bezieht sich auch diese kognitive Tätigkeit sowohl auf innermathematische als auch auf außermathematische Sachverhalte (in Form eines Beweises bzw. einer Begründung). In der Kategorie *Gebrauch von mathematischen Darstellungen* werden – in Anlehnung an Bruner (1972) – im Wesentlichen *ikonische* Ebenen der Repräsentation hinsichtlich kognitiver Komplexität unterschieden.

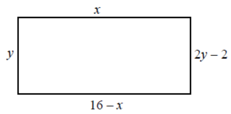
Die vorgestellten Kategoriensysteme beschreiben und erfassten das inhärente Aufgabenpotential mittels der Anspruchsmerkmale, die im Lösungsprozess identifiziert wurden. Die Ansprüche zeigen dabei insgesamt eine Spannweite, die von inhaltlich-curricularen Merkmalen über Elemente mathematischen Arbeitens bis hin zu kognitiven Kategorien reicht. Dabei ist innerhalb der Kategoriensysteme eine zunehmende Ausdifferenzierung feststellbar (kategoriale Breite). Letzteres gilt insbesondere –

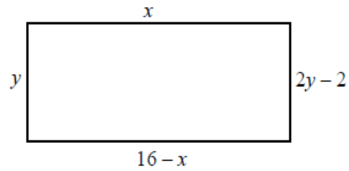
beispielsweise im Vergleich zum Kategoriensystem von Neubrand et al. (2002) – für die Kategorien der Dimension *Kognitive Elemente des Modellierungskreislaufs*, mit denen das kognitive Anspruchsniveau erfasst wird. Sie wurden im Rahmen des COACTIV-Projekts zum Teil neu ausgearbeitet bzw. erneut aufgegriffen und weiterentwickelt. Über all diese Kategorien hinweg zeigt sich zugeich ein Übergang von einer dichotomen Kodierung hin zu einer differenzierten Operationalisierung der kognitiven Komplexität auf vier Niveaus (Operationalisierungstiefe). Drüke-Noe (2014) ergänzt in der Folgezeit die Dimension Kognitive Elemente des Modellierungskreislaufs des COACTIV-Kategoriensystem um die Analysekategorie technisches Arbeiten, das entsprechend auf vier kognitiven Anspruchsniveaus operationalisiert wird. Die derart zusammengefassten kognitiven Tätigkeiten werden in dieser Arbeit dazu herangezogen, innerhalb der Spannweite des mathematischen Denkens, den Grad der kognitiven Aktivierung – das kognitive Anspruchsniveau – von Aufgaben zu erfassen. Dabei ermöglicht die gemeinsame Betrachtung der sieben Tätigkeiten eine deutlich differenziertere Formulierung der Ergebnisse zum kognitiven Anspruchsniveau von Lern- wie auch Testaufgaben als dies mittels der Kategoriensysteme von Neubrand et al. (vgl. 2002) wie auch des COACTIV-Projektes (vgl. Jordan et al. 2008) der Fall war (vgl. 1.3.2).

Um die Anwendung des Klassifikationschemas aus Tab. 1 beispielhaft zu verdeutlichen, werden zunächst ausgewählte Testaufgaben der polnischen Mittelschulprüfung vorgestellt (vgl. Abb. 2).

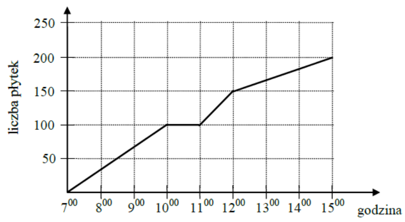
Abbildung 2: Beispiele für Testaufgaben der Mittelschulprüfung aus den Jahren 2014, 2012, 2006 und 2011 (von oben nach unten) (vgl. CKE 2012, S. 3; CKE 2014, S. 7; CKE 2006, S. 3; CKE 2011b, S. 2).

<table>
<tr>
<td>
Aufgabe 1 (1 Rohpunkt)[6]

Na rysunku przedstawiono prostokąt, którego wymiary są opisane za pomocą wyrażeń.


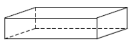
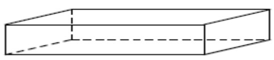
Oceń prawdziwość podanych zdań. Wybierz P, jeśli zdanie jest prawdziwe, lub F – jeśli jest fałszywe.
<table>
<tr><td>Jeden z boków prostokąta ma długość 8.</td><td>P</td><td>F</td></tr>
<tr><td>Obwód prostokąta jest równy 20.</td><td>P</td><td>F</td></tr>
</table>
</td>
<td>
In der Abbildung ist ein Rechteck dargestellt, dessen Maße als Terme angegeben wurden.


Bewerte die Wahrheit der Sätze. Wähle W, wenn der Satz richtig ist, F – wenn er falsch ist.
<table>
<tr><td>Eine Seite des Rechtecks hat die Länge 8.</td><td>W</td><td>F</td></tr>
<tr><td>Der Umfang des Rechtecks beträgt 20.</td><td>W</td><td>F</td></tr>
</table>
</td>
</tr>
<tr>
<td>
Aufgabe 2 (1 Rohpunkt)

Glazurnik układał płytki. Wykres przedstawia liczbę ułożonych płytek w zależności od czasu w trakcie ośmiogodzinnego dnia pracy.


Na podstawie wykresu wybierz zdanie <u>fałszywe</u>.
A. O godzinie 10^{00} glazurnik rozpoczął godzinną przerwę.
B. Od 7^{00} do 8^{00} glazurnik ułożył mniej płytek niż od 11^{00} do 12^{00}.
C. W ciągu każdej godziny glazurnik układał taką samą liczbę płytek.
D. Przez ostatnie trzy godziny pracy glazurnik ułożył 50 płytek.
</td>
<td>
Ein Fliesenleger verlegt Fliesen. Der Graph zeigt die Anzahl der verlegten Fliesen in Abhängigkeit von der Zeit während seiner achtstündigen Arbeitszeit.
Wähle eine <u>falsche</u> Antwort aus.


A. Um 10 Uhr hat der Fliesenleger eine einstündige Pause begonnen.
B. Von 7 Uhr bis 8 Uhr hat der Fliesenleger weniger Fliesen verlegt als zwischen 11 Uhr und 12 Uhr.
C. In jeder Stunde verlegt der Fliesenleger gleich viele Fliesen.
D. In den letzten drei Stunden hat der Fliesenleger 50 Fliesen verlegt.
</td>
</tr>
<tr>
<td>
Aufgabe 3 (1 Rohpunkt)

Cegła ma kształt prostopadłościanu o wymiarach 24 cm × 12 cm × 6 cm. Jakie są wymiary ścianki cegły, którą ta cegła powinna przylegać do podłoża, aby wywierać na nie jak największe ciśnienie?

A. 12 cm × 6 cm
B. 12 cm × 24 cm
C. 24 cm × 6 cm
D. Za mało danych, by odpowiedzieć.
</td>
<td>
Ein Ziegelstein hat die Form eines Quaders mit den Maßen
24 cm × 12 cm × 6 cm. Welche Maße hat die Seite, auf die der Ziegelstein auf dem Boden liegen sollte, um den größten Druck auszuüben?

A. 12 cm × 6 cm
B. 12 cm × 24 cm
C. 24 cm × 6 cm
D. Zu wenige Informationen, um zu antworten.
</td>
</tr>
<tr>
<td>
Aufgabe 4 (1 Rohpunkt)

Średnia arytmetyczna pięciu ocen cząstkowych Jacka jest równa 3,4. Jaką średnią ocen będzie miał Jacek, gdy otrzyma jeszcze czwórkę?
A. 4,2 B. 3,7 C. 3,5 D. 3,8
</td>
<td>
Das arithmetische Mittel von vier einzelnen Noten von Jacek beträgt 3,4. Welchen Durchschnitt wird Jacek haben, wenn er noch eine vier erhält.

A. 4,2 B. 3,7 C. 3,5 D. 3,8
</td>
</tr>
</table>

6 Bei allen Aufgaben aus Abbildung 2 handelt es sich um „geschlossene Aufgaben“ (vgl. CKE 2010), bei denen im Sinne einer dichotomen Auswertung zwischen einem vorgegebenen, richtigen (stets 1 Rohpunkt) und einem falschen Zielzustand (0 Rohpunkte) unterschieden wird. Im Falle von Aufgabe 1 bedeutet dies beispielsweise, dass beide Teilaufgaben korrekt gelöst sein müssen, um einen Rohpunkt zu erhalten (vgl. Scheja 2017a, S. 301 ff.).

Die nachstehende Tabelle 2 zeigt einen Überblick über die Zuordnung der in Abbildung 2 aufgeführten Testaufgaben zu den Niveaus der Kategorien, die für die vorliegende Untersuchung relevant sind.

Tabelle 2: Zuordnung der Testaufgaben aus Abb. 2 zu den Kategorien und ihren Ausprägungen aus Tab. 1.

	Kategorien und ihre Ausprägungen								
Aufgabe	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
1	3	2	2	0	2	0	1	1	1
2	2	3	2	2	0	0	1	2	1
3	3	2	3	2	0	0	0	1	1
4	4	2	2	2	0	0	0	2	2

Zuordnung der Zahlen aus Tabelle 2 zu den Kategorien:

1 – Stoffgebiet; 2 – curriculare Wissensstufe; 3 – Aufgabenklasse; 4 – außermathematisches Modellieren; 5 – innermathematisches Modellieren; 6 – mathematisches Argumentieren; 7 – Gebrauch math. Darstellungen; 8 – Umgang mit math. Texten; 9 – technisches Arbeiten

Die in Tabelle 2 dargestellte Klassifizierung innerhalb der hier betrachteten Kategorien wird im Folgenden anhand von Aufgabe 3 verdeutlicht: Den inhaltlichen Rahmen der Bearbeitung, das Stoffgebiet, bildet hier die Geometrie (Ausprägung 3), da sowohl konstituierende Eigenschaften eines Quaders als auch der Flächeninhalt ebener Figuren benötigt werden (thematische Bereiche: Geometrische Grundbegriffe sowie Flächen- und Rauminhalte (vgl. Jordan et al. 2006). Auf der Grundlage des Curriculums erfordert die Aufgabe damit „einfaches Wissen der Sekundarstufe I“ (Ausprägung 2). Im Verarbeitungsprozess wird in qualitativer Hinsicht vornehmlich konzeptuelles Denken verlangt (folglich Aufgabenklasse: begriffliche Modellierungsaufgabe (Ausprägung 3), da der „verdeckte“ antiproportionale Zusammenhang, zunächst allgemein, und anschließend zwischen den zu betrachtenden Größen „Seitenfläche des Quaders“ und „zugehöriger Druck der Seitenfläche auf die Unterlage“ als tragfähig zu identifizieren ist, bevor auf dieser Grundlage mittels einer Rückwärtsstrategie eine Antwort – 12 cm × 6 cm als Maße der kleinsten Fläche – zu den Maßen angegeben werden kann. Der somit mittels eines Modellierungskreises beschreibbare Lösungsprozess (vgl. Abb.1) wird durch eine Aufgabestellung initiiert, deren sprachlogische Komplexität eher dem einfachen Niveau zuzuordnen ist (Ausprägung 1). So entspricht die Reihenfolge der Sätze nicht unmittelbar den Schritten der Bearbeitung und die genannten Größen sind nicht unmittelbar für die Lösungsfindung zu übernehmen, da beispielsweise eine Zuordnung jeweils zweier Seitenlängen zum erzeugten Druck der zugehörigen Fläche nicht offen gelegt wird. Das anschließende, außermathematisch verortete Modellieren ist ebenfalls auf mittlerem Anspruchniveau angesiedelt (innermathematisches Modellieren wird damit Ausprägung 0 zugewiesen). So sind hier Verbindungen herzustellen, die über den in der Aufgabe angesprochenen Gegenstand hinausgehen; der Lösungsansatz ist aus den Vorgaben der Aufgabe somit nicht unmittelbar abrufbar, was insgesamt mehrere Modellierungsschritte erfordert (vgl. Jordan et al. 2006, S. 36). Mathematisches Argumentieren erfordert die Aufgabe nicht (Ausprägung 0); der Umgang mit mathematischen Darstellungen ist – trotz des abgebildeten Quaders – nicht erforderlich, da dieser keine über den Text hinausgehenden Informationen enthält. Somit wird diese Kategorie mit 0 klassifiziert. Schließlich sind die Anforderungen an den Umgang mit Kalkülen, das technische Arbeiten, als niedrig einzustufen, da keine hierarchischen Techniken (weder Punkt- vor Strich- noch Potenz- mit Punkt- und Strichrechnung) im Prozess der Lösungsfindung benötigt werden (vgl. Drüke-Noe 2014, S. 74 f.). Insgesamt kann damit festgestellt werden, dass das außermathematische Modellieren, der Umgang mit mathematischen Texten sowie das technische Arbeiten ausschließlich das kognitive Anspruchsniveau von Testaufgabe 3 bestimmen.

Aus dem vorgestellten, innerfachlich verorteten Klassifikationsansätzen wurden in der Folgezeit vor allem fächerübergreifende Kategoriensysteme entwickelt und empirisch erprobt. Hier ist zum einen das siebendimensionale Kategoriensystem von Maier et al. zu nennen (vgl. 2010), das unter Rückgriff auf allgemeindidaktische Lernzieltaxonomien wie auch „wiederkehrende Begriffe in fachspezifischen Kategoriensystemen" (Drüke-Noe et al. 2017, S. 213) das kognitive Potenzial von Aufgaben erfasst. Zum anderen haben Blömeke et al. (2006) ebenfalls ein fächerübergreifendendes Analysemodell entwickelt, das erweiternd „zwischen dem objektiven Potential einer Aufgabe, den intedierten Anforderungen seitens der Lehrkraft und der Realisierung dieser Anforderungen im Lehr-Lernprozess unterscheidet." (ebd., S. 330)

1.3.2 Befunde zum Umgang mit Kategoriensystemen zur Anspruchsbeschreibung von Mathematikaufgaben

Die in 1.1.1 beschriebene Forschungsperspektive des Angebot-Nutzen-Modells versucht modellhaft Faktoren der Unterrichtsqualität in ihrer Wirkungsweise auf Lernprozesse zu integrieren. Unter Einbezug der Ausführungen der Abschnitte 1.1 und 1.2 kann *nun* angenommen werden, dass Aufgaben zum einen in hohem Maße die Grundstruktur des Lern*angebots* bestimmen und zum anderen auch dessen *Wirkungen* erfassen. Wendet man sich vor diesem Hintergrund der Frage zu, welchen Personenkreisen eine Auswahl und Konzeption von Lern- wie auch Testaufgaben zukommt, die den Lernenden während ihrer Bildungsbiographie begegnen (vgl. Büchter & Pallack 2012), so sind hier im Wesentlichen

- Lehrkräfte an Einzelschulen und
- ausgebildete Aufgabenkonstrukteure damit beauftragter Institute zu unterscheiden (Polen: Zentrale Prüfungskommission (Centralna Komisja Egzaminacyjna (CKE) und NRW: (Qualitäts- und Unterstützungsagentur (QUA-LIS).

Zur Erfassung des kognitiven Anspruchsniveaus von Mathematikaufgaben, die beide Personenkreise umsetzen, erscheinen im Wesentlichen zwei Ansätze sinnvoll: So kann dieses anhand von curricularen Vorgaben, wie inhaltliche Standards und Richtlinien untersucht und bewertet werden. Hierbei geht man von einer weitest gehenden Passung der dort verankerten bildungspolitischen Wirkungserwartungen des Unterrichts und der getroffenen Themenauswahl aus. Da aktuelle empirische Befunde hierzu jedoch großflächige Diskrepanzen zwischen Wirkungs*erfahrungen* und Wirkungs*erwartungen* im Fach Mathematik aufzeigen, scheint der Ansatz der Erfassung eines curricular verankerten kognitiven Anspruchsniveaus als maßgeblicher und ausschließlicher Impulsgeber für den Unterricht offenbar nur eingeschränkt geeignet (vgl. Büchter & Pallack 2012; Neubrand & Neubrand 2010; Karpinski et al. 2013; MEN 2009; Kühn & Drüke-Noe 2013). Dem zweiten Ansatz folgend, wird das kognitive Anspruchsniveau in den eingesetzten Aufgaben selbst erkennbar, wobei Konzeptionsmerkmale von zentralen Testaufgaben zunehmend eine steuernde Wirkung auf das Lernarrangement besitzen (vgl. Drüke-Noe 2014, S. 150 ff.; Konarzewski 2008; Maier et al. 2011). Auf empirische Befunde, die eine derartige kategorial orientierte Konzeption von Aufgaben innerhalb der beiden Personenkreise beleuchten, wird im Folgenden kurz eingegangen.

Bisher analysieren nur wenige Studien das kognitive Anspruchsniveau von Lern- und schulinternen Testaufgaben (Auswahl/Konzeption: Lehrkräfte) ausschließlich mit dem Fokus auf ihre Merkmale (Neubrand 2002; Jordan et al. 2008; Drüke-Noe 2014). So bewertet J. Neubrand das Anspruchsniveau der deutschen Lernaufgaben im Rahmen der TIMSS-Videostudie insgesamt als niedrig. Dies führt sie darauf zurück, dass fast 90% der eingesetzten Aufgaben technischen Typs sind, Modellierungen insgesamt kaum gefordert werden und diese unausgewogen über die Stufen der kognitiven Komplexität streuen. Damit trägt der Grad der verlangten Komplexität von Fertigkeiten und Faktenwissen in sehr hohem Maße zum kognitiven Anspruchsniveau der deutschen Aufgaben bei und prägt diese zugleich wesentlich.

In die gleiche Richtung weisend, in Bezug auf das kognitive Anspruchsniveau jedoch differenzierter, sind die Ergebnisse der COACTIV-Studie (vgl. 1.3.1): Insgesamt zeigen die von den deutschen Lehrkräften eingesetzten Aufgaben und damit auch ihr Unterricht hier ein sehr niedriges Anspruchsniveau. Während die Autoren – vor dem Hintergrund einer ausgewogenen Realisierung der Winter'schen Grunderfahrungen (Winter 1995) – über alle kognitiv anspruchsgenerierenden Kategorien (Argumentieren, inner- und außermathematischen Modellieren, Umgang mit math. Darstellungen und Umgang mit math. Texten) hinweg Mittelwerte zwischen 1 und 2 erwarten (Neubrand et al. 2011, S. 126), erreichen diese in keiner Kategorie im Schnitt 0,5 (ebd., S. 129). Schulformeffekte treten dabei entgegen der Erwartungen nur punktuell auf. So wird beispielsweise das Argumentieren am Gymnasium in 10 % aller Klassenarbeitsaufgaben verlangt, während es in anderen Schulformen nur jede fünfzigste ist.

Die empirischen Befunde von Drüke-Noe (2014) zum kognitiven Anspruchsniveau (verwendete Kategorien: innermathematisches und außermathematisches Modellieren, mathematisches Argumentieren, Gebrauch von math. Darstellungen und technisches Arbeiten) von Klassenarbeitsaufgaben erweitern die COACTIV-Ergebnisse, indem zum einen Verteilungen der Aufgaben auf die kognitiven Komplexitätsniveaus der Tätigkeiten angegeben werden. Zum anderen liefert die Berücksichtigung der mathematischen Tätigkeit „technisches Arbeiten" wichtige Erklärungsansätze für lerntheoretische Orientierungen der Aufgaben und damit auch des erteilten Unterrichts (vgl. Neubrand 2002, S. 24 ff.). So verlangt die Bearbeitung der Klassenarbeitsaufgaben erwartungskonform nur in geringen Maße Tätigkeiten wie außer- oder innermathematisches Modellieren, Argumentieren oder auch den Gebrauch von Darstellungen. Schulformunabhängig werden diese Tätigkeiten auf dem mittleren Anspruchsniveau allenfalls nur punktuell, auf dem hohen Anspruchsniveau typischer Weise gar nicht verlangt. Sofern sie überhaupt vorkommen, genügen im Lösungsprozess lediglich Standardaktivitäten auf einfachem Anspruchsniveau. Dagegen kommt dem technischen Arbeiten, das in erster Linie in technischen Aufgaben gefordert wird, eine sehr hohe Bedeutung zu. Damit konnte in der Studie insgesamt gezeigt werden, dass der Anspruch an das kontextlose, mechanische „Umgehen mit Kalkülen einen sehr wesentlichen Anteil am insgesamt in diesen Aufgaben abgebildeten kognitiven Anspruchsniveau hat." (Drüke-Noe 2014, S. 243). Letzteres scheint die Studie von Karpinski et al. (2013), die landesweit an 20 polnischen Mittelschulen durchgeführt wurde, tendenziell auch für den polnischen Unterricht zu bestätigen. Hier zeigen dreimonatige Unterrichtsbeobachtungen, in denen der Implementierungsgrad „allgemeiner mathematischer Anforderungen" (vgl. MEN 2009) erfasst wurde,[7] dass die zentralen Merkmale dabei eingesetzter Lernaufgaben *ebenfalls* mit „einübend und algorithmusorientiert" (Karpinski et al. 2013, S. 25) umschrieben werden.

Studien, die das kognitive Anspruchsniveau von externen Testaufgaben (Auswahl/Konzeption: Aufgabenkonstrukteure) mittels des vorgestellten Konstruktes erfassen, sind bisher allenfalls punktuell vorhanden.[8] So untersucht aus qualitätsorientierter Perspektive lediglich Drüke-Noe (2014, S. 136 ff.) die Aufgaben der hessischen MSA-Prüfungen[9] aus dem Zeitraum zwischen 2007 und 2009. Hier zeigt sich, dass die Aufgaben insgesamt über eine relativ große Bandbreite von Tätigkeiten streuen und dabei vor allem beim Gebrauch von Darstellungen, dem außermathematischen Modellieren und beim Argumentieren zum Teil als hoch einschätzbare kategoriale Mittelwerte erreichen (vgl. ebd., S. 160 f.; Neubrand et al. 2011, S. 128). Gleichzeitig kommt dem technischen Arbeiten eine im Vergleich zu den oben beschriebenen Klassenarbeiten geringere Bedeutung zu, womit der Umgang mit Kalkülen keinen derart herausragenden Anteil am gesamten kognitiven Anspruchsniveau der hessischen MSA-Aufgaben hat.

7 Eine Unterscheidung in der Ausprägung der jeweiligen mathematischen Anforderungen wurde dabei nicht vorgenommen.

8 Zu den Studien, in denen Testaufgaben zentraler Abschlussprüfungen hinsichtlich ihrer Verteilung auf die *allgemeinen mathematischen Kompetenzen* und ihre Anforderungsbereiche untersucht werden, zählen bislang die von Neubrand und Neubrand (2010) sowie die von Kühn und Drüke-Noe (2014). Aufgrund ihrer Anlage sind diese im Wesentlichen als fachbezogene Implementationsstudien zu betrachten.

9 Schriftliche Abschlussprüfungen zum Erwerb des Mittleren Schulabschlusses, im Folgenden kurz: MSA.

Insgesamt wird aus den vorgestellten empirischen Ergebnissen deutlich, dass Lern- wie auch Testaufgaben, deren Konzeption und Einsatz durch Lehrkräfte erfolgt, ein sehr niedriges kognitives Anspruchsniveau aufweisen. Der unverhältnismäßig hohe Beitrag des technischen Arbeitens zum kognitiven Anspruchsniveau der Aufgaben kann dabei als ein Bruch zu den beschriebenen Annahmen über das Lernen und dessen qualitätsbeeinflussende Faktoren gesehen werden (vgl. 1.1, 1.2).[10] Erste, bisher jedoch als kaum belastbar einzustufende Ergebnisse zu Testaufgaben zentraler Prüfungen deuten hingegen auf ein höheres kognitives Anspruchsniveau hin, das über eine größere Bandbreite von eingeforderten Tätigkeiten und deren mittlere Anspruchsniveaus generiert wird.

1.4 Ziele und Aufbau der Arbeit

1.4.1 Erste Perspektive des Umgangs – Testaufgaben zentraler Prüfungen

Dass innerhalb des Angebot-Nutzungs-Modells durchaus eine gegenseitige Abhängigkeit zwischen der Art und Weise der Wirkungserfassung des Unterrichts und der Qualität der Lernangebote besteht (vgl. 1.1.1), zeigt sich unter anderem daran, dass aktuell zunehmend davon ausgegangen wird, dass Konzeptionsmerkmale von Testaufgaben eine qualitätsbeeinflussende Wirkung auf das Lernarragement besitzen (vgl. Büchter & Pallack 2012; Konarzewski 2008; Maier et al. 2011). Hierbei scheint das Wirkungsmaß bei den Tests, die primär die Funktion „Individualbewertung" (vgl. Drüke-Noe 2014, S. 150 ff.; Konarzewski 2004; Maier et al. 2011) ausweisen, am stärksten zu sein (Büchter & Pallack 2012). In Anbetracht dieser, offenbar weitflächig verbreiteten Form der impliziten Normierung von Gelegenheitsstrukturen für Lernprozesse gewinnt die Frage der *kognitiven Aufgabenmerkmale* und deren Komposition seitens der fachdidaktischen Forschung wie auch der Implementationsforschung auch in Bezug auf Testaufgaben zunehmend an Bedeutung (vgl. 1.3.1).

Den Ausführungen zum kognitiven Anspruch von Mathematikaufgaben zentraler Abschlussprüfungen zur Folge (vgl. 1.3.2), beleuchten bisherige Untersuchungen dieses Feld nur punktuell, wobei internationale Vergleichsstudien gänzlich fehlen. Dieses Forschungsdesiderat wird im Rahmen der vorliegenden kumulativen Dissertation aufgeriffen, indem das Konstrukt der kognitiven Aktivierung, wie auch das Teilkonstrukt des kognitiven Anspruchs, zugrunde gelegt wird (vgl. 1.2); in Anlehung an Hammer (2016) und Neubrand et al. (2011) werden diese Konstrukte jeweils als spezifische inhärente Aufgabenpotentiale definiert, die als Kategoriensysteme im Rahmen des COACTIV-Projekts entwickelt und anschließend von Drüke-Noe ergänzt wurden (vgl. 1.3.1). Die Anwendung dieser Klassifikationsschemata ermöglicht auf der Grundlage einer rationalen Aufgabenanalyse (vgl. Resnick & Ford 1981) eine schrittweise Erarbeitung des *ersten*

10 Zu einer Einschätzung der Ursachen für diese Befunde tragen Hospesova & Ticha (2010) oder auch Hillje (2012) bei. In die gleiche Richtung weisend zeigen sie in qualitativen Studien, dass (angehende) Lehrkräfte nur punktuell in der Lage sind, Aufgabenpotentiale zu erkennen.

zentralen Anliegens der Dissertation:

- ***Das kognitive Aktivierungspotential von Mathematikaufgaben der polnischen Mittelschulprüfung und der nordrhein-westfälischen ZP 10 soll erfasst, analysiert und bewertet werden.***

Dieses Anliegen wird im Einzelnen in drei Teilfragestellungen ausdifferenziert, indem sie zunächst einer qualitäts- und anschließend einer vergleichsorientierten Forschungsperspektiven folgen (vgl. 2.1 und 2.2):

- *Wie verteilen sich die Prüfungsaufgaben der polnischen Mittelschulprüfung und der ZP 10 auf verschiedene Aufgabenklassen und welche Anforderungsniveaus lassen sich in Bezug auf curriculare und kognitive Erfassungskategorien identifizieren?*
- *Lassen sich in einem Vergleich zwischen der Mittelschulprüfung und der nordrhein-westfälischen ZP 10 Unterschiede hinsichtlich der Verteilungen auf kognitive Kategorien ausmachen?*

Die ebenfalls am Ende des ISCED 2 Levels durchgeführte Zentrale Prüfung des Landes Nordrhein-Westfalen (kurz ZP 10) eignet sich für einen derart angelegten bewertenden Vergleich in hohem Maße. So ist der Anteil der ZP-10-Prüflinge an Prüflingen, die deutschlandweit an MSA teilnehmen, in Nordrhein-Westfalen im Vergleich der deutschen Bundesländer am höchsten. Darüber hinaus weisen beide Prüfungsinstrumente hinsichtlich Selektion und Vergleichbarkeit eine hohe funktionale Äquivalenz auf und der Umgang mit den Prüfungsergebnissen seitens der Bildungspolitik (z.B. Prozedur der Ergebnisrückmeldungen) zeigt zahlreiche Gemeinsamkeiten (vgl. Büchter & Pallack 2012; CKE 2010; CKE 2011a).

Die Arbeit geht über eine Erfassung und Analyse des „Ist-Zustandes" hinsichtlich der Testaufgaben der polnischen Mittelschulprüfung hinaus und integriert zusätzlich eine Entwicklungsperspektive, indem Untersuchungen als eine Längsschnittstudie mit folgender Fragestellung angelegt werden (vgl. 2.3):

- *Zeigen die Mathematikaufgaben der polnischen Mittelschulprüfung verschiedener Durchführungszeiträume (2002 – 2005 (Zeitraum nach der Einführung der Mittelschulprüfung) und 2012 – 2015 (Zeitraum nach der Reform der Mittelschulprüfung)) Veränderungstendenzen bezüglich des kognitiven Anspruchsniveaus und welche Erklärungsansätze lassen sich hierfür identifizieren?*

Mit dem *ersten* Anliegen der kumulativen Dissertation greift der Autor die skizzierten Forschungsdesiderata auf, um vornehmlich der allgemein- und fachdidaktischen Diskussion, die derzeit um die spannungsgeladenen Themen der Aufgabenkultur von Test- und Lernaufgaben geführt wird (Büchter & Pallack 2012; Drüke-Noe 2014; Krüger 2015; Kühn & Drüke-Noe 2013; Neubrand et al. 2011), ein anforderungsbezogenes Profil von Testaufgaben zweier kürzlich reformierter Bildungssysteme darzulegen und schließlich zu vergleichen. Von besonderem Interesse ist hierbei zugleich die Zugrundelegung der jeweils gültigen curricularen Vorgaben, wie inhaltlicher Standards und Richtlinien, die verbindliche

Kompetenzerwartungen setzen und eine Evaluationsfunktion für interne und externe Prüfungen erfüllen. Sie sind damit ein zentrales Instrument zur Qualitätssicherung und -entwicklung und werden – so die bildungspolitische Annahme hinsichtlich ihrer Implementation – auch in Bezug auf die kognitven Anforderungsmerkmale den Testaufgaben zugrunde gelegt. Vor dem Hintergrund der Ergebnisse von Längsschnittuntersuchungen von Neubrand und Neubrand (2010) sowie Kühn und Drüke-Noe (2013), in denen den deutschen Testaufgaben länderübergreifend a) eine geringe Passung mit den Bildungsstandards für den Mittleren Schulabschluss und zugleich b) eine konzeptionelle Konstanz offen gelegt wurden, ist es ebenfalls von Interesse, ob derartige Entkopplungstendenzen (Standards → Testaufgaben) zum einen bestätigt (im Falle der ZP 10) und zum anderen auch international feststellbar sind.

1.4.2 Zweite Perspektive des Umgangs – Lehrkräfte

Situationen, in denen die Wirkungen des Unterrichts erfasst und bewertet werden, sind ko-konstruktiv ablaufende Lerngelegenheiten vorgeschaltet (vgl. 1.1.1); gemäß des Angebot-Nutzungs-Modells und der Ausführungen zur funktionalen Typisierung darin verortbarer Aufgaben (1.2.2), sind Lernaufgaben ein zentrales Medium, mit dem neben Inhalten auch die kognitive Anregung der Lernenden gesteuert wird. In der Kommunikation zwischen Lehrkräften und Lernenden tragen sie damit wesentlich zur Art und Weise der Kompetenzprogression bei (vgl. 1.2.3). Dabei scheint offenbar folgender, empirisch belastbarer Zusammenhang auf der Hand zu liegen: Die Fähigkeiten zur Lösung kognitiv gehaltvoller Testaufgaben werden primär in einem kognitiv anregenden Unterricht erarbeitet (vgl. 1.2.3). Aktuelle innerdeutsche Studien zum kognitiven Anspruchsniveau von Lernaufgaben zeigen jedoch das weitgehende Fehlen von Aufgaben mit kognitiven Tätigkeiten auf mittlerem und auch auf hohem Komplexitätsniveau, die erst eine aktive Verknüpfung von Inhalten und Konzepten befördern würden (vgl. 1.3.2). Das erste Anliegen dieser Dissertation erweiternd, lässt dies nach den Gründen dieser Impuls- und Variationsarmut im Rahmen des Lernangebots fragen (vgl. 1.3.2). So vermutet Drüke-Noe mögliche Ursachen der vorwiegend kalkülorientierten Aufgabenkultur in lediglich vereinzelt in der Literatur zu findenden Hinweisen auf eine kriterial geleitete Konzeption von Lern- und Testaufgaben; dies lässt sie weitergehend auf fehlende Aus- und Fortbildungsinhalte mit einem derartigen Fokus schließen (2014, S. 249). Erste empirische Befunde legen bestätigend die Deutung nahe, dass *Lehrkräften* kognitive Aufgabenmerkmale und damit jene jenseits inhaltlicher nur wenig bewusst sind (vgl. Hillje 2012; Hospesova & Ticha 2010), obgleich das Kategoriensystem der Bildungsstandards als implementiert angenommen wird. Aktuelle Studien mit fächerübergreifenden Klassifikationen scheinen diese Vermutung zu bestätigen (Drüke-Noe et al. 2017), für das Fach Mathematik fehlen entsprechende Befunde annähernd gänzlich. Daraus resultiert das *zweite* Anliegen dieser Dissertation:

Kennen und verstehen Lehrkräfte kognitive Aufgabenmerkmale als Beschreibungsstrukturen von Mathematikaufgaben und inwiefern kann eine Lehrerfortbildung zu einer Steigerung dieser Kompetenz führen?

Dieses Forschungsdesiderat greift die abschließende Studie entlang folgender drei Fragen auf:

- *Welche Merkmale des kognitiven Anspruchs von Mathematikaufgaben kennen Lehrkräfte?*
- *Inwiefern sind Lehrkräfte in der Lage, die von ihnen genannten Merkmale des kognitiven Anspruchs hinsichtlich der kognitiven Komplexität gezielt zu variieren?*
- *Inwieweit wird das Wissen der Lehrkräfte um die Merkmale des kognitiven Anspruchs von Mathematikaufgaben erweitert und kann es flexibel angewendet werden?*

Auf eine Erfassung, Analyse und vor allem eine Steigerung der oft in der Literatur kritisierten kriterial geleiteten Aufgabenkompetenz von Lehrkräften zielend, zeigt die dazugehörige Studie, wie Lehrkräfte mittels einer videobasierten Fortbildung angeregt werden können, ihr kategoriales Wissen zum kognitiven Anspruch von Mathematikaufgaben zu erweitern und flexibel anzuwenden.

2 Publikationen

Zur Beantwortung der leitenden Fragen der Dissertation werden zwei Studien herangezogen, die sich auf die Konzepte der kognitiven Aktivierung, bzw. des Grades der kognitiven Aktivierung, des kognitiven Anspruchsniveaus, stützen. Sie zielen einerseits darauf ab, empirisch gesicherte Erkenntnisse über die Aufgabenmerkmale der polnischen Mittelschulprüfung sowie der ZP 10 zu gewinnen und in einem weiteren Schritt ggf. bisherige Konzeptionsmerkmale kritisch zu reflektieren (2.1, 2.2 und 2.3); andererseits soll mittels einer videobasierten Lehrerforbildung ein Beitrag dazu geleistet werden, dass Mathematiklehrkräfte das Anspruchsniveau von Aufgaben zunehmend als eine kriterial geleitete Fähigkeit auffassen und umsetzen (2.4).

2.1 Publikation I

Scheja, B. (2017). Kognitive Aktivierung durch Mathematikaufgaben zentraler Abschlussprüfungen – Eine Vergleichsanalyse der polnischen Mittelschulprüfung und der Zentralen Prüfung aus Nordrhein-Westfalen. *Journal für Mathematikdidaktik*, 38, 291 – 322.

Zusammenfassung

Die vorliegende Studie untersucht Mathematikaufgaben der im Zuge der Bildungssystemreform in Polen eingeführten Mittelschulprüfung im Hinblick auf ihr kognitives Aktivierungspotential. Dabei werden unter kognitivem Aktivierungspotential die Ausprägungen von curricularen, inhaltlichen und kognitiven Kategorien verstanden, die potentiell im Lösungsprozess aktiviert werden. Die Untersuchung beleuchtet hierzu einleitend den curricular-organisatorischen Rahmen der Mittelschulprüfung im Fach Mathematik, dessen Vorgaben in konzeptioneller Hinsicht für die Aufgabenentwicklung als grundlegend angesehen werden können. Danach wird zunächst das kognitive Aktivierungspotential aller Prüfungsaufgaben des Zeitraums zwischen 2002 und 2015 betrachtet. Anschließend wird der polnische Aufgabensatz des Zeitraumes 2007 – 2015 entsprechenden Aufgaben der Zentralen Prüfung ZP 10 des Landes Nordrhein-Westfalen gegenübergestellt. Es zeigt sich unter anderem, dass die in der polnischen Mittelschulprüfung eingesetzten Aufgaben insgesamt ein verhältnismäßig hohes Niveau im kognitiven Aktivierungspotential aufweisen. Hierzu tragen insbesondere sprachlogisch komplexe Aufgabenstellungen und mathematische Darstellungen bei, denen ebenfalls anspruchsvolle, da mehrschrittige Verarbeitungsprozesse nachgeordnet werden. Die Befunde des nordrheinwestfälisch-polnischen Vergleichs weisen auf zwei mit Einschränkungen vergleichbare Aufgabenprofile hin. So zeigt der polnische Aufgabensatz ein vergleichsweise ausgewogeneres Anforderungsspektrum, was primär durch eine im Hinblick auf Maß und Niveau unterschiedliche Berücksichtigung der Aufgabenkontexte (innermathematisch und außermathematisch) bedingt wird.

J Math Didakt (2017) 38:291–322
DOI 10.1007/s13138-017-0119-7

ORIGINALARBEIT/ORIGINAL ARTICLE

Kognitive Aktivierung durch Mathematikaufgaben zentraler Abschlussprüfungen

Eine Vergleichsanalyse der polnischen Mittelschulprüfung und der Zentralen Prüfung aus Nordrhein-Westfalen

Bruno Scheja

Eingegangen: 10. August 2016 / Angenommen: 19. April 2017 / Online publiziert: 26. Mai 2017

Zusammenfassung Die vorliegende Studie untersucht Mathematikaufgaben der im Zuge der Bildungsystemreform in Polen eingeführten Mittelschulprüfung im Hinblick auf ihr kognitives Aktivierungspotential. Dabei werden unter kognitivem Aktivierungspotential die Ausprägungen von curricularen, inhaltlichen und kognitiven Kategorien verstanden, die potentiell im Lösungsprozess aktiviert werden. Die Untersuchung beleuchtet hierzu einleitend den curricular-organisatorischen Rahmen der Mittelschulprüfung im Fach Mathematik, dessen Vorgaben in konzeptioneller Hinsicht für die Aufgabenentwicklung als grundlegend angesehen werden können. Danach wird zunächst das kognitive Aktivierungspotential aller Prüfungsaufgaben des Zeitraums zwischen 2002 und 2015 betrachtet. Anschließend wird der polnische Aufgabensatz des Zeitraumes 2007–2015 entsprechenden Aufgaben der Zentralen Prüfung ZP 10 des Landes Nordrhein-Westfalen gegenübergestellt. Es zeigt sich unter anderem, dass die in der polnischen Mittelschulprüfung eingesetzten Aufgaben insgesamt ein verhältnismäßig hohes Niveau im kognitiven Aktivierungspotential aufweisen. Hierzu tragen insbesondere sprachlogisch komplexe Aufgabenstellungen und mathematische Darstellungen bei, denen ebenfalls anspruchsvolle, da mehrschrittige Verarbeitungsprozesse nachgeordnet werden. Die Befunde des nordrhein-westfälisch-polnischen Vergleichs weisen auf zwei mit Einschränkungen vergleichbare Aufgabenprofile hin. So zeigt der polnische Aufgabensatz ein vergleichsweise ausgewogeneres Anforderungsspektrum, was primär durch eine im Hinblick auf Maß und Niveau unterschiedliche Berücksichtigung der Aufgabenkontexte (innermathematisch und außermathematisch) bedingt wird.

Mein besonderer Dank gilt den Gutachterinnen und Gutachtern des Artikels.

B. Scheja (✉)
St.-Ursula-Gymnasium Brühl, Kaiserstr. 22, 50321 Brühl, Deutschland
E-Mail: schejab@uni-koeln.de

Schlüsselwörter Sekundarstufe · Kognitive Aktivierung · Zentrale Prüfungen · Empirische Bildungsforschung

Mathematical Subject Classification B13 · C33 · D13

Cognitive activation througt mathematics tasks of centralised final examinations

A comperative analysis of the Polish middle School examination and the centralised examination administered in North-Rhine Westphalia

Abstract This study analyses mathematical tasks of the Middle School test, which were introduced as a part of the education reform in Poland, with regard to their quality of cognitively activating students. The term quality of cognitive activation, here, denotes the different possible manifestations of categories such as curricular standards, subject-specific contents, and cognitive abilities, which are potentially activated in the course of solving the respective task at hand. In its introductory part, this study examines the curricular and organisational framework of the Middle School test in mathematics, whose standards – on a conceptual level – are fundamental for the development of tasks. The introduction is then followed by the examination of the complete set of tasks employed between 2002 and 2015. In a final step, the set of tasks employed in Poland between 2007 and 2015 is compared to the respective tasks used for the centralised test that is part of tenth grade (ZP 10) in North Rhine-Westphalia. First, it becomes evident that the tasks that are part of the Middle School test in Poland exhibit a comparatively high potential in terms of their cognitive requirements. This is especially due to task designs characterised by complex linguistic and logical constructions as well as mathematical representations that require demanding cognitive processing operations. Within the limits of the aforementioned period, the chronological perspective shows that the tasks that are part of the reformed Polish test concept are consistently more balanced and cognitively more challenging than those employed in the centralised tests in NRW. The findings resulting from the comparison of the Polish and North Rhine-Westphalian tasks suggest two task profiles, which are primarily due to a difference in emphasis put on activation within the framework of the overall task design.

Keywords Secondary education · Cognitive activation · Centralised tests · International educational research

1 Einleitung

Zu den wesentlichen Elementen des in den vergangenen drei Jahrzehnten international vermehrt vollzogenen Wandels von Bildungssystemen zählen unter anderem zentrale (Abschluss-)Prüfungen, die zumeist schulstufenabschließend durchgeführt werden (vgl. Krüger 2015; Kühn 2013). In den darin eingesetzten, in erster Linie fächergebundenen Aufgabensets einerseits und den zugrundeliegenden Standards andererseits manifestieren sich – so die verbreitete Annahme – bildungspolitische

Wirkungserwartungen im Hinblick auf Lernprozesse in den einzelnen Fächern (vgl. Kühn und Drüke-Noe 2013). Eines der Länder, das infolge seiner Bildungsreform im Jahre 1999 auf der Grundlage dieses Wirkungsgefüges zentral die Leistungen seiner Schülerinnen und Schüler überprüft, ist Polen. Tatsächlich weisen die dort vor und nach der Reformumsetzung erfassten Mathematikleistungen positive Veränderungen auf (vgl. Bialecki und Haman 2001; MEN 2013). Vor dem Hintergrund des beobachteten Leistungszuwachses im Rahmen der PISA-Studien (PISA 2000: 470 Punkte; PISA 2012: 518 Punkte; ebd.) ist es aus fachdidaktischer Sicht von Interesse, den Fokus auf qualitätserfassende und qualitätsbeeinflussende Instrumente, wie die am Ende der neunten Klasse landesweit verpflichtende Mittelschulprüfung zu legen. Von ihr kann angenommen werden, gerade für die Sekundarstufe I eine der zentralen Einflussgrößen für den polnischen Mathematikunterricht zu sein. In den wenigen bisher veröffentlichten Studien zu diesem 2002 eingeführten Steuerungsinstrument liegt der Fokus überwiegend auf systemischen, also organisatorisch-administrativen Betrachtungen (vgl. Kupisiewicz 2006; Zahorska 2002; Scheja 2007), wohingegen innerfachliche Aspekte, wie beispielsweise das mit der Reform einhergehende, ebenfalls gewandelte Verständnis von Denk- und Arbeitsweisen in der Schulmathematik, vergleichsweise wenig Berücksichtigung finden (vgl. Karpinski et al. 2013; Konarzewski 2008). Um Fragen dieses Forschungsfeldes näher betrachten zu können, sind im Wesentlichen zwei Ansätze denkbar: So kann zum einen das innerfachliche, intendierte Konzept anhand von curricularen Vorgaben, wie zum Beispiel den inhaltlichen Standards und Richtlinien erfasst und beurteilt werden (vgl. Konarzewski 2004; Kühn und Drüke-Noe 2013; Schupp 1982). Da aktuelle empirische Befunde hierzu jedoch großflächige Diskrepanzen zwischen Wirkungserfahrungen und Wirkungserwartungen im Fach Mathematik aufzeigen, scheint der Ansatz der Erfassung einer curricular verankerten Mathematikauffassung als maßgeblicher und ausschließlicher Impulsgeber für den Unterricht offenbar nur eingeschränkt geeignet (vgl. Neubrand und Neubrand 2010; Kühn und Drüke-Noe 2013). In der aktuellen Literatur geht man einem zweiten Ansatz folgend zunehmend davon aus, dass formale, inhaltliche und prozessartige Konzeptionsmerkmale von zentralen Prüfungsaufgaben eine steuernde Wirkung auf das Lernarrangement besitzen (vgl. Büchter und Pallack 2012; Konarzewski 2008; Maier et al. 2011), wobei das Wirkungsausmaß bei den Tests, die primär die Funktion „Individualbewertung“ (vgl. Drüke-Noe 2014, S. 150 ff; Maier et al. 2011) ausweisen, am stärksten zu sein scheint. In Anbetracht dieser, offenbar weitflächig verbreiteten Form der impliziten Normierung von Lehr- und Lernprozessen im Hinblick auf Inhalte, Methoden und Aufgabenformate gewinnt die Frage der schwierigkeitsbestimmenden Aufgabenmerkmale und deren Orchestrierung auch in Bezug auf Testaufgaben zunehmend eine tragende Bedeutung (vgl. Jordan et al. 2008; Büchter und Pallack 2012; Neubrand et al. 2002). Ein umfassendes Instrument zur Erfassung dieses Konstrukts, aufgefasst als inhärentes Potential der Aufgaben zur gezielten kognitiven Aktivierung, liefert – auf den kognitionstheoretisch basierten Vorarbeiten von Neubrand et al. (2002) gründend – das Kategoriesystem, das im Rahmen des COACTIV-Projekts entwickelt und anschließend von Drüke-Noe ergänzt wurde (vgl. Jordan et al. 2006; Drüke-Noe 2014).

Die Anwendung dieses Klassifikationsschemas ermöglicht auf der Grundlage der rationalen Aufgabenanalyse (vgl. Resnick und Ford 1981) eine schrittweise Erarbeitung des zentralen Anliegens dieser Arbeit: Das kognitive Aktivierungspotential durch Mathematikaufgaben der polnischen Mittelschulprüfung zu erfassen, zu analysieren und zu bewerten. Dieses wird dabei wie folgt ausdifferenziert: (a) Wie verteilen sich die Prüfungsaufgaben auf verschiedene Aufgabenklassen und welche Anforderungsniveaus lassen sich im Bezug auf curriculare und kognitive Erfassungskategorien identifizieren? (b) Lassen sich in einem Vergleich mit Zentralen Abschlussprüfungen des Landes Nordrhein-Westfalens (ZP 10) Unterschiede hinsichtlich der Verteilungen auf schwierigkeitsgenerierende Kategorien ausmachen? Mit diesen Anliegen greift der Autor das einleitend skizzierte Forschungsdesiderat auf, um vornehmlich der allgemein- und fachdidaktischen Diskussion, die derzeit in Deutschland um die spannungsgeladenen Themen der Aufgabenkultur und speziell Testaufgaben geführt wird (Büchter und Pallack 2012; Drüke-Noe 2014; Krüger 2015; Kühn und Drüke-Noe 2013), ein umfassendes inhaltlich-konzeptionelles Profil von Testaufgaben eines ebenfalls kürzlich reformierten Bildungsystemes darzulegen und schließlich mit dem von NRW zu vergleichen. Die ebenfalls am Ende des ISCED 2 Levels[1] durchgeführte ZP 10 eignet sich für eine derartige Gegenüberstellung in besonderem Maße. So ist der Anteil der ZP-10-Prüflinge an Prüflingen, die deutschlandweit an *Abschlussprüfungen zum Erwerb des Mittleren Schulabschlusses* (kurz MSA) teilnehmen, in Nordrhein-Westfalen im Ländervergleich am höchsten. Darüber hinaus weisen beide Prüfungsinstrumente eine hohe funktionale Äquivalenz (Selektion und Vergleichbarkeit) auf und der Umgang mit den Prüfungsergebnissen seitens der Bildungspolitik zeigt zahlreiche Parallelen (Büchter und Pallack 2012; CKE 2011).

Der Beitrag beginnt mit einer Beschreibung der Reform des polnischen Bildungssystems (2.1) und des Vergleichs der Mathematikleistung polnischer Schülerinnen und Schüler vor und nach der Reform anhand von PISA-Ergebnissen der Jahre 2000 und 2012 (2.2). Anschließend werden die curricular-organisatorischen Grundlagen der Mittelschulprüfung zum Zeitpunkt ihrer Einführung (2002) und der einschneidenden Reform im Jahre 2012 dargelegt (Kap. 3). Inwiefern der Begriff der kognitiven Aktivierung zur Analyse der im Fokus stehenden Aufgaben geeignet ist, wird in Kap. 4 dargelegt und das zugehörige Kategorienschema in Kap. 5 vorgestellt. In Kap. 6 erfolgt dann die Untersuchung des vollständigen polnischen Aufgabensatzes des Testzeitraums 2002–2015 (6.1) und anschließend die vergleichende Untersuchung der Aufgabensätze der polnischen Mittelschulprüfung und der zentralen Abschlussprüfung des Landes Nordrhein-Westfalen im Zeitraum 2007–2015 (6.2). Von besonderem Interesse ist es hierbei, in welcher Weise sich die eingesetzten Prüfungsaufgaben unterscheiden, um gegebenenfalls auf dieser Grundlage Testprofile beider Länder herausarbeiten zu können.

[1] International Standard Clasiffication of Education, lower Secondary (vgl. UNESCO 2011).

2 Bildungssystemreform in Polen

2.1 Notwendigkeit der Bildungssystemreform

Das schlagartig einsetzende Freiwerden bildungspolitischer Kräfte im Zeitraum der frühen 1990er-Jahre führte in Polen zu einem radikalen Verwurf tief verankerter kommunistischer Trends und damit einer Abkehr von ideologiegeladenen, anachronistischen Ideen innerhalb des Schulsystems.[2] Die erste Zuwendung sollte vor allem Erfordernissen gerecht werden, die an ein Bildungssystem im Wirkungsfeld von Globalität gestellt werden. Die im Zuge des angestoßenen Wandels rasch implementierten Neuordnungen beurteilten zahlreiche Bildungsbeteiligte zum einen als unumgänglich, zum anderen jedoch als unzeitgemäß und darüber hinaus völlig unzureichend (vgl. Banach 1999, S. 59 ff; Zahorska 2002, S. 113 f). So zeigte sich beispielsweise, dass die Leitideen des Wandels leistungsstarker Länder, die sich seit den 1980er-Jahren zunehmend unter den Schlagwörtern Qualitätssteigerung, Evaluation, Wettbewerb oder auch der Effektivität eines Bildungssystems zusammenfassen lassen (vgl. van Ackeren 2003), dabei in den initiierten Umgestaltungen weitestgehend ohne Berücksichtigung blieben. Fasst man die vor allem ab 1997 seitens der Bildungsbeteiligten rasch anwachsende Kritik am Systemzustand zu Problemfeldern zusammen, so zählten dazu im Wesentlichen

1. das allgemein thematisierte niedrige Leistungsniveau der polnischen Lernenden,
2. eine ungleiche Verteilung von Bildungschancen und damit die Festigung existierender sozialer Schichtung,
3. die allgemein geringe Effektivität des Bildungswesens sowie
4. eine im Hinblick auf den demografischen Wandel innerhalb der Schülerschaft ungeeignete Schulstruktur (Zahorska 2002, S. 278).

Alle Problemfelder eint dabei eine – zum Teil durch entsprechende Studien bestätigte – Kluft zwischen der wahrnehmbaren Wirkung von Inputs bzw. Prozessen und den zu diesem Zeitpunkt bereits rege diskutierten Wirkungserwartung an diese (vgl. Banach 1999; Białecki und Haman 2001). Der wachsende Druck auf die Bildungspolitik führte mit der Veröffentlichung des Dokuments „Reforma systemu edukacji – projekt“ (Reform des Bildungssystems – Projekt) (MEN 1998a) zur Einleitung einer umfassenden Bildungssystemreform, mit deren Umsetzung 1999 begonnen wurde.

Um nun einerseits die geringe Systemeffektivität dieses Zeitraumes im Hinblick auf das erste Problemfeld und andererseits die leistungsbezogene Wirkung der Bildungssystemreform aufzuzeigen, werden im Folgenden die Leistungen der polnischen Lerndenden im Fach Mathematik bei der PISA-Studie skizziert (2000 und 2012), bevor anschließend, nach einer groben Beschreibung der Reformelemente, der Fokus auf „egzamin gimnazjalny“, die Mittelschulprüfung, und somit einen Aspekt der Reform gerichtet wird.

[2] Als innerfachlicher Niederschlag dieser können beispielsweise die Ergebnisse der Lehrplananalysen von Molenda und Piesyk (1993) betrachtet werden. Diese zeigen, dass die zu erwerbenden Fähigkeiten und Fertigkeiten im Mathematikunterricht der 1960er-1980er-Jahre annähernd ausschließlich innermathematisch verortet wurden.

2.2 Zentrale Ergebnisse im Bereich der mathematischen Grundbildung als Maß für den Leistungsstand polnischer Schülerinnen und Schüler

2.2.1 Ergebnisse der Studie PISA 2000: Leistungsstand vor der Bildungssystemreform

Um die Mathematikleistungen von Lernenden des polnischen Bildungssystems vor der Reform zu erfassen, wird hier mit PISA 2000 auf ein Instrument zurückgegriffen, dessen Alterskohorte überwiegend 1991 eingeschult wurde. Damit durchliefen die daran teilnehmenden Lernenden „alte“ Lern- und Förderstrukturen (vgl. Kupisiewicz 2006) und die Wirkung der seit September 1999 sukzessive umgesetzten Bildungssystemreform kann als vernachlässigbar betrachtet werden.

Die PISA-Studie 2000, die zugleich die erste internationale Studie zur Leistungsmessung war, an der Polen teilnahm, zielt im Wesentlichen darauf, den teilnehmenden Ländern Daten über die Funktions- und Leistungsfähigkeit ihrer Bildungssysteme zu liefern. Der mathematische Teil des PISA-Tests untersucht dabei, in welchem Maße mathematisches Wissen funktional, flexibel und mit Einsicht zur Bearbeitung vielfältiger, kontextbezogener Probleme eingesetzt werden kann (Klieme et al. 2001, S. 141). Inwieweit die Jugendlichen, die das „an demokratische Anforderungen nicht angepasste polnische Bildungssystem“ (Banach 1999, S. 7) durchlaufen haben, derartige Anforderungen bewältigen konnten, kann anhand der Leistungsergebnisse im Bereich mathematischer Grundbildung aufgezeigt werden:

1. Der Mittelwert polnischer Schülerleistungen lag mit 470 Punkten signifikant unterhalb des OECD-Mittelwerts von 500 Punkten. Deutschland weist hier mit 490 Punkten ein vergleichsweise höheres Leistungsniveau auf.
2. Der Anteil schwacher und schwächster Schülerinnen und Schüler, die die „Kompetenzstufe I und darunter“ nicht überschreiten, besetzen etwa 30 % der Lernenden. Damit lag es unterhalb des, als besonders schlecht eingeschätzten, deutschen Ergebnisses (unter 25 %).
3. Lediglich 10 % der polnischen Lernenden besetzen die Kompetenzstufe IV und V, wovon etwa nur jeder zehnte der zuletzt genannten Stufe der „komplexen Modellierung und des innermathematischen Argumentierens“ zuzuordnen ist.
4. Im Vergleich zum OECD-Durchschnitt streuen die Leistungen überdurchschnittlich. Folglich ist die „Fähigkeit“ des „alten“ polnischen Bildungssystems mathematische Kompetenz zu homogenisieren gering (vgl. Białecki und Haman 2001).

Bei der in PISA 2012 getesteten Schülerpopulation handelt es sich annähernd ausschließlich um Jugendliche, die bereits gänzlich das reformierte Schulsystem Polens durchlaufen haben und zum Zeitpunkt der Erhebung unmittelbar vor dem Abschluss der Mittelschule standen (vgl. MEN 2013, S. 8 ff). Das Leistungsbild auf dem Gebiet der mathematischen Grundbildung weist dabei folgende, an PISA 2000 angelehnte Merkmale auf:

1. Der Mittelwert polnischer Schülerleistungen im Fach Mathematik lag mit 518 Punkten signifikant oberhalb des OECD-Durchschnitts (494 Punkte). Mit dieser unter den Teilnehmern seit 2000 stärksten Leistungszunahme überholt Polen

Deutschland (514) und liegt nun beispielsweise gleichauf mit Ländern wie Finnland.
2. Den Anteil schwacher und schwächster Schülerinnen und Schüler, die die „Kompetenzstufe I und darunter“ nicht überschreiten, besetzen 14,4 % der Lernenden. Damit hat sich der Anteil im Vergleich zu 2000 etwa halbiert und liegt nun sowohl unterhalb des OECD-Mittelwertes (23 %) als auch des deutschen Ergebnisses (17,7 %).
3. Insgesamt 16,7 % der polnischen Lernenden besetzen die beiden höchsten Kompetenzstufen V und VI. Damit scheinen die polnischen Schülerinnen und Schüler nun verhältnismäßig gut auf die Bewältgung komlexer Anforderungen innerhalb der mathematischen Grundbildung vorbereitet zu sein (OECD-Mittelwert: 12,6 %; Deutschland: 17,5 %).
4. Die Streuung der Kompetenzverteilung ist im Vergleich zum OECD-Durchschnitt kleiner, ohne jedoch signifikant nach unten abzuweichen. Demnach scheint das reformierte Bildungssystem Leistungen im Bereich der mathematischen Grundbildung befriedigender homogenisieren zu können (vgl. ebd.).

3 Die Mittelschulprüfung als Element der Bildungssystemreform

Die Verbesserung der in 2.1 angeführten Problemfelder kennzeichneten zugleich im Wesentlichen die Ziele der seit 1999 umgesetzten Reform (vgl. Zahorska 2002). Zu den zentralen Elementen der Restrukturierungen zählten

1. die Umgestaltung der Schulstruktur, d. h. Einführung der Grundschule (Klassen 1–6), der Mittelschule (Klassen 7–9) sowie der nach Neigung differenzierenden Schulformen der ISCED 3 Stufe (Klassen 10–12),
2. die Einführung eines Systems externer Prüfungen,
3. eine neue, dezentrale Verwaltungs- und Finanzierungsstruktur des Bildungswesens,
4. die Einführung von Prüfungsstandards, Rahmenplänen und des Lehrplanpluralismus auf Einzelschulebene sowie
5. die grundlegend umgestaltete, leistungsorientierte Organisation der Lehrerbildung im Zeitraum zwischen 2000 und 2004.

Wie Scheja (2007, S. 252 ff) gezeigt hat, war gerade die Vielschichtigkeit der Reformelemente eine tragende Voraussetzung dafür, das Instrument der Evaluation als einen zentralen qualitätssteigernden und zugleich alle Reformelemente in ihrer Wirkung einenden Lenkmechanismus zu installieren. Es findet Anwendung im Hinblick auf die Leistungen der Schülerschaft, die Kompetenzen der Lehrkräfte oder die Qualität der Schulen. Insofern wird Qualität aller Ebenen des polnischen Bildungssystems (Mikro-, Meso- und Systemebene) nun mithilfe von Feedbackschleifen *im Prozess* beschrieben. Ähnlich dem deutschen Bildungssystem „nach PISA“ wurde damit eine „Umsteuerung“ vollzogen.

In welchem Maße das Instrument der externen Leistungsmessungen in den letzten drei Jahrzehnten im europäischen Raum an Bedeutung gewonnen haben, belegt vor allem Krüger (2015) am Beispiel ausgewählter PISA-Teilnehmerländer. Dass auch Polens Bildungspolitik der Implementierung dieses Instruments einen hohen Stellenwert beigemessen hat, zeigt seine Explizierung als eines der fünf Reformelemente. Kupisiewicz (2006, S. 74) spricht in organisatorisch-funktionalem Kontext des Instuments von der Einführung eines „Systems externer Leistungsbewertung", d. h., er fasst den Reformpunkt als einen organisierten, aus Einzelelementen bestehenden Komplex, der sowohl institutionelle als auch konzeptionelle Reformaspekte in sich vereinheitlicht. Die Bezeichnung gründet dabei auf der Aufgliederung des Systems externer Leistungsmessung in folgende Elemente:

1. Prüfungskommissionen, die externe Prüfungen konzipieren, organisieren und zum Teil auch durchführen.
2. Externe Leistungsmessungen am Ende jeder Schulstufe, d. h. der Grundschule, der Mittelschule sowie auch aller darauf aufbauenden Schulformen.
3. Standards als Vorgaben für externe Prüfungen.

Das im Zuge des zweiten Reformpunkts eingeführte System externer Leistungsmessungen setzt sich im Einzelnen aus

- dem Grundschultest (polnisch: „sprawdzian", Durchführung: Ende der sechsten Klasse),
- der Mittelschulprüfung (polnisch: „egzamin gimnazjalny", Durchführung: Ende der neunten Klasse),
- der Berufsprüfung an Berufskollegs (polnisch: „egzamin potwierdzający kwalifikacje zawodowe") sowie
- der Abiturprüfung an (Berufs-)Lyceen („egzamin maturalny") zusammen.

Am Ende der dritten Mittelschulklasse (neunte Klassenstufe) und damit kurz vor dem Ende der Pflichtschulzeit führen die zuständigen *Regionalen Prüfungskommissionen* (Okregowa Komisja Egzaminacyjna, Abk.: OKE) in Zusammenarbeit mit den Einzelschulen die zweite, zentral konzipierte Laufbahnprüfung durch. Dies ist die landesweit einheitliche Mittelschulprüfung, an der seit der erstmaligen Durchführung im Jahre 2002 jährlich alle polnischen Jugendlichen im Alter zwischen 15 und 16 Jahren verpflichtend[3] teilnehmen.

Der erste Testteil (gleichzeitig erster Prüfungstag) überprüft das Leistungsniveau aus dem Bereich humanistischer Fächer („część humanistyczna"), zu denen zum einen Polnisch (Bearbeitungsdauer seit 2012: 90 min) und zum anderen Geschichte und Gesellschaftslehre (Bearbeitungsdauer insg. 60 min) zählen. Am zweiten Prüfungstag wird der Fokus auf mathematisch-naturwissenschaftliche („część matematyczno-przyrodnicza") Fächer wie Mathematik, Biologie, Geografie, Physik, Astronomie und Chemie gelegt. Am dritten Prüfungstag werden die Lernenden schließlich in einer modernen Fremdspache ihrer Wahl getestet.

Für die Untersuchung der zentralen Leitfragen der in diesem Artikel präsentierten Studie wird nun für den Prüfungsteil Mathematik offen gelegt,

[3] Nimmt eine Schülerin/ein Schüler an der Mittelschulprüfung nicht teil, so wird ihm – unabhängig seiner bisherigen Noten – der Abschluss nicht bescheinigt. In diesem Fall muss die dritte Klasse der Mittelschule wiederholt werden.

- welche inhaltliche Ausrichtung die Prüfung hat,
- welches Bewertungskonzept der Prüfung zugrundegelegt wird sowie
- in welcher Art und Weise Aspekte dieser beiden Punkte transparent gemacht werden.

3.1 Curriculare und konzeptionelle Grundlagen der Mittelschulprüfung Mathematik

Im Zeitraum zwischen der Einführung der Mittelschulprüfung im Jahre 2002 und ihrer Neuausrichtung 2012 umfassten die landesweit einheitlichen Mathematikaufgaben[4] einen Anteil zwischen 40 % und 50 % der Aufgaben im mathematisch-naturwissenschaftlichen Testteil (Bearbeitungsdauer: 120 min, insgesamt ca. 35 math.-naturwiss. Testaufgaben).

Die inhaltlichen Grundlagen der Aufgaben bildeten als landesweit geltende Vorgaben der Rahmenplan „Podstawa Programowa" sowie die „Prüfungsstandards mathematisch-naturwissenschaftlicher Fächer" (vgl. CKE 2007). Im Sinne eines Erwartungshorizonts werden in den Prüfungsstandards fächerübergreifende „Gebiete der Standards" aufgeführt (vgl. Tab. 1). Seitens der Zentralen Prüfungskommission (Abk.: CKE) wird beansprucht, dass die Prüfungsstandards durch die Prüfungsaufgaben angemessen operationalisiert werden (vgl. MEN 1998b).

Die landesweit durchgeführte Mittelschulprüfung wurde 2012 vornehmlich aufgrund erheblicher konzeptioneller Unzulänglichkeiten hinsichtlich des Zusammenspiels zwischen dem Rahmenplan und der in Tab. 1 dargestellten Prüfungsstandards reformiert (vgl. Konarzewski 2004). So wird nun das Fach Mathematik *separat* abgeprüft. Die Prüfung besteht aus 23 Aufgaben, umfasst dabei eine Dauer von 90 min und ist ein Bestandteil des nun zweigeteilten mathematisch-naturwissenschaftlichen Prüfungstages (bestehend aus „Mathematik" (90 min) und „Naturwissenschaften" (60 min)). Die inhaltlichen Grundlagen bildet nun ausschließlich der landesweit geltende, im Jahre 2008 überarbeitete Rahmenplan „Podstawa Programowa", in dem

1. fünf prozessartige Ziele unter den Begriffen „Gebrauch und die Erzeugung von Informationen" (1.1), „Nutzung und Interpretation von Repräsentationen" (1.2), „mathematisches Modellieren" (1.3), „Nutzung und Erzeugung von Strategien" (1.4) und „Schlussfolgerungen und Argumentation" (1.5) sowie
2. detaillierte *performance* Standards[5] in elf Inhaltsfeldern („Rationale Zahlen", „Potenzen", „Wurzeln", „Prozente", „Algebraische Ausdrücke", „Gleichungen", „Funktionsgraphen", „Beschreibende Statistik und Einführung der Wahrscheinlichkeitsrechnung", „Ebene Figuren" sowie „Körper") der dreijährigen Schulform benannt werden (vgl. MEN 2009, S. 35 ff).[6]

[4] Im Gegensatz dazu werden in Systemmonitoringstudien vergleichsweise viele Aufgaben eingesetzt, die auf der Grundlage eines Multi-Matrix-Designs gemeinsam ausgewertet werden.

[5] Die zugrunde gelegten Niveauanforderungen entsprechen Regelanforderungen und können somit als *Regelstandards* identifiziert werden.

[6] Hierbei geht man zugleich von der Annahme aus, dass die Inhalte des Grundschulunterrichts als gesichert gelten und somit auch Gegenstand der Mittelschulprüfung sein können.

Tab. 1 Prüfungsstandards des Zeitraums 2002–2011 (s. CKE 2007, Übersetzung durch Scheja)

Gebiet der Standards	Standards
I. Die korrekte Anwendung von Begriffen, Formeln und Prozeduren mathematisch-naturwissenschaftlicher Fächer, die für den Alltag und spätere Weiterbildung erforderlich sind	Der Schüler – wendet mathematisch-naturwissenschaftliche Begriffe und Verfahren an, – führt Berechnungen in unterschiedlichen praktischen Situationen durch, – nutzt die Eigenschaften von Figuren
II. Das Suchen und Anwenden von Informationen	Der Schüler – entnimmt Informationen, die in Form einer Tabelle, eines Textes, einer Karte oder eines Graphen dargestellt wurden, – nutzt Informationen, indem er sie selektiert, vergleicht, analysiert, umwandelt und interpretiert
III. Aufzeigen und Beschreiben von Tatsachen, Zusammenhängen und Abhängigkeiten, insbesondere in ursächlicher, funktionaler, räumlicher und zeitlicher Hinsicht	Der Schüler – zeigt die Richtigkeit von Prozessen auf, – gebraucht algebraische Ausdrücke und die Sprache der Symbole, – nutzt Funktionen in Form einer Gleichung, einer Tabelle und eines Graphen, – nutzt fachübergreifendes Wissen zur Erklärung von naturwissenschaftlichen Phänomenen
IV. Die Anwendung integralen Wissens und Fähigkeiten im Prozess der Problemlösung	Der Schüler – nutzt Techniken kreativer Problemlösung, – analysiert eine problematische Situation, – wählt ein Modell der problematischen Situation, – wählt und realisiert einen Lösungsweg, – bewertet, interpretiert und stellt die Ergebnisse dar

Ein Vergleich der prozessartigen Ziele des aktuellen Rahmenplans mit den allgemeinen mathematischen Kompetenzen der *Bildungsstandards für den Mittleren Schulabschluss* zeigt eine hohe Übereinstimmung (Ausnahme: K5 findet keine Entsprechung), wobei in beiden Fällen kein ausgearbeiteter Bezug zu spezifischen Lerntheorien vorfindbar ist (vgl. MEN 2009; Reiss 2004). In der hohen Deckungsgleichheit der Kompetenzen wird im neuen Rahmenplan zugleich der annähernd gleichverteilte Niederschlag der drei Grunderfahrungen eines allgemeinbildenden Mathematikunterrichts nach Winter (1995) erkennbar. Deren intendierte Realisierung in den prozessartigen Zielen kann – wenngleich nicht trennscharf – durch folgende Zuordnungen aufgezeigt werden: Strukturorientierung (Ziele 1.2, 1.3, 1.4), Anwendungsorientierung (Ziele 1.1, 1.2, 1.4), Problemorientierung (Ziele 1.4, 1.5).

Demgegenüber werden in den Prüfungsstandards der Jahre 2002–2011 ausschließlich die Grunderfahrungen der Anwendungs- und Problemorientierung als zentrale Erträge der Bildungsprozesse aufgefasst. Die Ausrichtung der verlangten Fähigkeiten in den Prüfungsstandards von 2002–2011 korrespondiert dabei zugleich mit dem Ziel der CKE: Im Unterschied zur Aufgabenkonzeption vor 2012 soll nun ein kleinerer Anteil an Aufgaben einbezogen werden, welche „die Kenntnis und den Umgang mit Algorithmen in ihrer typischen Anwendung verlangen" (CKE 2010, S. 61), wohingegen Problemstellungen, in denen das Verständnis mathematischer Begriffe (polnisch: „rozumienie pojęć matematycznych") (ebd.) sowie die Fähigkeit zur geeigneten Auswahl von mathematischen Lösungsstrategien in untypischen Problemsituationen vermehrt verlangt werden sollen (polnisch: „umiejętności dobierania własnych strategii matematycznych do nietypowych warunków“) (ebd.).

Die beiden Anforderungsklassen – prozessartige Ziele und Standards der Inhaltsfelder – des neu konzipierten Rahmenplans haben den Charakter von Normierungs- und zugleich Vergleichsmaßstäben. Sie dienen der innerfachlichen Schul- und Unterrichtsentwicklung der Mikro- und Mesoebene, indem sie den Bildungsauftrag, den die Mittelschule im Fach Mathematik zu erfüllen hat, auf zwei unterschiedlichen Abstraktionsstufen benennen (*Orientierungsfunktion*). Die zweite Funktion der Anforderungsvorgaben besteht darin, der Zentralen Prüfungskommission – analog zu den Prüfungsstandards 2002–2011 – eine Grundlage zur jährlichen Entwicklung der Testaufgaben zu liefern. Zielsetzung der damit betrauten CKE-Gremien ist hierbei, eine Erfassung „der Leistung (der Lernenden) in Bezug auf die Standards für Anforderungen“ (CKE 2011, S. 5) sicherzustellen.

Wie dieser Schöpfungsprozess konkret gestaltet wird, d. h. welche konzeptionellen Aspekte beispielsweise im Hinblick auf Existenz von Niveauabstufungen oder gar die Verortung dieser im Bezug auf die definierten Normstandards dabei einfließen, wird zum Missfallen einzelner Bildungsbeteiligter nicht in die Öffentlichkeit getragen (vgl. Konarzewski 2004, S. 126 ff). So zeigt Konarzewski in seiner Interviewstudie zum neuen Rahmenplan, dass die verantwortlichen Mitarbeiterinnen und Mitarbeiter der CKE auf Fragen nach der Konzeption der Testaufgaben stets mit „esoterischer“ (ebd. S. 127) Verschwiegenheit reagieren, indem sie sich auf das Dienstgeheimnis berufen. Da die strikte Geheimhaltung ein zentrales Prinzip bei der Entwicklung von Prüfungsaufgaben ist, werden die eingesetzten Aufgaben analog zur ZP 10 und anders als bei Systemmonitoringstudien vor ihrem Einsatz nicht empirisch geprüft (vgl. Büchter und Pallack 2012; Konarzewski 2004).

Während der Durchführung der Mittelschulprüfung sind folgende Hilfsmittel zugelassen: Füllhalter mit schwarzer Tinte (verpflichtend), Lineal, Bleistift, Radiergummi, Geodreieck und ein Zirkel. Nicht zugelassene Hilfsmittel sind hingegen Taschenrechner, Formelsammlung, Wörterbücher oder Smartphones, was im Sinne der nordrhein-westfälischen Vorgaben der *Hilfsmittelfreiheit* enspricht.

3.2 Bewertungskonzept der Mittelschulprüfung Mathematik

Der von der CKE beschriebene Bewertungsprozess der Aufgaben kann in schematischer Form wie folgt gegliedert werden (vgl. CKE 2010):

Aufgabenbearbeitung → Auswertungshorizont → Rohpunkte
→ Prozent-und Perzentilangabe

Die Bewertung der Mittelschulprüfung erfolgt durch externe Prüferinnen und Prüfer der Regionalen Prüfungskommissionen (OKE). Die Testaufgaben werden im Hinblick auf ihre Merkmale offiziell in „offene“ und „geschlossene“ Formate unterteilt. Bei den geschlossenen Aufgaben des Formats Multiple-Choice sieht der Korrekturbogen im Sinne einer dichotomen Auswertung lediglich die Unterscheidung zwischen einem vorgegebenen, richtigen (stets 1 Rohpunkt) und einem falschen Zielzustand vor (0 Punkte) (vgl. Greefrath 2004, S. 17 f; CKE 2010, S. 61), während für offene Aufgaben (hier zumeist charakterisierbar durch mehrdeutige Transformationen und unklare Zielzustände) zum einen höhere Punktezahlen (zwischen 2 und 5 Rohpunkten) und zum anderen bis zu sechs unterschiedliche Lösungsniveaus („Poziom rozwiazania“) einbezogen werden können (vgl. CKE 2010, S. 61). Die offenen Aufgaben, die sich an tradierte Bewertungsformate aus der Schulpraxis orientieren, machen im betrachteten Zeitraum einen Anteil von 7 bis 13 % aller Aufgaben aus. Ähnlich wie bei der ZP 10 folgt die getroffene Gewichtung bei den Lösungsniveaus keinem etablierten didaktischen Konzept (vgl. Konarzewski 2004; Büchter und Pallack 2012). So sind die Lösungsniveaus offener Aufgaben zunächst allgemein definiert (*Lösungsniveau 0*: „Kein Fortschritt in der Lösung“; *Lösungsniveau 1*: „Unwesentlicher, jedoch notwendiger Fortschritt auf dem Weg zur Lösung“; (...); *Lösungsniveau 5*: „Wesentliche Schwierigkeiten der Aufgabe wurden fehlerfrei überwunden, weitere Teile der Lösung enthalten jedoch Fehler (Rechenfehler u. ä.)“) (CKE 2010, S. 61) und erfahren in den Erwartungshorizonten der einzelnen Prüfungsaufgaben eine Konkretisierung.

Den individuell erzielten Rohpunktwerten werden schließlich Prozent- wie auch Perzentilwerte zugeordnet, die mit dem Abschlusszeugnis der Mittelschule ausgehändigt werden. Diese fließen bei der Notengebung nicht ein. Die Tatsache jedoch, dass

- alle Lernenden verpflichtend teilnehmen müssen, um die Mittelschule abzuschließen (vgl. 3.1) und
- beim Auswahlverfahren für die sich unmittelbar anschließenden ISCED 3 Schulformen die erreichten Ergebnisse mit einem Anteil von 50 % bei der Ermittlung der Rekrutierungspunktzahl („punkty rekrutacyjne“) einbeziehen,[7] kann der Mittelschulprüfung neben einer Informations- auch eine Selektionsfunktion zugeschrieben werden (vgl. CKE 2011).

3.3 Umgang mit Ergebnissen der Mittelschulprüfung

Die Zentrale Prüfungskommission (CKE) sowie die zuständige Regionale Prüfungskommission (OKE) bieten vor wie auch nach der Durchführung der Prüfung eine mit der nordrhein-westfälischen ZP 10 vergleichbare Transparenz. So können im Vorfeld in den Informationsmaterialien („Informator o egzaminie gimnazjalnym“; Informationsmaterial zur Mittelschulprüfung) und „Materialy dodatkowe“ (Zusätzliches Material; veröffentlicht 2010 bis 2012) sowie in den „Arkusze egzaminacyjne“ (Prüfungsbögen) verwendete Aufgabenformate, konkrete Aufgabestellungen sowie Bewertungsaspekte anhand von Beispiel[8]- und eingesetzten Prüfungsaufgaben im Detail eingesehen werden. Nach der Durchführung der Mittelschulprüfung veröffentlichen sowohl die CKE als auch die zuständige OKE Ergebnisberichte, die den Leistungsstand eines Jahrgangs landesweit, regional (Wojewodschaft, Bezirk) und lokal (Kreis) abbilden. Als vergleichbar kann hier die Praxis der Transparenzschaf-

[7] Die übrigen 50 % der Rekrutierungspunktzahl werden aus den Zeugnisnoten gebildet.

[8] Aufgaben, deren Ziel es ist, Transparenz im Hinblick auf Anspruch, Form und Bewertung von Prüfungsaufgaben zu schaffen (CKE 2010, S. 62 ff).

Tab. 2 Verteilung von Prüfungs- und Beispielaufgaben innerhalb der veröffentlichten Aufgabensammlung (Stand: 12.03.2016)

Zweck der Aufgabe	Prüfungsaufgabe	Beispielaufgabe
Anzahl	249	89
Anteil in %	73,7	26,3

fung von Polen und NRW auf Landes- und Individualebene (öffentlich, individuell) betrachtet werden[9].

Mit Blick auf eine Unterrichtssteuerung durch (Prüfungs-)Aufgaben ist in diesem Zusammenhang von besonderem Interesse, a) welchen Anteil Prüfungs- und Beispielaufgaben innerhalb der zur Verfügung stehenden Aufgabensammlung haben (vgl. Tab. 2) und b) inwieweit Prüfungsaufgaben die vorbereitenden Beispielaufgaben der CKE-Informationsmaterialien lediglich duplizieren.

Im Hinblick auf die erste Fragestellung zeigt sich, dass die Aufgaben des jährlich ergänzten Satzes der Prüfungsaufgaben gegenüber den Beispielaufgaben der Informationsmaterialien einen größeren Anteil in der veröffentlichten Aufgabensammlung ausmachen. Die begrenzte Zeitspanne der Veröffentlichung der Beispielaufgaben (2010–2012) deutet dabei eher auf Transparenzschaffung im Bezug auf die reformbedingte, „neue Aufgabenkultur" im Prüfungsfach Mathematik hin (vgl. 3.1). Eine weitergehende Betrachtung der beiden Aufgabensätze im Hinblick auf Fragestellung b) zeigt darüber hinaus, dass eine Duplizierung von Aufgaben in (jeweils) nur einer Aufgabenstellung vorfindbar ist. Aufgrund des hohen Maßes an verfahrensbedingter Unabhängigkeit in beiden Aufgabensätzen kann damit den Beispielaufgaben keine intendierte Steuerung seitens der CKE auf den Unterricht zugeschrieben werden. Ob die veröffentlichten Prüfungsaufgaben (bzw. die Beispielaufgaben) eine implizite *Backwash-Wirkung* erwarten lassen, hängt nach Konarzewski (2008, S. 6) unter anderem davon ab, ob die Prüfungen mechanisches Lernen betonen, was primär Gegenstand der Betrachtungen von Kap. 6 sein wird.

4 Zur Rolle von Mathematikaufgaben

4.1 Funktionale Typisierung von Mathematikaufgaben unter der Betrachtung kognitiver Aktivierung

Aufgaben spielen in der Unterrichtspraxis stets eine tragende Rolle, wobei sie in besonderem Maße die Lern- und Leistungssituationen im Fach Mathematik dominieren. Hier sind sie stets „eine Aufforderung zur gezielten Bearbeitung eines eingegrenzten mathematischen Themas." (Neubrand et al. 2011, S. 115). Unabhängig davon, ob Lern- oder Leistungsaufgaben, weisen sie dabei vielfältige strukturelle Gemeinsamkeiten auf, die nun vor dem Hintergrund der nachfolgend beschriebenen Untersuchungsmethode diskutiert werden.

[9] vgl. www.standardsicherung.schulministerium.nrw.de/cms/upload/zp10/berichte/ZP10_Ergebnisbericht_2014.pdf.

In Lernsitutationen sind *Unterrichtsaufgaben* eine Schnittstelle der Aktivität und Kommunikation zwischen Lehrkräften und Lernenden und damit ein zentrales Werkzeug der Unterrichtsgestaltung. Lehrkräfte steuern mit ihnen den Unterricht und damit auch die Aktivitäten der Lernenden. In diesem Prozess sind sie Träger mathematischer Inhalte und bestimmen diese weitgehend, sodass sie als das zentrale Steuerungsmedium für Kompetenzprogression aufgefasst werden können. Die von den Lernenden verlangten bzw. erlebten Anforderungen können folglich an den im Unterricht eingesetzten Aufgaben identifiziert werden. Wie im Falle der Einstiegs- und Hausaufgaben, die im Rahmen der COACTIV-Studie untersucht wurden (vgl. Jordan et al. 2006), setzen diese einen Rahmen für die Genese angestrebter mathematischer Fähigkeiten, die von den Lernenden beim aktiven Beschreiten des Lösungsprozesses zugleich bewusst erlebt werden können. Damit stellen Unterrichtsaufgaben im Prozess des Wissensaufbaus ein Vehikel vielfältiger kognitiver Aktivitäten bei Lernenden dar. Diese Aktivitäten können – wie beispielsweise bei COACTIV – in Form von Kategorien charaktersiert werden. Die Untersuchung der kognitiven Aktivierung durch Unterrichtsaufgaben erfasst somit die „Facetten kognitiver Prozesse beim Lösen der Aufgaben" (Neubrand et al. 2011, S. 118) innerhalb unterrichtlicher „Lerngelegenheiten" (ebd.).

Leistungs- bzw. Testaufgaben haben hingegen „die Funktion zu zertifzieren, welches Können im Bildungsprozess erworben wurde" (Terhart et al. 2009, S. 24). Testaufgaben zentraler Prüfungen werden extern konzipiert. Abhängig von ihrer Funktion wird bei der Aufgabenentwicklung beansprucht, die in den Vorgaben verankerten Anforderungen in unterschiedlicher Weise zu operationalisieren (vgl. Büchter 2007). Wider Erwarten lassen sich auf der Grundlage vertiefter Struktur- und Anforderungsanalysen jedoch deutliche Hinweise dafür identifizieren, dass Unterrichts- wie auch Testaufgaben „weitgehend identische kognitive Aufgabenmerkmale aufweisen" (Drüke-Noe 2014, S. 9) und sich der Unterschied lediglich im Umgang mit den Lösungen und deren Bewertung zeigt (vgl. Büchter und Leuders 2005). Vor dem Hintergrund der Funktion von Testaufgaben ist somit festzuhalten, dass hier die von den Lernenden im Lösungsprozess potentiell erlebten kognitiven Tätigkeiten als „erwartete Unterrichtswirkung" gedeutet werden können. Zusammenfassend lässt sich festhalten, dass Testaufgaben, wie sie im Rahmen dieser Studie untersucht werden sollen, in gleicher Weise wie Unterrichtsaufgaben, als Träger kognitiver Aktivierung aufgefasst und untersucht werden können. Im Bezug auf den hier betrachteten Aufgabentyp beschreibt dieses *kognitive Aktivierungspotential* somit analog die Facetten kognitiver Prozesse beim Lösen der Aufgaben.

4.2 Unterrichtsziele und kognitive Aktivierung

Die intendierten kognitiven Prozesse beim Lösen von Testaufgaben sollten die zentralen Ziele des Mathematikunterrichts hinreichend gut abbilden. In den curriularen Vorgaben Polens und Nordrhein-Westfalens werden – implizit (Polen, vgl. 3.2) und explizit (Nordrhein-Westfalen) – auf Winter (1995) zurückgehende allgemeine Bildungsziele des Mathematikunterrichts als zentral erachtet. Dieser geht davon aus, dass der allgemeinbildende Mathematikunterricht jedem Lernenden drei ausgewogen verteilte *Grunderfahrungen* ermöglichen sollte:

1. *Anwendungsorientierung*: Technische, soziale und kulturelle Erscheinungen und Vorgänge mithilfe der Mathematik wahrnehmen, verstehen und unter Nutzung mathematischer Gesichtspunkte beurteilen.
2. *Strukturorientierung*: Mathematik mit ihrer Sprache, ihren Symbolen, Bildern und Formeln in der Bedeutung für die Beschreibung und Bearbeitung von Aufgaben und Problemen inner- und außerhalb der Mathematik kennen und begreifen.
3. *Problemorientierung*: In der Bearbeitung von Fragen und Problemen mit mathematischen Mitteln allgemeine Problemlösefähigkeit erwerben (ebd., S. 38 ff).

Aus ihrer in konzeptioneller Hinsicht fundamentalen Bedeutung für die hiesigen Bildungsstandards und damit auch die im Zusammenhang mit dem Unterricht stehenden Prozesse folgt eine entsprechende Orchestrierung der zu vermittelnden Kompetenzen. Mit Blick auf Aufgaben als ein zentrales Mittel der Steuerung (Vorgaben abbildend) und Qualitätserfassung in einem Schulsystem (Umsetzung der Vorgaben erfassend) bedeutet dies: Unabhängig ihrer Funktion, d. h. sowohl als Test- als auch Unterrichtsaufgaben (vgl. 4.1), sollten die verankerten Charakteristika der kognitiven Aktivierung ein verhältniswahrendes Abbild dieser Vorgaben sein, was sich primär in einer ausgewogenen Verteilung der drei Grunderfahrungen im verankerten Anforderungsprofil zeigt. Hinsichtlich einer konkreten Zuordnung (Grunderfahrung → kognitive Tätigkeit als Kategorie), die eine derartige Realisierung der Grunderfahrungen sichtbar werden lässt, muss an dieser Stelle auf Drüke-Noe verwiesen werden (vgl. 2014, S. 248 ff).

Ziel ist es nun, die Aufgaben der Mittelschulprüfung so zu klassifizieren, dass ihr potentieller Beitrag zur kognitiven Aktivierung erkennbar wird. Die zur Umsetzung dieses Ziels herangezogenen Beurteilungskategorien haben sich empirisch bewährt und entstammen einschlägiger didaktischer und pädagogisch-psychologischer Literatur (vgl. 5.2).

5 Methode

5.1 Stichprobe und Datensatz

Bei den beiden im Rahmen dieser Studie untersuchten Datensätzen handelt es sich zum einen um die Aufgabensätze der polenweit im Zeitraum zwischen 2002 und 2015 durchgeführten Mittelschulprüfung. Hierzu zählen insgesamt 249 (Teil-)Aufgaben[10], die ausschließlich dem Testformat „Aufgabenset für Lernende ohne Einschränkungen und für Lernende mit spezifischen Lernschwierigkeiten“ angehören. An dieser Form der Mittelschulprüfung nehmen jährlich etwa 98 % der Schülerinnen und Schüler eines getesteten Jahrgangs teil. Da hiesige Prüfungen vergleichbaren Typs und Erhebungszeitpunkts mit Ausnahme von Rheinland-Pfalz zwar flächendeckend installiert sind, diese jedoch im Gegensatz zu Polen bundeslandintern konzipiert und organisert werden (vgl. Kühn 2013), wird dazu mit

[10] Aufgrund sprachlicher Vereinfachung ist stets die Rede von „Aufgaben“, wobei jedoch immer „Teilaufgaben“ (eine Analyseeinheit) gemeint sind, sofern nichts anderes ausdrücklich gesagt ist.

den schriftlichen „Zentralen Prüfungen 10“, kurz ZP 10, des Landes Nordrhein-Westfalen auf ein vergleichbares Instrument der Landesebene zurückgegriffen (N = 1556). Neben der höchsten Teilnehmerquote innerhalb Deutschlands und einer hohen funktionalen Äquivalenz beider Instrumente (vgl. 3.1–3.3 und Büchter und Pallack 2012; Büchter 2007) wird die Vergleichbarkeit der Befunde sichergestellt durch folgendes Vorgehen:

- Da die polnische Mittelschulprüfung verpflichtend ist und diese keiner vorgeschalteten Selektion bezüglich des Leistungsspektrums unterliegt, werden die nordrhein-westfälischen Aufgabensätze aller teilnehmenden Schulformen vereinigt. Damit setzt sich dieser untersuchte Aufgabensatz aus den Teilaufgabensätzen der Hauptschule (bestehend aus „Typ A“ und „Typ B“), der Gesamtschule (bestehend aus „Grundkurs“ und „Erweiterungskurs“), der Realschule, der Abendrealschule sowie des Gymnasiums zusammen (am Gymnasium nur zwischen 2007 und 2010, da die Prüfung an dieser Schulform danach abgeschafft wurde).
- Beide Datensätze bestehen gleichermaßen aus den Aufgaben, die zwischen 2007, dem Einführungsjahr der ZP 10, und 2015 eingesetzt wurden. Die erheblich kleinere Aufgabenzahl der polnischen Mittelschulprüfung hat dabei zur Folge, dass hier möglichen Fehlern in der vorgenommenen Kodierung eine höhere prozentuelle Auswirkung zukommt. Getroffene Aussagen sind hier somit mit Einschränkungen generalisierbar.

5.2 Aufgabenkategorien

5.2.1 Aufgabenkategorien auf der Grundlage der COACTIV-Studie

Um das kognitive Aktivierungspotential der Testaufgaben zu erfassen, stützt sich der Autor auf eine Auswahl empirisch bewährter Klassifikationskategorien, die im Rahmen des Projektes COACTIV (vgl. Jordan et al. 2006, 2008) formuliert und in den Analysen von Drüke-Noe (vgl. 2014) weiterentwickelt wurden. Die konzeptionellen Vorarbeiten von Neubrand et al. (2002) erweiternd untersuchten beide Studien großflächig die kognitive Aktivierung durch Unterrichts- wie auch Testaufgaben, die in neunten und zehnten Klassen an deutschen Schulen eingesetzt wurden. Das hier gewählte Klassifikationsschema fasst Analysekategorien von Aufgaben in drei Dimensionen zusammen: Inhaltlicher Rahmen, Kognitiver Rahmen sowie Kognitive Elemente des Modellierungskreislaufs. Diese werden weiter im Sinne einer Indikatorisierung in Kategorien mit zumeist vier Ausprägungsstufen ausdifferenziert (vgl. Tab. 3). Die Kategorien reichen dabei von niedrig inferenten (z. B. Stoffgebiete) bis hin zu hoch inferenten Klassifikationen (z. B. „Innermathematisches Modellieren“ oder „Typ mathematischen Arbeitens“). Weiterführendes ist im folgenden Klassifikationsschema und bei Jordan et al. (2006) nachzulesen.

Im Rahmen der Dimension „Inhaltlicher Rahmen“ werden einer Aufgabe zunächst ein oder mehrere Stoffgebiet/e zugeordnet, die den inhaltlichen Rahmen bei ihrer Bearbeitung ausmachen. Daneben wird in der Kategorie „Curriculare Wissensstufe“ das kognitive Aktivierungspotential erfasst, sodass aus der Zuweisung des

Tab. 3 Zentrale Analysekategorien der Studie (vgl. Jordan et al. 2008)

Dimension	Kategorie	Ausprägung
Inhaltlicher Rahmen	Stoffgebiet	1- Arithmetik, 2- Algebra, 3- Geometrie, 4- Stochastik
	Curriculare Wissenstufe	1- Grundkenntnisse (Stufe I, in der Grundschule erworben), 2- Einfaches Wissen der Sek. I, 3- Anspruchsvolles Wissen der Sek. I
Kognitiver Rahmen	Typ mathematischen Arbeitens	1- technische Aufgabe, 2- rechnerische Aufgabe, 3- begriffliche Aufgabe
Kognitive Elemente des Modellierungskreislaufs	Außermathematisches Modellieren	0- nicht benötigt, 1- Standardmodellierungen, 2- Mehrschrittige Modellierungen, 3- Modellreflexion, -validierung oder -eigenentwicklung
	Innermathematisches Modellieren	0- nicht benötigt, 1- Standardmodellierungen, 2- Mehrschrittige Modellierungen, 3- Modellreflexion, -validierung oder -eigenentwicklung
	Mathematisches Argumentieren	0- nicht benötigt, 1- Standardbegründungen, 2- Mehrschrittige Argumentation, 3- Entwicklung komplexer Argumentationen oder Beurteilen von Argumenten
	Gebrauch mathematischer Darstellungen	0- nicht benötigt, 1- Standarddarstellungen, 2- Wechsel zwischen Darstellungen, 3- Beurteilen von Darstellungen
	Umgehen mit mathematikhaltigen Texten	0- nicht bzw. kaum benötigt, 1- Unmittelbares Textverstehen, 2- Textverstehen mit Umorganisation, 3- Verstehen logisch komplexer Texte

Behandlungszeitraums der Inhalte die stoffliche Breite und damit das curriculare Anspruchsniveau eine Berücksichtigung findet (Jordan et al. 2008, S. 87).

Die nachfolgende Dimension „Kognitiver Rahmen“ erfasst unter der Kategorie „Typ mathematischen Arbeitens“ die Aufgabenklasse, indem der Typus des antizipierten Lösungsprozesses aufgezeigt wird. Hier kann zunächst zwischen zwei Aufgabenlösungsprozessen unterschieden werden: a) Aufgrund des Fehlens jeglicher Kontextanbindung ist der durchschrittene Lösungsprozess ohne eine Mathematisierung und ohne innermathematische Strukturierungsleistung beschreibbar (vgl. Abb. 1). b) Das Lösen der Aufgabe erfordert die Überführung des Problems „in einen verarbeitenden Ansatz und den Rückbezug des Ergebnisses auf die Ausgangssituation“ (Jordan et al. 2006, S. 31). Die in Abb. 1 verdeutlichte Kategorieausprägung „Technische Aufgabe“ dieser Analysekategorie weist dabei einen vorgegebenen Ansatz auf und erfordert im Lösungsprozess lediglich einen kontextlosen Abruf von „Fertigkeiten“ im Sinne eines bekannten Bearbeitungsalgorithmus (wie z. B. Rechnen oder Konstruieren nach vorgegebenen Regeln). Einem ebensolchen Aufgabentyp werden auch Aufgaben zugeordnet, die auf bereits vorgegebene Voraussetzungen die Anwendnung von „Faktenwissen“ verlangen (vgl. ebd., S. 47).

Modellierungsaufgaben, die beispielhaft in Abb. 2 dargestellt sind, erfordern hingegen die Überführung eines Problems in einen verarbeitenden Ansatz und den Rückbezug der Ergebnisse auf die Ausgangssituation. In dieser modellhaft als Modellierungskreislauf angenommenen Weise (vgl. Neubrand et al. 2011, S. 120) strukturierte Lösungprozesse finden ihren Niederschlag in rechnerischen und begrifflichen Modellierungsaufgaben. Rechnerische Modellierungsaufgaben werden dadurch cha-

Beende die Aufgabe mit dem richtigen Ergebnis	**Beende die Aufgabe mit dem richtigen Ergebnis**
Die Zahl $\frac{3^2+3^2+3^2}{3^3}$ ist gleich	Die Zahl $\frac{3^3 \cdot 3^4}{(3^3)^4}$ ist gleich
A. 3^{-5} B. 3^0 C. 3^5 D. 3^{-1}	A. 3^{-5} B. 3^0 C. 3^5 D. 3^{-1}

Abb. 1 Technische Aufgaben (vgl. CKE 2012a, 2012b)

Die Schüler der 3. Klasse haben eine Klassensprecherwahl durchgeführt. Dabei gab es drei Kandidaten: Ola, Pawel und Romek. In der Klasse sind 32 Schüler und jeder von ihnen hat eine gültige Stimme abgegeben. Gewonnen hat Ola, die weniger als die Hälfe aller Stimmen erhielt. Die restlichen Stimmen verteilen sich gleichmäßig auf die übrigen Kandidaten. Wie viele Stimmen erhielt Ola und wie viele jeweils die übrigen Kandidaten? Schreibe alle möglichen Wahlausgänge auf und begründe, dass es keine weiteren gibt.	Ein quaderförmiges Glasbehältnis mit den Maßen 6cm, 15cm und 18cm wurde zum Teil mit Wasser gefüllt und dicht verschlossen. Anschließend wurde es auf die Seite gelegt, die den größten Flächeninhalt besitzt. In dieser Lage reicht das Wasser 4cm hoch. Beende die Aufgabe. Wähle aus den vorgegebenen Antworten die richtige aus. Als das Glasbehältnis auf der Seite mit dem kleinsten Flächeninhalt aufgestellt wurde, reichte das Wasser bis zu einer Höhe von A. 8cm B. 10Cm C. 12cm D. 16cm

Abb. 2 Aufgaben des Testjahres 2010 und 2015 (CKE 2010, S. 12, 2015, S. 7)

rakterisiert, dass in ihrem Lösungsprozess das prozedural-algorithmische Denken dominiert (s. Abb. 2, rechte Aufgabe), bei den begrifflichen Modellierungsaufgaben steht im Verarbeitungsprozess das begriffliche Denken im Mittelpunkt (ebd., linke Aufgabe). Beide Aufgabentypen können dabei sowohl in einen inner- als auch einen außermathematischen Kontext eingebettet sein.

Die Kategorien der dritten Dimension „Kognitive Elemente des Modellierungskreislaufs" operationalisieren das Anspruchsniveau der kognitiven Prozesse und Aktivitäten zunächst innerhalb der (bisher) fünf mathematischen Tätigkeiten. Von grundlegender Bedeutung ist hierfür die Tatsache, dass der oben angesprochene problemhaltige Aufgabenkontext – unabhängig davon, ob inner- oder außermathematisch – in ein strukturgleiches Verarbeitungsmodell überführt wird, in dem Prozesse wie das Übersetzen/Übertragen, das Verarbeiten bzw. Interpretieren/Validieren kognitiv analog, wenngleich mit unterschiedlicher Ausprägung, vorkommen können (Jordan et al. 2006, S. 32). Der Vollzug dieser beiden Tätigkeiten kann – analog zum mathematischen Argumentieren sowie dem Gebrauch von mathematischen Darstellungen – auf vier im Hinblick auf Komplexität unterscheidbaren Niveaus stattfinden: 1. Keine Ausprägung/Zuweisung, 2. einfaches Anforderungsniveau, 3. mittleres Niveau und 4. hohes Niveau (vgl. Neubrand et al. 2011, S. 123). Für die Kategorie „Gebrauch von mathematischen Darstellungen", die im Wesentlichen auf ikonische Repräsentationsformen in einer Aufgabe Bezug nimmt, werden diese wie folgt operationalisiert:

- 0. ohne (ikonische) Darstellung,
- 1. einfaches Anforderungsniveau: „Informationen aus gegebenen Darstellungen (Tabelle, Graph oder Diagramm) entnehmen; Standarddarstellungen mittels gegebener Informationen anfertigen oder fortsetzen" (Jordan et al. 2006, S. 42);

Berechne: $2^3 \cdot 2^7 \cdot 4^2$

Abb. 3 Technische Aufgabe (Drüke-Noe 2014, S. 73)

- 2. mittleres Anforderungsniveau: „Darstellungsformen je nach Situation und Zweck auswählen; zwischen verschiedenen Darstellungen wechseln (Übersetzen); Zusammenhänge zwischen gegebenen Darstellungen herstellen." (ebd.)
- 3. hohes Niveau: Beurteilen von Darstellungen

Innerhalb der Kategorie „Umgehen mit mathematikhaltigen Texten" wird das Fähigkeitsniveau erfasst, welches notwedig ist, um aus dem Aufgabentext Informationen für die Struktur des Lösungsprozesses geordnet zu filtern. Hier verlangen beispielsweise die beiden Modellierungsaufgaben aus Abb. 2 ein sprachlogisch mittleres Niveau (Textverstehen mit Umorganisation), da die jeweilige Aufgabenstellung mehrere Haupt- und Nebensätze aufweist, wobei die Reihenfolge der Sätze bzw. Satzteile in erschwerter Weise den Schritten der Bearbeitung entspricht. Dies wird unter anderem auch daran deutlich, dass die vorgegebene sprachliche Konstruktion logische Implikationen in Form sprachlicher Rückbezüge aufweist (vgl. ebd.).

Demgegenüber sind beispielsweise die in Abb. 1 vorgestellten Testaufgaben auf dem Anspruchsniveau 0 angesiedelt, da die mathematische Bearbeitung hier ohne eine Verstehensleistung im Sinne der Kategorie „Umgang mit Textverstehen" erfolgen kann.

5.2.2 Erweiterungen des Klassifikationsschemas

Wie man beispielsweise aus dem Vergleich der Testaufgabe in Abb. 3, in der der Wert eines Terms mit einer multiplikativen Struktur berechnet werden soll, und den Aufgaben aus Abb. 2 entnehmen kann, können diese im Hinblick auf den Typ mathematischen Arbeitens der gleichen Aufgabenklasse (Technische Aufgabe) zugeordnet sein.

Die Lösungsfindung im Sinne der Aufgabenstellung ist jedoch ungleich „schwierig"; Tab. 4 verdeutlicht, dass die dabei geforderten kognitiven Handlungen in den Analysekategorien aus Tab. 3 dennoch ungestuft erfasst werden.

Die von Drüke-Noe entwickelte Analysekategorie „Technisches Arbeiten" ergänzt das bisherige Klassifikationsschema (vgl. Tab. 5), indem es erforderliche Tätigkeiten und deren Anforderungen an technischer Performanz expliziert. Da allen Technischen Aufgaben wie auch der Mehrzahl der Aufgaben, deren Lösungsprozess durch einen Modellierungskreis beschrieben werden kann, ein Umgang mit Kalkülen unterschiedlicher Komplexität innewohnt (vgl. Abb. 4.4 in Drüke-Noe 2014, S. 71), eröffnet sich damit die Möglichkeit, die Kategorie auf alle drei Aufgabenklassen anzuwenden. Sie erfasst hier – unterschiedlich im Lösungsprozess verortet (vgl. Abb. 4.3, S. 70 und Abb. 4.4, S. 71, ebd.) – gleichermaßen das Anspruchsniveau des Umgangs mit Kalkülen. Dieser Umgang findet gemäß des in Tab. 3 vorgestellten Klassifikationsschemas auf einem der vier Stoffgebiete (Artihmetik, Algebra, Geometrie und Stochastik) statt, innerhalb derer wie folgt vier Niveaus ausgewiesen werden.

Tab. 4 Kategorie-Ausprägungen von Aufgaben aus Abb. 1 und 3 in den Analysekategorien aus Tab. 3

	Stoffgebiet	Curriculare Wissensstufe	Aufgabenklasse	I. Modellieren	A. Modellieren	Argumentieren	Umgang mit Texten	Umgang mit Darstellungen
Aufgaben aus Abb. 1	1	3	1	0	0	0	0	0
Aufgabe aus Abb. 3	1	3	1	0	0	0	0	0

Über den COACTIV-Ansatz hinaus erlaubt eine derart operationalisierte mathematische Tätigkeit, nun das technische Anforderungsniveau im Hinblick auf das im Lösungsprozess eingebettete Technische Arbeiten zu verorten. Angewendet auf die beiden Aufgaben aus Abb. 1 und die Aufgabe aus Abb. 3, die in Tab. 3 hinsichtlich der technischen Performanz noch ungestuft erfasst wurden, zeigt die Zuordnung innerhalb der neuen Erfassungskategorie Unterschiede. So sind die Lösungsanforderungen der Aufgaben in Abb. 1 auf dem höchsten Anspruchsniveau der Arithmetik angesiedelt, da hier komplexe hierarchische Techniken angewendet werden, die sich durch eine (beliebige) Kombination von Potenz-, Punkt- und Strichrechnungen kennzeichnen. Dagegen ist die Technische Aufgabe aus Abb. 3 aufgrund der Struktur des Terms dem mittleren Anspruchsniveau zuzuordnen, da hier lediglich „einfache Vorrangregeln", wie Potenz- vor Punktrechnung, beachtet werden müssen.

5.3 Übersetzung und Kodierung

Im Vorfeld der Untersuchung wurden sämtliche 249 Aufgaben der vierzehn Aufgabensätze durch einen vereidigten, u. a. auf dem mathematisch-naturwissenschaftlichen Gebiet tätigen Übersetzer ins Deutsche übertragen.

Die Kodierung der übersetzten Aufgaben wurde unabhängig voneinander vom Autor (Kodierer 1) und einer erfahrenen Lehrerin (Kodiererin 2) vorgenommen, die im Vorfeld mit der Konzeption der Studie eingehend bekannt gemacht wurde. Da der Zuverlässigkeit der Bewertung eine grundlegende Bedeutung zukommt, wurde im Vorfeld ein Kodiertraining anhand von solchen COACTIV-Aufgaben durchgeführt, denen

- bereits im Rahmen des Projektes Ausprägungen in den COACTIV-Erfassungskategorien zugewiesen wurden,
- die Kodierung des „Technischen Arbeitens" durch Frau Drüke-Noe und ihre Mitarbeiter zugrunde lag.

In einem weitergehenden Schritt war es zunächst von Interesse, die Variabilität des *interrater agreements* zu quantifizieren, um so aus der Übereinstimmung Aufschluss über Stabilität und Zuverlässigkeit der Probekodierung zu erhalten. Als Maß zur Bestimmung dieser Übereinstimmungsgüte wurde der Cohens-Kappa-Koeffizient κ herangezogen. Als Minimalwert der Reliabilität und damit den Übergang zur Kodierung der Aufgaben aus der Mittelschulprüfung wurde $0{,}7 < \kappa$ festgelegt. In

Tab. 5 Niveaubeschreibungen des Technischen Arbeitens (vgl. Drüke-Noe 2014, S. 78)

Tätigkeit/ Kategorie	Stoffgebiet	Niveau	Bedeutung der Ausprägung und Beispiele
Technisches Arbeiten	Arithmetik	0	Nicht benötigt, z. B. a^0
		1	Nur Punkt- bzw. nur Strichrechnen
		2	Einfache hierarchische Techniken (Punkt- vor Strichrechnung, Potenz- vor Punktrechnung)
		3	Komplexe hierarchische Techniken (Potenz- mit Punkt- und Strichrechnung)
	Algebra	0	Nicht benötigt, z. B. Ist die Zuordnung ... eine Funktion?
		1	Nur Punkt- bzw. nur Strichrechnen, basales Punkt- vor Strichrechnen
		2	Einfache hierarchische Techniken (komplexe Punkt- vor Strichrechnung, Potenz- vor Strichrechnung)
		3	Komplexe hierarchische Techniken (Potenz- mit Punkt und Strichrechnung)
	Geometrie	0	Nicht benötigt, z. B. wie groß ist ein rechter Winkel?
		1	Nur Punkt- oder nur Strichrechnen, z. B. sin(23°) errechnen, mit Strahlensatz rechnen
		2	Einfache hierarchische Techniken (u. a. Potenz- vor Strichrechnung)
		3	Komplexe hierarchische Techniken (Potenz- mit Punkt- und Strichrechnung kombiniert)
	Stochastik	0	Nicht benötigt, z. B. alle Kombinationen notieren
		1	Nur Punkt- oder nur Strichrechnen
		2	Einfache hierarchische Techniken (Punkt- vor Strichrechnung, Pozenz- vor Punktrechnung)
		3	Komplexe hierarchische Techniken (Potenz- mit Punkt- und Strichrechnung kombiniert)

Anlehnung an Wirtz und Caspar (2002) nimmt man hierbei an, dass $0{,}75 < \kappa$ eine sehr gute, $0{,}6 < \kappa < 0{,}75$ eine gute und $0{,}4 < \kappa < 0{,}6$ eine akzeptable Konkordanz der Bewertungen beschreibt.

Die paarweisen Vergleiche der zwei durchgeführte Trainingskodierung, der COACTIV-Kodierungen bzw. der von Drüke-Noe und ihren Mitarbeitenden zeigten, dass die Übereinstimmungsgüte der Trainingskodierung ($N = 291$) in allen Kategorien/Ausprägungen nicht schlechter als „gut" zu beurteilen ist ($\kappa \geq 0{,}7$). Insgesamt lässt sich somit feststellen, dass die Güte der Kodierung des polnischen und des nordrhein-westfälischen Aufgabensatzes in hinreichendem Maße anschlussfähig an die Kodierung der Kategorien des COACTIV-Schemas und des „Technischen Arbeitens" ist.

6 Ergebnisse

Im folgenden Abschnitt werden die Ergebnisse der Studie zum kognitiven Aktivierungspotential der Aufgaben aus den polnischen Mittelschulprüfung 2002–2015 dargestellt. Die Ergebnisdarstellung erfolgt zunächst entlang der drei Dimensionen, die in Kap. 5 definiert wurden, innerhalb derer die Spannweite mathematischen Denkens bei der Aufgabenbearbeitung erfasst wird (6.1). Anschließend steht ein Vergleich mit nordrhein-westfälischen Aufgaben der Zentralen Prüfung (ZP 10) im Vordergrund der Betrachtungen (6.2).

6.1 Inhaltlich-kognitive Aktivierungsmerkmale von Mittelschulprüfungsaufgaben

Die Testaufgaben werden hier zunächst in den Kategorien der beiden Dimensionen inhaltlicher und kognitiver Rahmen erfasst (6.1.1). Anschließend richtet sich der Blick auf die Untersuchung des kognitiven Anspruchs, den Lösungsprozesse der Testaufgaben erfordern (6.1.2).

6.1.1 Inhaltlich-kognitiver Rahmen

Ein Ziel der Mittelschulprüfung ist es, das Kenntnis- und Fähigkeitsniveau von Neuntklässlern im Hinblick auf die Vorgaben des Rahmenplans der dritten Schulstufe zu erfassen. Die dort vorgenommene Schwerpunktsetzung auf die Stoffgebiete der Arithmetik, Algebra und Geometrie findet ebenfalls in der Verteilung auf die getesteten Gebiete einen entsprechenden Niederschlag. So zeigt sich, dass insgesamt 97,6 % aller Aufgaben dem Bereich der Arithmetik, der Algebra und der Geometrie zuzuordnen sind (vgl. Tab. 6). Hier ist es vor allem das Stoffgebiet der Arithmetik, das mit 44,2 % aller Testaufgaben mit deutlichem Abstand den größten Anteil ausmacht, während Algebra und Geometrie mit 24,9 % bzw. 28,5 % in etwa gleich häufig abgetestet werden, im Vergleich zu Arithmetik jedoch erheblich seltener vorkommen.

Das Stoffgebiet der Stochastik macht mit 5,2 % der Aufgaben den kleinsten Anteil aus. Betrachtet man jedoch die Inhaltsfelder des Rahmenplans im Fach Mathematik, in denen „Beschreibende Statistik und Einführung der Wahrscheinlichkeitsrechnung" lediglich einen äußerst beschränkten von insgesamt elf Lernbereichen ausmacht (vgl. 3.1), so kann dies folglich weniger als eine „Vernachlässigung" dieses Stoffgebiets in der Mittelschulprüfung als vielmehr in seinen inhaltlich-konzeptionellen Vorgaben gedeutet werden.

Tab. 6 Verteilung der polnischen Mittelschulprüfungsaufgaben auf die vier Stoffgebiete und die curricularen Wissensstufen

Stoffgebiet	Anzahl der Aufgaben ($N = 249$)	Anteil in %	Curriculare Wissensstufe $\bar{x}$
Arithmetik	110	44,2	2,07
Algebra	62	24,9	2,52
Geometrie	71	28,5	2,65
Stochastik	13	5,2	2,08
–	Summe der Zuordnungen bei $N = 249$:256	103	2,39

Einzelne Aufgaben können mehr als einem Stoffgebiet zugeordnet werden

Untersucht man nun die genannten Stoffgebiete im Hinblick darauf, welches curriculare Anspruchsniveau das darin getestete mathematische Wissen aufweist, so zeigt sich zunächst, dass das mittlere Niveau zwischen „einfachem Wissen aus der Sekundarstufe I“ und „anspruchsvollem Wissen der Sekundarstufe I“ angesiedelt ist. Kennzeichnend ist im Einzelnen zudem, dass zum einen keine der 249 Aufgaben ausschließlich Grundkenntnisse der deutschen Grundschule und somit nicht des „I Etap edukacyjny“ der polnischen Grundschule verlangt; zum anderen wird die 3. Wissensstufe aufgrund des Zeitpunktes der Durchführung und damit einhergehender erheblicher Einschränkung der Themenauswahl erwartungskonform schwächer besetzt.

Dass innerhalb der Stoffgebiete dennoch zum Teil erhebliche Unterschiede im curricularen Anspruchsniveau zu verzeichnen sind, belegt Tab. 6. Hier wird deutlich, dass beispielsweise die Aufgaben der Stoffgebiete Arithmetik und der Stochastik im Schnitt nur selten über das einfache Sekundarstufenniveau hinausgehen, wohingegen Aufgaben aus dem Bereich der Algebra und der Geometrie Wissen erfordern, das eher einer höheren curricularen Wissensstufe zuzuordnen ist. Unabhängig des curricularen Anspruchsniveaus konnten auf keinem der vier Stoffgebiete Vorgriffe im Curriculum festgestellt werden, wohingegen Rückgriffe im Hinblick auf den Rahmenplan der sechsjährigen polnischen Grundschule punktuell vorfindbar sind.

Die bisherigen Beschreibungen zeigen, dass das durchschnittliche curriculare Anspruchsniveau der polnischen Testaufgaben zwischen dem mit 2,02 als gering eingeschätzen Wert von PISA-E 2003 und dem im Rahmen der COACTIV-Studie erreichten Wert für neunte Klassen (GY: 2,80; nicht GY: 2,46) angesiedelt ist (vgl. Neubrand et al. 2011, S. 127). Darin wird zugleich eine Funktion dieses Instruments im System externer Leistungsmessung bestätigend aufgezeigt. Die Mittelschulprüfung erfasst das Wissens- und Fähigkeitsniveau auf der Basis des Curriculums der 7. bis 9. Klasse, das häufig genau den Übergang zwischen den Wissensstufen markiert. Andererseits findet man Rückgriffe auf die Stufe der Grundkenntnisse nicht vor, da diese vermutlich trennscharf im Rahmen des „sprawdzian“, dem zentralen Test am Ende der Grundschule, erfasst werden.

Wendet man sich nun der Frage zu, durch welche kognitiven Merkmale sich das mathematische Arbeiten beim Durchlaufen des Lösungsprozesses kennzeichnet, so kann hier annähernd ausschließlich eine Aufteilung auf rechnerische und begriffliche Modellierungsaufgaben festgestellt werden. Diese kognitiv „an sich“ als anspruchsvoller geltenden Aufgabenklassen machen zusammen mehr als 98 % aller Testaufgaben aus, wobei erstere mit mehr als zwei Drittel aller Aufgaben einen deutlichen Schwerpunkt der Mittelschulprüfung ausmachen. Sowohl bei begrifflichen als auch rechnerischen Aufgabenstellungen dominieren anwendungsorientierte Problemkontexte. Lediglich vier der 249 Aufgaben verlangen hingegen, dass Algorithmen bzw. Rechenverfahren auf einen vorgegebenen Ansatz angewendet werden, also ausschließlich „technisch abgearbeitet“ werden (vgl. Tab. 7).

Die Verteilung der Aufgaben auf die Aufgabenklassen macht deutlich, welche Ausrichtung der unterrichtlichen Aktivierung der Zentralen Prüfungskommission in den Jahren zwischen 2002 und 2015 wünschenswert erschien. So fand offenbar eine „von oben“ gelenkte Abkehr von einem traditionellem Selbstverständnis dieser Domäne statt, das häufig als „gedankenlos“ (Konarzewski 2004, S. 13), da rein

Tab. 7 Verteilung der polnischen Mittelschulprüfungsaufgaben auf die Typen mathematischen Arbeitens

	Typ mathematischen Arbeitens		
Ausprägung	Technische Aufgabe	Prozedurale/Rechnerische Modellierungsaufgabe	Konzeptuelle/Begriffliche Modellierungsaufgabe
Anzahl bei N = 249	4	173	72
Anteil in %	1,6	69,5	28,9

instrumentell und mit äußerst beschränktem Anwendungsbezug charakterisiert wurde (vgl. Moleda und Piesyk 1993; Ehrenfeucht und Stande 1970). Mathematische Begriffe und Verfahren werden im untersuchten Aufgabensatz annähernd ausschließlich als „Werkzeuge" (Neubrand et al. 2002, S. 102) eingesetzt. Damit scheint die vorfindbare Aufgabenkonstruktion a) die Prüfungsstandards, in denen annähernd ausschließlich anwendungs- bzw. problemorientiertes Wissen als verbindlich aufgeführt wird (vgl. Tab. 1), angemessen zu operationalisieren und b) Merkmale des Literacy-Konzepts aufzuweisen, das den Übergang vom Phänomen zu mathematischen Begriffen – im Gegensatz zu skills and routines der Lernenden – erfasst (PISA 2000: Aufgabenanteil technisch: 3,2 %; Aufgabenanteil rechnerisch: 45 %; Aufgabenanteil begrifflich: 52 %; vgl. Neubrand et al. 2002, S. 101). Dass diese Konzeptionsausrichtung tendentiell durchaus wünschenswert ist, in der vorfindbaren Ausprägung jedoch in qualitativer Hinsicht als unausgewogen erachtet werden kann, wird beispielsweise aus der Betrachtung des Rahmenplans der Mittelschule deutlich (vgl. MEN 1999, S. 609), der auch derartiges Fakten- und Fertigkeitswissen in den Fokus unterrichtlicher Bemühungen stellt. Dies scheint die Vermutung von Konarzewski (2004, S. 126 ff) zu untermauern, dass weniger dem Rahmenplan als vielmehr den Prüfungsstandards eine höhere potentielle Steuerungswirkung zugeschrieben werden kann.

Zusammenfassend lässt sich somit im Bezug auf den untersuchten inhaltlich-kognitiven Rahmen der Mittelschulaufgaben feststellen, dass die curriculare Wissensstufe insgesamt ein curriculumkonformes Aktivierungsniveau aufweist, wobei innerhalb der vier Stoffgebiete eine zum Teil erhebliche curriculare Streuung besteht. Die geforderten Aufgaben-Verarbeitungsprozesse lassen sich annähernd ausschließlich in begriffliche und rechnerische Modellierungsaufgaben trennen.

6.1.2 Kognitiver Anspruch

Die in 6.1.1 betrachteten Typen des mathematischen Arbeitens legen dar, welche Charakteristik der antizipierte Lösungsprozess einer Aufgabe aufweist, nicht jedoch welche kognitiven Komplexitäts-Ansprüche diesen im Einzelnen kennzeichnen. Um diese Untersuchungsdimension differenziert abzubilden, werden hierzu in den anspruchgenerierenden Kategorien neben dem im Durchschnitt erreichten Anspruchsniveau auch Verteilungen auf die zugehörigen vier Niveaus betrachtet. Tab. 8 zeigt, wie sich kognitive Niveaus innerhalb der sechs Analysekategorien im untersuchten Aufgabensatz verteilen.

Hier zeigt sich zunächst, dass die Untersuchung von Übergangsprozessen zwischen Realität und Mathematik (außermathematisches Modellieren) mit insgesamt

Tab. 8 Verteilungen der polnischen Mittelschulprüfungsaufgaben auf die anspruchsgenerierenden Kategorien und ihre Niveaus im Zeitraum zwischen 2002 und 2015

Mittelschulprüfungsaufgaben 2002–2015: Anspruchsgenerierende Kategorien und ihre Niveaus						
Kategorie	Niveau	0	1	2	3	$\overline{x}$
Außermathematisches Modellieren	Anzahl	58	63	122	6	1,304
	Anteil in %	23,2	25,6	48,8	2,4	
Innermathematisches Modellieren	Anzahl	194	4	39	12	0,474
	Anteil in %	77,9	1,6	15,7	4,8	
Mathematisches Argumentieren	Anzahl	238	1	10	0	0,086
	Anteil in %	95,5	0,4	4,1	0,000	
Gebrauch math. Darstellungen	Anzahl	112	105	31	1	0,682
	Anteil in %	45	42,2	12,4	0,4	
Technisches Arbeiten	Anzahl	39	130	55	25	1,264
	Anteil in %	15,6	52,2	22,1	10,0	
Umgang mit math. Texten	Anzahl	15	162	59	13	1,281
	Anteil in %	6,0	65,1	23,7	5,2	

76,8 % aller Aufgaben den Schwerpunkt der Mittelschulprüfung ausmacht. Darunter sind insbesondere Aufgaben des mittleren Niveaus (48,8 %) vertreten, wohingegen einfaches und besonders anspruchsvolles außermathematisches Modellieren 25,6 % bzw. 2,4 % aller Aufgaben ausmachen. Mit 22,1 % aller Aufgaben entfällt dagegen ein erheblich kleinerer Anteil auf das innermathematische Modellieren. Abweichend zur Kategorie außermathematischen Modellierens gilt hier: Verlangt eine Aufgabe das innermathematische Modellieren, so findet es vorwiegend auf dem mittleren oder hohen Niveau statt. Damit scheint die vorfindbare Aufgabenkonstruktion im Hinblick auf die kontextuelle Einbettung der Aufgaben in dieser Ausprägung eher unausgewogen. So stellen die Rahmenpläne vor und nach der Reform den Aufbau innermathematischer Wissensnetze bzw. Fähigkeiten gleichermaßen als einen zentralen Aspekt des Mathematikunterrichts heraus, der hier jedoch eher in geringerem Maße vertreten scheint (vgl. MEN 1999, S. 609). Diese Diskrepanz kann vor allem als ein weiterer Beleg für ein eingeschränktes Maß an Korrespondenz zwischen dem Rahmenplan und den abbildenden Prüfungsstandards gedeutet werden (vgl. 3.1).

Das im Mittel erreichte Anpruchsniveau beim Argumentieren kann sowohl im Vergleich zu den übrigen Tätigkeiten als auch zu dem mittleren Anspruchsniveau der PISA-Aufgaben als vergleichsweise niedrig eingeschätzt werden (Argumentieren bei PISA 2003 international: 0,15; PISA 2003 national: 0,26; vgl. Neubrand et al. 2011, S. 128). Entgegen der Erwartung zeigt sich damit zugleich, dass konzeptionelle Vorgaben der Testaufgaben, die Argumentieren im Zeitraum 2002–2011 implizit (u. a. Standard: „Der Schüler zeigt die Richtigkeit von Prozessen auf.“; vgl. Tab. 1) und ab 2012 explizit aufführen, offenbar nicht hinreichend berücksichtigt werden. Wird Argumentieren im Lösungsprozess der Testaufgaben verlangt, so handelt es sich in zehn von den insgesamt 11 Fällen um mehrschrittige Argumentationen mittleren Niveaus, die im Lösungsprozess sowohl inner- als auch außermathematischer Kontexte vorwiegend offener Aufgabenformate verlangt werden. Auch das im Mittel erreichte Anspruchsniveau beim Umgang mit mathematischen Darstellungen trägt in vergleichsweise geringem Maße zur kognitiven Komplexität der Testaufgaben bei, da es zumeist nicht erforderlich ist (45 %) bzw. auf einfachem Niveau verlangt wird (42,2 %). Dass der Befund gegenüber anderen Tätigkeiten dennoch relativert werden kann und auch sollte, zeigt beispielsweise ein Vergleich mit dem als hoch eingeschätzen Kategorieergebnis von PISA (PISA 2003 international: 0,49; PISA 2003 national: 0,69; vgl. ebd.).

Das Technische Arbeiten wird in 16,9 % der Aufgaben nicht verlangt. Das mittlere Anspruchsniveau zeigt dabei, dass es im Vergleich zu den bisher beschriebenen Kategorien eher in hohem Maße zum kognitiven Aktivierungspotential beiträgt. Ist es für den Lösungsprozess erforderlich, so findet es zumeist auf niedrigem (52,6 %) oder mittlerem (22,1 %) Niveau statt. Eine ähnliche Verteilung zeigen die Aufgaben innerhalb der Kategorie „Umgang mit mathematischen Texten“, die neben dem außermathematischen Modellieren und dem Technischen Arbeiten den höchsten Kategorien-Schnitt und damit – relativ betrachtet (vgl. ebd.) – als besonders potentiell kognitiv aktivierend einzustufen sind.

Zusammenfassend kann somit gesagt werden, dass die in der polnischen Mittelschulprüfung eingesetzten Aufgaben insgesamt ein verhältnismäßig hohes kognitives Anspruchsniveau aufweisen. Hierzu tragen insbesondere vergleichweise anspruchsvolle mathematische Texte und Darstellungen sowie das Technische Arbeiten bei, das in anspruchsvollen, da zumeist mehrschrittigen Verarbeitungsprozessen eingebettet ist. Das Argumentieren bestimmt das kognitve Anspruchsniveau hingegen eher in geringerem Maße.

6.2 Vergleich des kognitiven Anspruchs der Testaufgaben der Mittelschulprüfung und der ZP 10

Um den kognitiven Anspruch der polnischen Mittelschulprüfung (A2) im Sinne der zweiten erkenntnisleitenden Frage zu untersuchen, wird nun der Aufgabensatz der ZP 10 (A1) zum Vergleich herangezogen. Tab. 9 verdeutlicht die Verteilung der Aufgaben beider Aufgabensätze über die Niveaus der mathematischen Tätigkeiten.

Es zeigt sich, dass die Testaufgaben der ZP 10 gegenüber denen der Mittelschulprüfung tendentiell einen höheren Anteil außermathematischen Modellierens verlangen (75,2 % gegenüber 67,5 %), wenngleich die zugehörigen Mittelwerte durchaus vergleichbar sind. Betrachtet man nun die beiden Verteilungen der Tätigkeitsniveaus innerhalb dieser Kategorie, so wird deutlich, dass diese Verteilungsmerkmale primär dadurch bestimmt werden, dass zum einen im Aufgabensatz A2 außermathematische Modellierungen auf einfachem Niveau (24,1 %) gegenüber dem A1-Aufgabensatz (32,4 %) merklich schwächer besetzt werden; zum anderen weisen Modellierungsaufgaben des mittleren und hohen Niveaus jeweils vergleichbare Anteile auf. Das innermathematische Modellieren erfordert im Vergleich zum außermathematischen Modellieren ein geringerer, wenngleich unterschiedlich hoher Aufgabenanteil in den beiden Aufgabensätzen. Es wird in insgesamt 16,2 % der ZP 10 Aufgaben und 31,1 % der polnischen Mittelschulaufgaben verlangt. Ein Vergleich der zugehörigen Verteilungen zeigt darüber hinaus, dass die nordrhein-westfälischen Aufgaben hauptsächlich auf einfachem und mittlerem Niveau angesiedelt sind (zusammen 15,4 % von insg. 16,2 %), wohingegen die Mittelschulaufgaben annähernd ausschließlich dem mittleren und hohen Niveau zuzuordnen sind (zusammen 30,5 % von insg. 31,7 %). Während das kognitive Anspruchsniveau innerhalb der beiden bisher ver-

Tab. 9 Verteilung der Aufgaben der ZP 10 aus Nordrhein-Westfalen (A1) und der polnischen Mittelschulprüfung (A2) auf die anspruchsgenerierenden Kategorien und ihre Niveaus im Zeitraum zwischen 2007 und 2015

		ZP 10 (A1) und Mittelschulprüfung (A2): Verteilung der Aufgaben des Zeitraum zwischen 2007 und 2015 auf die Kategorien und ihre Niveaus									
		Aufgabensatz 1 (A1: $N = 1556$)					Aufgabensatz 2 (A2: $N = 166$)				
Kategorie	Niveau	0	1	2	3	x_{A1}	0	1	2	3	x_{A2}
Außermathematisches Modellieren	Anzahl	386	504	654	12	1,188	54	40	67	5	1,137
	Anteil in %	24,8	32,4	42	0,8		32,5	24,1	40,3	3,0	
Innermathematisches Modellieren	Anzahl	1304	67	174	12	0,289	113	2	43	8	0,670
	Anteil in %	83,8	4,3	11,1	0,8		68,1	1,2	25,9	4,8	
Mathematisches Argumentieren	Anzahl	1402	78	76	0	0,148	155	1	10	0	0,126
	Anteil in %	90,1	5,0	4,9	0,0		93,4	0,6	6	0,0	
Gebrauch math. Darstellungen	Anzahl	550	880	126	0	0,729	74	66	25	1	0,719
	Anteil in %	35,3	56,5	8,2	0,0		44,4	39,9	15,1	0,6	
Technisches Arbeiten	Anzahl	304	720	296	236	1,299	29	83	37	16	1,241
	Anteil in %	19,5	46,3	19,0	15,2		17,5	50,0	22,2	10,2	
Umgang mit math. Texten	Anzahl	335	980	241	0	0,94	16	101	43	6	1,234
	Anteil in %	21,5	63	15,5	0		9,64	60,8	25,9	3,6	

glichenen Tätigkeiten auf dem mittleren bzw. hohen Niveau angesiedelt ist (PISA 2003 international: 1,52 (Außermath. Modellieren) und 0,20 (Innermath. Modellieren); vgl. Neubrand et al. 2011, S. 127 f), wird das mathematische Argumentieren, das 9,9 % bzw. 6,6 % der jeweiligen Aufgaben erfordern, in einem relativ geringen Umfang verlangt (vgl. ebd.). Dabei charakterisiert die Verteilung des ZP 10 Aufgabensatzes eine nahezu gleichmäßige Besetzung des einfachen und mittleren Niveaus, wohingegen die Aufgaben der Mittelschulprüfung annähernd ausschließlich Argumentationen mittleren Niveaus verlangen. Die bisherigen Aussagen über die Aufgaben der Mittelschulprüfung sind im Bezug auf die Berücksichtigung der Aufgabenkontexte und des Argumentierens jedoch eher als diffus einzustufen. So kann vermutlich davon ausgegangen werden, dass infolge der Reformen der Mittelschulprüfung im Jahre 2012 auch längsschnittliche Auswirkungen auf die Aufgabenkonzeption zu erwarten sind. So wird aus dem Vergleich der beiden konzeptionellen Grundlagen vor und nach der Reform (vgl. 3.1) deutlich, dass zum einen innermathematische Problemkontexte gegenüber außermathematischen und zum anderen das Argumentieren an Bedeutung gewonnen haben. Eine zu erwartende Umsetzung der Vorgaben würde jüngst tendentiell a) eine stärkere Divergenz hinsichtlich der Berücksichtigung der Kontexte und b) eine Annäherung des Anspruchs an das Argumentieren in beiden Aufgabenprofilen nach sich ziehen. Eine erste längsschnittliche Analyse der Mitteschulprüfungsaufgaben vor und nach der Reform bestätigt diese Vermutung.

Der Umgang mit mathematischen Darstellungen weist in beiden Aufgabensätzen ein vergleichbares mittleres Anspruchsniveau auf, das als relativ hoch einzustufen ist (vgl. Neubrand et al. 2011, S. 128 f). Im nordrhein-westfälischen Aufgabensatz dominieren dabei vor allem Darstellungen auf einfachem Niveau, wohingegen die Aufgaben der Mittelschulprüfung tendentiell gleichmäßiger verteilt sind. Das Anspruchsniveau, auf dem das Technische Arbeiten im Schnitt verlangt wird, ist in beiden Aufgabensätzen im Vergleich zu den anderen Tätigkeiten vergleichsweise hoch. Wie es jedoch im Verhältnis zu anderen zentralen Abschlussprüfungen zu beurteilen ist, kann derzeit nicht verlässlich beantwortet werden, da entsprechende Vergleichswerte über längere Zeiträume nicht vorliegen. Das im Mittel erreichte Anspruchsniveau der polnischen Testaufgaben entspricht dabei etwa dem der ZP 10, wobei zugleich die drei Niveaus in beiden Aufgabensätzen nur mit geringfügigen Abweichungen besetzt werden. Interessant ist hierbei, dass der aus dem bildungsbiographischen Durchführungszeitpunkt der ZP 10 resultierende curriculare Wissensvorsprung, der in erster Linie Themengebiete wie quadratische Funktionen/Gleichungen, Exponentialfunktionen oder Aspekte der Trigonometrie umfasst, damit nicht mit einer Erhöhung des mittleren Anspruchsniveaus im Umgang mit Kalkülen einhergeht. Der Umgang mit Texten weist im Schnitt in beiden Aufgabensätzen ein vergleichsweise hohes (vgl. ebd.), jedoch eingeschränkt vergleichbares Anspruchsniveau auf. So zeichnen sich die Aufgaben der Mittelschulprüfung durch eine sichtlich gleichmäßigere Verteilung über alle Niveaus aus, was nur bedingt mit einem kleineren Anteil von „textarmen“ Technischen Aufgaben erklärt werden kann.

Insgesamt sind die Verteilungen der Tätigkeiten und ihrer Niveaus in beiden Aufgabensätzen mit Einschränkungen vergleichbar. Zu den gemeinsamen Aufgabenmerkmalen zählt das recht hohe kognitive Anspruchsniveau, das über eine große Bandbreite von eingeforderten Tätigkeiten und deren mittlere Anspruchsniveaus generiert wird. Ein Unterschied, der vermutlich seit der Reform der Mittelschulprüfung im Jahre 2012 als fundamental einzustufen ist, kann dagegen im Maß und Anspruchsniveau der Berücksichtigung innermathematischer Modellierungensaufgaben ausgemacht werden.

7 Zusammenfassung und Diskussion der Befunde

Die Ergebnisse der Studie zeigen, dass die Mathematikaufgaben der polnischen Mittelschulprüfung im Zeitraum beachtlicher Leistungszuwächse Polens insgesamt ein vergleichweise hohes kognitives Aktivierungspotential aufweisen. Kennzeichnend ist hierbei für den gesamten Untersuchungszeitraum (2002–2015), dass der zentrale Typ mathematischen Arbeitens in den Testaufgaben durch begriffliche und vor allem rechnerische Modellierungen charakterisierbar ist. Kontextlose, Technische Aufgaben bleiben dagegen annähernd unberücksichtigt, womit die Prüfungs- und vor allem Unterrichtsvorgaben im Bezug auf die Aufgabenklasse nur eingeschränkt abgebildet werden.

Mit Ausnahme der Auswahl der Prüfungsinhalte (curriculare Wissensstufe) weisen die Aufgaben darüber hinaus in der durch außer- und innermathematische Modellierungen, mathematische Darstellungen, das Technische Arbeiten und den Umgang mit mathematischen Texten vertretenen Spannweite anspruchsgenerierender Kategorien eine breites Aktivierungspotential auf. Die vorfindbare Ausprägung deckt sich hierbei in auffällig hohem Maße mit der bildungsbiographisch annähernd zeitgleich stattfindenden PISA-Studie. In Anlehnung an die aktuelle Literatur scheint es vor diesem Erkenntnishintergrund zunächst plausibel, die Ursachen der aufgezeigten Leistungszunahme polnischer Schülerinnen und Schüler beispielsweise auf dem Feld einer sich reformbedingt gleichgerichtet wandelnden Unterrichtskultur zu verorten, zu der ebenfalls die Unterrichtssteuerung durch Test- bzw. Beispielaufgaben zählt. Während erste Studien bereits Hinweise auf einen sich derart wandelnden Mathematikunterricht und seine Rahmenbedingungen liefern (vgl. Karpinski et al. 2013; Konarzewski 2008), konnte hier eine „Aufgabensteuerung" mittels Beispielaufgaben allenfalls als ein nicht intendierter Effekt identifiziert werden. So zeigten die Analysen, dass der festgestellte Anteil der Prüfungsaufgaben, die diese lediglich duplizieren, vernachlässigbar ist. Inwiefern jedoch Prüfungs- und Beispielaufgaben der Mittelschulprüfung tatsächlich auf die Unterrichtsgestaltung Einfluss nehmen, bleibt zunächst offen. Vor dem Hintergrund des festgestellten kognitiven Aktivierungspotentials erscheint es jedoch zweifelhaft, ob das Wirkungsmaß im Vergleich zu Prüfungen und Beispielaufgaben, in denen vorwiegend das mechanische Lernen betont wird, als vergleichbar einzustufen ist.

Dennoch sind die Ergebnisse der vorgenommenen Analysen der Testaufgaben in Teilen kritisch zu sehen. So wird in den Aufgaben die Breite der drei Grunderfahrungen des allgemeinbildenden Mathematikunterrichts nach Winter zunächst in zweifacher Hinsicht unverhältnismäßig abgebildet. Obwohl die Bedeutung innermathematischer Aspekte in beiden Rahmenplänen besonders hervorgehoben wird und damit offensichtlich als steuerungswirksam für die Unterrichtsgestaltung angesehen werden kann, würde man neben substantiellen Anteilen des Technischen Arbeites und einem Umgehen mit Darstellungen auf verschiedenen Niveaus auch höhere Anteile des innermathematischen Modellierens erwarten. Hier scheint folglich die Strukturorientierung einen, beispielsweise gegenüber der Anwendungsorientierung, zu eingeschränkten Beitrag zur Erfassung der Allgemeinbildung zu leisten. Ein analoges Bild zeigt sich im geringen Erfassungsmaß der „Reflexion über das eigene Denkhandeln" (Winter 1995, S. 41) in Form des mathematischen Argumentierens, die ihrerseits einen substantiellen Aspekt der Grunderfahrung Problemorientierung darstellt. Die Analysen zeigen interessanterweise, dass die beschriebene Schwerpunktsetzung der Mittelschulprüfung dabei weniger auf eine eingeschränkte Konstruktvalidität des Instruments zurück zu führen ist. So deutet ein Vergleich der Anforderungen der Prüfungsstandards (Wirkungserwartungen) und der Testaufgaben (Wirkungserfahrungen) vielmehr darauf hin, dass erstere erwartungskonform operationalisiert werden. Hier scheint vielmehr eine eingeschränkte Korrespondenz zwischen Prüfungsstandards und dem Rahmenplan offenbar eine unausgewogene Berücksichtigung der Winterschen Grunderfahrungen in den Prüfungsaufgaben zu bedingen, was anschlussfähig an bisherige Befunde aus fachdidaktischer Forschung ist (vgl. Konarzewski 2004).

Diese Steuerungsproblematik scheint mit der Einbettung der Standards in den Rahmenplan (2012) erkannt und behoben; sinnvollerweise wäre es ebenfalls wünschenswert, die Aufgabenkompetenz, d. h. die inhaltlich-kognitive Konzeption der

Testaufgaben, im Sinne der kognitiven Aktivierung in den Förderungsfokus verantwortlicher CKE-Gremien zu nehmen. Auch vor dem Hintergrund der Steuerungswirkung durch Testaufgaben wären hierzu zunächst Fortbildungen wünschenswert, deren Ziel es sein sollte, die hier betrachteten Aufgabenmerkmale bewusst zu machen und bewusst im Prozess der Aufgabenkonzeption zu verankern.

In der Gegenüberstellung der Aufgaben der ZP 10 und der Mittelschulprüfung im Zeitraum zwischen 2007 und 2015 zeigt sich, dass die untersuchten mathematischen Tätigkeiten und, mit erheblichen Einschränkungen, auch ihre Niveaus in beiden Aufgabensätzen vergleichsweise ausgewogen verteilt sind (vgl. Drüke-Noe 2014, S. 160 f; Neubrand et al. 2011, S. 127 f). Damit sind die Ergebnisse hinsichtlich des kognitiven Aufgabenanspruchs des ersten Aufgabensatzes anschlussfähig an die wenigen bislang vorliegenden Forschungsergebnisse zu innerdeutschen MSA-Prüfungen (vgl. Drüke-Noe 2014, S. 160 ff). Wesentliche Unterschiede zeigen sich zunächst in der Berücksichtigung der Tätigkeiten des innermathematischen und des außermathematischen Modellierens; hier greifen die nordrhein-westfälischen Aufgaben häufiger außermathematische Kontexte auf, wohingegen die Aufgaben der Mittelschulprüfung ausgewogener über die beiden Tätigkeiten verteilt sind. Lässt man in dieser Betrachtung die beiderseits geringfügig berücksichtigte Aufgabenklasse „Technischen Aufgaben“ einfließen, so können die genannten kontextuellen Einbettungen tendentiell gar insgesamt als profilbildend angesehen werden. Darüber hinaus lässt die im Zuge der Reform verankerte curriculare Schwerpunktbildung der Mittelschulprüfung wie auch die aktuellen längsschnittlichen Entwicklungsbefunde der innerdeutschen MSA diesbezüglich offenbar auf eine sich weiter verstärkende Divergenz beider Profile schließen (vgl. 3.1; Kühn und Drüke-Noe 2013). Die Aufgabenverteilung auf die drei Niveaus des inner- und außermathematischen Modellierens zeigt dabei, dass das kognitive Anspruchsniveau beider Tätigkeiten bei der ZP 10 geringfügig niedriger ausfällt, die Stufe der Verallgemeinerung und Bewertung bleibt darüber hinaus annähernd gänzlich unberücksichtigt. In beiden Aufgabensätzen weisen Argumentationen einen vergleichsweise geringen Grad an kognitiver Aktivierung auf, wohingegen der Umgang mit mathematischen Darstellungen bzw. Texten ein, beispielsweise mit PISA-Aufgaben, vergleichbares Anspruchsniveau verlangt (vgl. Neubrand et al. 2011, S. 128 f). Damit ist die im nordrhein-westfälischen Kernlehrplan explizit postulierte Umsetzung der drei auf Winter zurückführbaren Grunderfahrungen sowie die vernetzt und auf allen Anspruchsniveaus erfolgende Kompetenzentwicklung (MSW 2004, S. 11 ff), trotz Verbindlichkeit, interessanterweise im hiesigen Aufgabensatz in einem geringeren Maße als im polnischen realisiert. Vielmehr scheint das Anforderungsspektrum der Mittelschulprüfungsaufgaben und das der nordrhein-westfälischen Unterrichtsvorgaben aktuell zunehmend zu korrespondieren. Die bereits von Neubrand und Neubrand festgestellte (2010), eingeschränkte Konstruktvalidität, d. h. die Diskrepanz zwischen Wirkungserwartung und Wirkung, wird im deutschen Aufgabensatz primär darin deutlich, dass „Mathematik als Anwendung“ in erheblich stärkerem Maße eine Brücksichtigung erfährt, die Grunderfahrungen „Mathematik als Struktur“ oder auch „Mathematik als kreatives und intellektuelles Handlungsfeld“ dagegen hinsichtlich Breite und Aktivierungsniveau offenbar nur defizitär erfasst werden. Letzteres stützt darüber hinaus für ZP 10 die Interpretation, dass hier ein durchschnittlich geringes Aktivierungsniveau einer bestimmten Tätigkeit nicht durch ein hohes einer anderen „erkauft“ wird, was bei polnischen Mittelschulaufgaben im Hinblick auf inner- und außermathematische Modellierungen zutrifft.

Literatur

van Ackeren, I. (2003). *Evaluation, Rückmeldung und Schulentwicklung. Erfahrungen mit zentralen Tests, Prüfungen und Inspektionen in England, Frankreich und den Niederlanden*. Münster: Waxmann.
Banach, C. (1999). *Ku dobrej edukacji*. Torun: Adam Marszalek.
Białecki, I., & Haman, J. (2001). *Program Międzynarodowej Oceny Umiejętności Uczniów OECD/PISA*. Warszawa: Wydawnictwo Naukowe Scholar.
Büchter, A., & Leuders, T. (2005). *Mathematikaufgaben selbst entwickeln. Lernen fördern – Leistung überprüfen*. Berlin: Cornelsen Scriptor.
Büchter, A. (2007). Zentrale Prüfungen am Ende der Klasse 10 – Wirkungen auf Unterricht und Leistungsbewertung. In *Beiträge zum Mathematikunterricht 2007* (S. 503–506). Hildesheim: Franzbecker.
Büchter, A., & Pallack, A. (2012). Zur impliziten Standardsetzung durch zentrale Prüfungen – methodische Überlegungen und empirische Analysen. *Journal für Mathematik-Didaktik*, *33*, 59–85.
CKE (2007). Standardy wymagan bedace podstawa przeprowadzania egzaminu gimnazjalnego w ostatnim roku nauki w gimnazjum. https://www.cke.edu.pl/images/stories/000011_gim_pr/mas_gimn.pdf. Zugegriffen: 12.7.2016.
CKE (2010). Informator o egzaminie gimnazjalnym od roku szkolnego 2011/12. http://www.cke.edu.pl/images/files/Gimnazjum_2011_2012/Informator_G1.pdf. Zugegriffen: 7.10.2015.
CKE (2011). Ramowy opis polskiego systemu egzaminów zewnętrznych. www.bip.cke.edu.pl/bip_download.php?id=1293. Zugegriffen: 20.10.2015.
CKE (2012a). Egzamin w klasie trzeciej gimnazjum. Czesc matematyczno-przyrodnicza- matematyka. https://www.cke.edu.pl/egzamin-gimnazjalny/arkusze/. Zugegriffen: 14.02.2016.
CKE (2012b). Badanie Diagnostyczne w klasie trzeciej gimazjum. Czesc matematyczno-przyrodnicza-matematyka. https://www.cke.edu.pl/egzamin-gimnazjalny/materialy-dodatkowe/przykladowe-arkusze-2/. Zugegriffen: 20.07.2016.
CKE (2015). Egzamin w klasie trzeciej gimazjum. Czesc matematyczno-przyrodnicza- matematyka. https://www.cke.edu.pl/images/_EGZAMIN_GIMNAZJALNY/Arkusze-egzaminacyjne/2015/matematyka/Zeszyt_zadan_GM-MX1-152.pdf. Zugegriffen: 12.10.2015.
Drüke-Noe, C. (2014). *Aufgabenkultur in Klassenarbeiten im Fach Mathematik. Empirische Untersuchungen in neunten und zehnten Klassen*. Wiesbaden: Springer Spektrum.
Ehrenfeucht, A., & Stande, O. (1970). *Algebra dla Klasy II Liceum Ogolnoksztalcacego*. Warszawa: PZWS.
Greefrath, G. (2004). Offene Aufgaben mit Realitätsbezug. Eine Übersicht mit Beispielen und erste Ergebnisse aus Fallstudien. *mathematica didactica*, *27*(2), 16–38.
Jordan, A., Ross, N., Krauss, S., Baumert, J., Blum, W., Neubrand, M., Löwen, K., Brunner, M., & Kunter, M. (2006). *Klassifkationsschema für Mathematikaufgaben: Dokumentation der Aufgabenkategorisierung im COACTIV-Projekt*. Berlin: Max-Planck-Institut für Bildungsforschung.
Jordan, A., Krauss, S., Löwen, K., Blum, W., Neubrand, M., Brunner, M., Kunter, M., & Baumert, J. (2008). Aufgaben im COACTIV-Projekt: Zeugnisse des kognitiven Aktivierungspotentials im deutschen Mathematikunterricht. *Journal für Mathematikdidaktik*, *29*, 83–107.
Karpinski, M., Grudniewska, M., & Zambrowska, M. (2013). *Nauczanie matematyki w gimnazjum- Raport z badania*. Warszawa: IBE.
Klieme, E., Neubrand, M. & Lüdtke, O. (2001). Mathematische Grundbildung: Testkonzeption und Ergebnisse. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, P. Stanat, K.-J. Tillmann & M. Weiß (Hrsg.), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 139–190). Opladen: Leske + Budrich.
Konarzewski, K. (2004). *Reforma oswiaty. Podstawa programowa i warunki kształcenia*. Warszawa: ISP.
Konarzewski, K. (2008). *Przygotowanie uczniów do egzaminu: pokusa łatwego zysku. Raport badawczy*. Warszawa: ISP.
Krüger, M. (2015). *Aufgabenkultur in zentralen Abschlussprüfungen: Exploration und Deskription naturwissenschaftlicher Aufgabenstellungen im internationalen Vergleich*. Münster: Waxmann.

Kühn, S., & Drüke-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.

Kühn, S. (2013). Vergleichbarkeit beim Mittleren Schulabschluss? Ein Überblick über die Vielfalt schulstrukturell möglicher Bildungswege und Prüfungsverfahren in den deutschen Ländern. *Die Deutsche Schule*, *105*, 87–101.

Kupisiewicz, C. (2006). *Projekty reform edukacyjnych w Posce*. Warszawa: PWN.

Maier, U., Bohl, T., Metz, K., & Kleinknecht, M. (2011). Einflüsse von Merkmalen des Testsystems und Schulkontextfaktoren auf die Akzeptanz und Rezeption von zentralen Testrückmeldungen durch Lehrkräfte. *Journal of Educational Research Online*, *3*, 62–93.

MEN (1998a). *Reforma systemu edukacji– projekt*. Warszawa: WSiP.

MEN (1998b). *Ustawa z dnia 25 lipca 1998 roku o zmianie ustawy o systemie oświaty*. Warszawa: Dz.U. Z 1998r. Nr 117, poz. 759.

MEN (1999). *Rozporzadzenie Ministra Edukacji Narodowej w sprawie podstawy programowej*. Warszawa: MENIS.

MEN (2009). *Podstawa programowa z komentarzami. Tom 6: Edukacja matematyczna i techniczna w szkole podstawowej, gimnazjum i liceum*. Warszawa: MEN.

MEN (2013). Programme for international student assessment. Wyniki badania 2012 w Polsce. www.ifispan.waw.pl. Zugegriffen: 14.12.2016.

MSW (2004). *Kernlehrplan für die Realschule in Nordrhein-Westfalen. Mathematik*. Frechen: Ritterbach.

Moleda, A., & Piesyk, Z. (1993). Przeglad zmian programow nauczania matematyki w szkole podstawoej w latach 1963–1990. *Folia Mathematica*, *6*, 25–57.

Neubrand, M., Klieme, E., Lüdtke, O., & Neubrand, J. (2002). Kompetenzstufen und Schwierigkeitsmodelle für den PISA-Test zur mathematischen Grundbildung. *Unterrichtswissenschaft*, *30*, 100–119.

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 und Oldenburg*

Neubrand, M., Jordan, A., Krauss, S., Blum, W., & Löwen, K. (2011). Aufgaben im COACTIV-Projekt: Einblicke in das Potenzial für kognitive Aktivierung im Mathematikunterricht. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 115–132). Münster: Waxmann.

Reiss, K. (2004). Bildungsstandards und die Rolle der Fachdidaktik am Beispiel der Mathematik. *Zeitschrift für Pädagogik*, *50*, 635–649.

Resnick, L. & Ford, W. (1981). *The Psychology of Mathematics for Instruction*. Hillsdale: Lawrence Erlbaum Associates.

Scheja, B. (2007). *Der Wandel von Bildungssystemen am Beispiel von Deutschland und Polen unter besonderer Berücksichtigung der PISA-Studie*. Dissertation, Universität zu Köln.

Schupp, H. (1982). Zum Verhältnis statistischer und wahrscheinlichkeitstheoretischer Komponenten im Stochastik-Unterricht der Sekundarstufe I. *Journal für Mathematikdidaktik*, *3*, 207–226.

Terhart, E., Baumgart, F., Meder, N. & von Sychowski, G. (2009). *Standardisierte Prüfungsverfahren in der Erziehungswissenschaft: Kontext, Formen, Konsequenzen. Erziehungswissenschaft*, 20, 9–36.

UNESCO (2011). *Revision of the International Standard Classification of Education (ISCED)*. 36C/19 of the 36th Session of the General Conference based on 34C/Resolution 20, Paris.

Winter, H. (1995). Mathematikunterricht und Allgemeinbildung. *Mitteilungen der Gesellschaft für Didaktik der Mathematik*, *61*, 37–46.

Wirtz, M., & Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität*. Göttingen: Hogrefe.

Zahorska, M. (2002). *Szkola miedzy panstwem, spoleczenstwem a rynkiem*. Warszawa: Zak.

2.2 Publikation II

Scheja, B. (2016). Cognitive activation through mathematics tasks in the context of centralised examinations on ISCED 2 Level, using the example of North-Rhine Westphalia. *Didactica Mathematicae*, 38, 175 – 202.

Abstract

This study analyses the cognitive activation through mathematics tasks in the context of the centralised examination – Zentrale Prüfung 10 (ZP 10) – administered in North-Rhine Westphalia. Cognitive activation, here, is understood as being one characteristic within a range of curricular, content-related and cognitive categories that can potentially be activated in the course of the solution process of a given task. Here, it is fundamental to first consider obligatory curricular standards as well as to categorise mathematics tasks in terms of their function. This first section is then followed by the analysis of the cognitive requirements of the complete set of test tasks used between 2007 (introduction of ZP 10) and 2016 (N=1706).

DIDACTICA MATHEMATICAE
Vol. 38, 2016, p. 175–202

Bruno Scheja (Köln, Germany)

Cognitive activation through mathematics tasks in the context of centralised examinations on ISCED 2 Level, using the example of North-Rhine Westphalia

Abstract: This study analyses the cognitive activation through mathematics tasks in the context of the centralised examination- Zentrale Prüfung 10 (ZP 10) – administered in North-Rhine Westphalia. Cognitive activation, here, is understood as being one characteristic within a range of curricular, content-related and cognitive categories that can potentially be activated in the course of the solution process of a given task. Here, it is fundamental to first consider obligatory curricular standards as well as to categorise mathematics tasks in terms of their function. This first section is then followed by the analysis of the cognitive requirements of the complete set of test tasks used between 2007 (introduction of ZP 10) and 2016 ($N = 1706$).

1 Introduction

One of the major elements of the changes of educational systems that started internationally in the 1980s is the implementation of centralised (final) examinations, which are in most cases administered at the end of a specific grade (cf. Krüger, 2015; Zabala et al., 2008). The sets of tasks used for these examinations as well as the standards that are at their basis reflect consequences expected by education policy with regard to educational processes in the context of individual subjects. One of the countries that began testing its pupils' performance in the wake of the educational reforms that started in 2001 on the basis of the complex interaction of the factors mentioned above is the Federal

Key words: secondary education, examinations, task design, educational research

Republic of Germany (cf. KMK, 2001). The centralised examinations are gradually administered on both national and federal levels. In the years following the successively implemented reform, Germany's performance in mathematics, as it has been gauged, for example, in the context of the international PISA test (2003-2012), shows an above-average increase (cf. OECD, 2004; OECD, 2014). Against the backdrop of this, it is of special subject-didactic interest to focus on instruments that both capture as well as influence quality, as for example the "centralised examination 10 (ZP 10)", which is obligatorily administered at the end of ISCED (*International Standard Clasiffication of Education*) level 2 (*secondary education*) in North-Rhine Westphalia. It can be assumed that, due to both its function as well as the specific point in time in learners' education biography at which the test is taken, the ZP 10 is a major factor of influence that affects German mathematics education.[1] In the few studies on this controlling instrument that have been published thus far, the focus is most frequently put on systemic, i.e. organisational and administrative considerations (cf. Klein et al., 2014). However, subject-internal aspects that reflect, for example, the new post-reform understanding of modes of thinking and working in mathematics education, are rarely considered (cf. Büchter & Pallack, 2012; Kühn & Drüke-Noe, 2013). A pivotal instrument for standardisation, which can be identified as being a carrier of these modes of thinking and working, are the *tasks* used to test learners. Against this backdrop, the characteristics of the requirements of said tasks, which define their level of difficulty/complexity, and their orchestration as a reflection of their quality are of major importance (cf. Neubrand et al., 2013). The category-system, which is based on the cognitivist-based groundwork done by Neubrand et al. (cf. 2002) and which was developed in the context of the COACTIV-project (cf. Neubrand et al., 2013) and then modified by Drüke-Noe (cf. 2014), is a comprehensive tool that captures the potential for cognitive activation of tasks. On the basis of a *rational task analysis* (cf. Resnik & Ford, 1981), using such a classification scheme allows for a faceted development of the main objective of this paper, which is to capture, anaylse and evaluate the level of cognitive activation of mathematics tasks of the North-Rhine Westphalian ZP 10.

By choosing this objective, the author takes up the desideratum of research outlined above in order to demonstrate and compare a comprehensive, content-related and conceptual profile of test tasks of a educational system that has recently been reformed. The realization of the chosen objective first introduces the discussion on the curricular standards of the ZP 10, or, the

[1] 23% of all German pupils who attended a general education school did so in North-Rhine Westphalia.

scholastic standards for Mathematics for the intermediate secondary school level certificate in the theroetical part of this paper. In the following chapter, tasks, which can be understood as being pivotal vehicles for both the implementation and examination of such standards, will be discussed in terms of their type and their quality as carriers of the potential for cognitive activation. The data collection scheme presented in the chapter on methodology, i.e. chapter 3, which is composed of categories and their respective properties and which is used for the subsequent analysis of the quality of cognitive activation, then allows for the implementation of the main objective of this paper in chapter 4. Finally, the results are summarised and dicsussed in chapter 5. This discussion also includes suggestions which are supposed to contribute to the development of task culture, since test tasks should be closely connected to classroom tasks and exam tasks.

2 Theoretical Background

2.1 Curricular Standards for the Centralised Examination ZP 10 in Mathematics

As part of the seven-step reformation scheme of the Conference of the Ministers of Education and Cultural Affairs – German abbreviation: KMK – nationwide effective scholastic standards that are differentiated for different grades were mandatorily implemented between 2004 and 2005 as primary measures of the fifth field of educational action (cf. KMK, 2001). Connected with their implementation was the goal to "establish a transparent system of Quality Assurance in Germany. Furthermore, they are supposed to improve teaching processes in order to attain higher educational returns" (Blum at al., 2006, p. 9; author's translation). For the intermediate secondary school level certificate (ISCED 2), they were first limited to the subjects German, Mathematics, L2 (English/ French) as well as the scientific domains. Irrespective of the subject, the scholastic standards are based on "educational goals that are broadly understood, embedded in subject matter and reflecting the structure of the respective subject. They are supposed to be attained in school and are furthermore embedded in the framework of interdisciplinary educational goals" (ibid., p. 15; author's translation). In this respect, the scholastic standards are *Performance Standards,* which list precisely those subject-specific competences at the end of secondary level I that are considered pivotal for a seamless continuation of any further educational as well as professional qualification.

With the obligations assumed by the federal states to implement and execute the scholastic standards, the latter affect everyone involved in education on all levels of the education system (macro-level (e.g. federal state level), meso-level (school level) as well as micro-level (school class level)). Here, these standards fulfil a guiding function for the different subjects but are also decisive for internal as well as external evaluations and have evaluation function and monitoring function respectively (Klieme et al., 2003, p. 99f.). The latter imply, for example, standardised tests, which have certification function, at the end of secondary level I in 15 federal states[2]. The concept of the *scholastic standards for Mathematics for the intermediate secondar y school level certificate* is in part based on already established competence models. Here, as with PISA, three different dimensions are distinguished:

1. Comprising what is called *general mathematical knowledge*, the first dimension implies pivotal mathematical activities, meaning the performance while dealing with mathematical content, namely “mathematical argumentation”, “solving problems mathematically”, “modeling mathematically”, “using mathematical representations”, “using mathematics on a symbolical/formal/technical level” as well as “communicating mathematically”. A comparison with the *Process Standards* as the respective standards of the *Principle and Standards for School Mathematics* of the NCTM as the most prominent of its kind shows that *Modeling* is incorporated while *Connections* is not (cf. Reiss, 2004).
2. *Mathematical content competences*, which are distinguished with regard to the following central ideas, i.e. subject matters, “numbers”, “measuring”, “space and form”, “functional correlation” as well as “data and contingency”. Concerning this aspect, these competences and the *Content Standards* of the NCTM are congruent, although the scholastic standards provide requirements that are phrased more explicitly and thus more concrete (ibid.).
3. In regard to the *requirement levels* (*Anforderungsbereiche*), three levels of cognitive requirement are distinguished, which can be identified by referring to the following, already established taxonomies of educational goals when it comes to solving problems: “reproducing”, “establishing connections/correlations” as well as “generalising and reflecting” (cf. Anderson & Krathwohl et al., 2001). From one level to the next, there is an increase in cognitive complexity, which manifests itself in the respective empirical difficulty (cf. KMK 2003, p. 6ff.). Similar nuances regarding the level of requirements cannot be found in the standards of the NCTM.

[2]Only Rhineland-Palatinate refrains from the implementation of such tests (cf. Kühn, 2013).

Instead, they name goals and contents of an "outstanding mathematics education" as ideal standards (Reiss, 2004, p. 637).

This concept of different dimensions of the scholastic standards makes it possible to make the general educational goals of mathematics education, which in part trace back to Freudenthal (1983) but mostly Winter (1995), more concrete. Thus, the function of mathematics to broaden general knowledge as presented by Freudenthal, which, besides formal educational goals, especially emphasizes the didactic principle of connectivity (*Beziehungshaltigkeit*) is explicated. In his explication, Winter assumes that the mathematics education that broadens general knowledge should enable all learners to make three well-balanced *basic experiences* listed and explained in the following table:

Type of basic experience	**Characterisation of basic experience**
Application-Oriented-Learning (Anwendungsorientierung)	Learners should be enabled to "perceive and understand technical, social and cultural phenomena and processes with the help of mathematics and to evaluate them with the aid of mathematical principles" (KMK, 2003, p. 6; author's translation).
Structure-Oriented-Learning (Strukturorientierung)	Learners should be enabled to "know and understand mathematics and its language, symbols, images and formulas and their significance for the description and performance/solution of tasks and problems within and outside the field of mathematics" (ibid.; author's translation).
Problem-Based-Learning (Problemorientierung)	Learners should be enabled to "acquire a general capacity to solve problems while performing tasks and solving problems with the aid of mathematical means" (ibid.; author's translation).

Table 1. The three basic experiences according to Winter (cf. 1995).

In terms of concept and design, the fundamental significance of these basic experiences for the scholastic standards and thus the respective processes in mathematics education is reflected in the orchestration of the competences imparted to learners. Hence, a unilateral orientation towards rules and algorithms in the context of mathematics education carries the risk of learners "developing a unilateral or even contorted conception of mathematics since the three *basic experiences* as put forward by Winter are not sufficiently taken into account in a balanced manner" (Drüke-Noe, 2014, p. 25, author's translation). With regard to the tasks used in mathematics education as pivotal means to control teaching-processes in the classroom (reflecting standards) as well as the quality of teaching (capturing the implementation of standards), this means: As key elements of control, regardless of their function, – i.e. both as learning tasks as well as assessment tasks – they should appropriately reflect the given standards for mathematics education. This is primarily reflected in a well-balanced distribution of the three basic experiences in the established requirement profile.

2.2 Types of Mathematical Tasks Against the Backdrop of Cognitive Activation

In the context of everyday teaching, tasks always play an important role, although in mathematics education they mostly dominate in situations of learning and testing. Here, they always ask learners "to specifically deal with a clearly defined mathematical issue" (Neubrand, 2002, p. 16). Whether they are learning tasks or assessment tasks, they have various structural similarities and it is thus appropriate to categorize them in terms of their intended purpose (cf. Drüke-Noe, 2014, p. 8f), i.e. in terms of their function.

In situations of learning, *classroom tasks* (*Unterrichtsaufgaben*) are an interface of the activity and communication between teachers and learners and thus a pivotal tool for the preparation and conduct of lessons. Teachers use them to control and manage their lessons and hence their learners' activities. In the context of this process, tasks carry mathematical contents and can therefore be understood as being the main means of control for the progression of the development of competences. This is mostly ensured by tasks, which, according to Jordan et al. (2006, p. 13), can be used as flexible, widely applicable and actively controllable means of controlling didactic aspects and content. Given an appropriate level of awareness, teachers can also influence learners' understanding of a mathematical term or process, the structure of a dense network of terms and ultimately their overall conception of mathematics (Neubrand et al., 2013, p. 126). The tasks used in class are hence often a means of identifying the requirements set out for and experienced by learners. These tasks set a framework for the formation of mathematical skills, which can be consciously experienced by learners while being actively involved in the solution process. In the process of building knowledge, classroom tasks are thus a vehicle of learners' cognitive activities.

Assessment tasks, on the other hand, are a means to attest "which skills have been acquired in the course of the education process" (Terhart et al., 2009, p. 24; author's translation). They are usually used for examinations or other forms of assessment in the context of internal and external evaluation of learner-performance. Internal evaluation comprises, for example, class-tests or informal tests that are used for individual classes or forms as a whole respectively; external forms of evaluation are then based on/guided by overarching standards (used for the assessment of learners of one federal state or assessment on a national level respectively as well as final examinations). One major characteristic of assessment tasks of the first category is that the teachers, who base their decision on the type of classroom tasks used prior to the assessment of performance, always choose and design them. Hence, these tasks make it possible to draw valid conclusions about the main emphasis chosen during as well as the quality of the mathematical requirements that characterises the lessons taught prior to the respective assessment (as can be seen, for example, in the context of the COACTIV-study (cf. Kunter et al., 2013)). Furthermore, they provide suggestions for possible action or behaviour in regard to the respective group of learners. Studies on the implementation of the diagnostic potential of these tasks, however, reveal that this potential only selectively leads to consequences on teaching. Assessment tasks that are part of external forms of assessment - mostly *Large Scale Assessments* – however, are designed by exteranl institutions and should fulfil the quality criteria of test theory (objectivity, validity, reliability). Depending on their function, the development of tasks requires the variable operationalisation of requirenments as presented in the standards (cf. Büchter & Pallack 2012). In the case of German, national forms of assessment, these tasks are based on standards (under the overall control of the IQB/ Institute for Quality Development), i.e. individually and in combination as tests, they are in line with the competence standards outlined above. Assessments conducted in individual federal states, however, reflect the requirements of the respective curricula (in North Rhine-Westphalia these standards are under the overall control of QUA-LIS/ Agency for Quality and Support of the State Institute for Schools). Following the resulution of the KMK, these should be in line with the educational overarching standards. Thus, the tasks used for these forms of assessment are indirectly based on standards. Through thorough analyses of the structure and requiremens of tasks, Drüke-Noe has shown that, contrary to expectations, classroom tasks as well as assessment tasks "mostly have identical cognitive charactersitics" (Drüke-Noe, 2014, p. 9), so that the difference only becomes apparent in how teachers deal with and assess learners' results. Here, classroom tasks allow for deviations and divergenes to happen while assessment tasks are designed with a focus on performance and results and thus require convergence (cf. Büchter & Leuders, 2005, ibid). In summary, it can be said that the assessment tasks to be analysed in this study can be understood as also being carriers of cognitive activation. With regard to the task type analysed in this study, cognitive activation means the facets of cognitive processes that unfold while solving the task at hand within the "anticipated effect of teaching" (Neubrand et al. 2013, p. 129).

As a means to answer the pivotal research question of this study, one objective is to classify the tasks of ZP 10 in a way that their potential for cognitive activation is revealed. The evaluation categories used to realise this objective are empirically tried and tested and are taken from relevant pedagogical/pschological literature. These categories make it possible to and analyse cognitive activation guided by three questions:

- How are the ZP 10 tasks distributed across the different subject areas and how can the curricular level of requirement be evaluated?
- How are the ZP 10 tasks distributed across the different types of mathetmatical activities?
- How can the level of cognitive requirement of the ZP 10 tasks be evaluated?

3 Method

3.1 Methodological aspects of this study

By focussing on the idea of cognitive activation as well as by employing a certain method of classifying tasks, this study links to essential characteristics of the COACTIV project and its conceptional extensions.Thus, after presenting the data sets in 4.1, the conceptional basics and the method of classifying tasks will be explained in 4.2. Finally, the process of coding is will be reported in 4.3.

3.1.1 Sample and Dataset

The dataset analysed within the scope of this study is the set of tasks used for the centralised examination that is part of tenth grade (ZP 10) and conducted in all of North Rhine-Westphalia between 2007 (introduction of the ZP 10) and 2016. It comprises the tasks for all ISCED 2 school types (N=1706) and consists of the respective task sets used for *Hauptschule* (consisting of subsets of tasks for “Hauptschulabschluss” after year 10 (secondary school qualification) and “Realschulabschluss” (secondary school certificate)), *Gesamtschule* (“Elementary Course” and “Advanced Course”), *Realschule*, *Abendrealschule* (night school) as well as *Gymnasium* (here, this test was only conducted from 2007 to 2010).

3.1.2 Task Categories

3.1.2.1 Task Categories Based on the COACTIV-Study

In order to capture the level of cognitive activation of the tasks used, the author draws on a range of empirically established categories for classification drafted in the context of the COACTIV-project (Professional Competence of Teachers, Cognitively Activating Instruction, and the Development of Students' Mathematical Literacy) (cf. Kunter et al., 2013) and developed further in the context of the analyses of tasks by Drüke-Noe (cf. 2014).

Successively extending the conceptional groundwork done by Neubrand et al. (2002), both studies extensively analyse, among other things, classroom tasks and classtest tasks respectively, which were used in grades nine and ten at German schools. The classification scheme chosen for this paper subsumes assessment criteria of tasks under three dimensions/ frameworks: content framework, cognitive framework as well as cognitive elements of the mathematical modeling cycle. These dimensions are further differentiated into categories with, mostly, four different property values (cf. Table 2). The values of these categories, as for example in the case of "topic area," reach from highly inferential to lowly inferential ratings (cf. Rosenshine, 1970).

In the context of the dimension "Content framework", a task is allocated one or more *topic areas*, which constitute the framework of content knowledge in the course of its processing. The category *Curricular knowledge level*, then, captures cognitive activation so that, through the respective allocation, the periods in which a topic is covered, the range of different topic areas and thus the curricular requirement level are explicitly taken into account (Jordan et al., 2008, p. 87). This criterion is captured – and this should be noted explicitly – by using German mathematics curricula as a guidance for rating (cf. Jordan et al., 2006, p. 29).

Falling under the category "Type of mathematical activity", the subsequent dimension, "Cognitive framework", captures the task type by showing the type of the corresponding solution process. First, there is a disticntion made between different types of solutions, which are characterised by the employment of one the following forms of modeling: 1. *intra-mathematical modeling*; 2. *extra-mathematical modeling*. The first merely requires the contextless application of skills or factual knowledge to already given assumptions (cf. Fig. 1), whereby processing forms of performance such as modeling or structuring are not required for the solution.

Dimension	Category	Properties
Content framework	Topic area	1 – arithmetic, 2 – algebra, 3 – geometry, 4 – stochastics
	Curricular knowlegde level	1 – elementary knowledge (level I), 2 – basic knowledge at lower secondary level (level II), 3 – advanced knowledge at lower secondary level (level III)
Cognitive framework	Type of mathematical activity	1 – technical task, 2 – procedural task, 3 – conceptual task
Elements of the modeling cycle	Extra-mathematical modeling	0 – not required, 1 – standard modeling, 2 – multistep modeling, 3 – reflection on a model, development and validation of complex models
	Intra-mathematical modeling	0 – not required, 1 – standard modeling, 2 – multistep modeling, 3 – reflection on a model, validation, strategy development
	Processing mathematical texts	0 – not required, 1 – direct text comprehension, 2 – text comprehension with reorganisation, 3 – comprehension of logically complex texts
	Mathematical argumentation	0 – not required, 1 – standard reasoning, 2 – multistep argumentation, 3 – development of complex argumentation, proofs, evaluation of arguments
	Using mathematical representaions	0 – not required, 1 – standard represenataions, 2 – switching between different representations, 3 – evaluating representations

Table 2. Pivotal assessment criteria of this study (cf. Neubrand et al., 2013, p. 134).

Task 3	**Task 4**
Solve the following equation:	Solve the following equation system:
$2x + 1.8 = 3.6 - x$	(I) $2x + y = 2$ (II) $x - 0.5y = 2$

Figure 1. Examples of a technical tasks taken from the examination administered in 2015 (MSW, 2015).

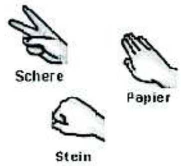
Mathematical modeling tasks, by contrast, require learners to transfer a problem to a processing ansatz and to refer back the solutions to the given initial situation. Thus structured solution processes – exemplarily understood as modeling cycle (cf. Jordan et al., 2006, p. 31) – are reflected in procedural and conceptual modeling tasks. The solving of procedural modeling tasks primarily requires procedural and algorithmic thinking/knowledge, while the processing of conceptual modeling tasks mainly requires conceptual thinking/knowledge (ibid). Here, both types of tasks can be embedded into a context located either inside or outside the field of mathematics. Thus, the three task types cover different fields of mathematical competence. In accordance with section 2, it is, among other things, the goal of German mathematics education to ensure that the different ways of thinking/types of knowledge that underlie these three task types are adequately reflected in the tasks used in class.

The categories of the third dimension "Cognitive elements of the mathematical modeling cycle" operationalise the requirement level of the cognitive processes and activities corresponding to the five types of mathematical activities considered in this study. For this, it is of fundamental importance that the task context presenting a mathematical problem – whether it lies inside or outside the field of mathematics – is transfered into a processing model with the same structure (*intra-mathematical modeling* and *extra-mathematical modeling*), in which processes such as translating/ transforming, processing or interpreting/validating can occur in a cognitively analogous manner, although with different characteristics (ibid., p. 32). The performance of these two activities can occur – analogous to *mathematical argumentation* as well as *using mathematical representations* – on four different levels of complexity: 0. no allocation of a property value 1. low, 2. moderate, 3. high (cf. Neubrand et al., 2013, p. 133). For the mathematical activity "mathematical argumentation", which, in the anlysis of PISA tasks, has also been verified to be a characteristic that generates complexity, these are operationalised as follows:

- 0. no argumentation required;
- 1. Standard reasoning: "Simple reproduction of standard forms of argumentation; perform forms of argumentation that only require common knowledge; developing one-step or purely procedural forms of argumentation" (Jordan et al., 2006, p. 40/ author's translation);
- 2. Multistep argumentation: "Developing straightforward multistep and also conceptual mathematical forms of argumentation and presenting them in writing or, if required, understanding them" (ibid./ author's translation);
- 3. Devoloping of complex argumantation, proofs: "Developing complex mathematical forms of argumentation (explanations, proofs, strategies, generalisations) and presenting them in writing or, if required, understanding them; comparing or evaluating different types of mathematical argumentation (ibid./ author's translation).

The category *Processing mathematical texts* captures the skill level that is required to extract information from the text of a given task, which are necessary for structuring the solution process. Here, test tasks 3 and 4 (Fig. 1), for example, requires the lowest level of verbal argumentation (property = 0) while the level that is required for the task shown in Fig. 2 can be classified as moderate.

“**Part 2, task 3**
Anne and Paul are playing ”Rock-Paper-Scissors”. After counting to three, both form one of the three signs with their hands. The following applies:



- Rock beats scissors because it blunts the scissors.
- Scissors beats paper because they cutt he paper.
- Paper beats rock because it wraps the rock.
- If both players form the same sign, the round ends with a draw.

In the following, it is assumed that both players make their decisions by pure chance. (. . .)

b) 1) Fill in the table below: Who wins which game? Where does the game end in a draw?”

Paul

Anne

	Stein	Schere	Papier	
Stein		Anne gewinnt		
Schere				
Papier				

Figure 2. Example of a task taken from the task set used in 2013 (cf. MSW, 2013, p. 4).

An analysis of the task reveals that the sequence of the (parts of the) sentences, in contrast to a task with a “low requirement level”, are not in accordance with the steps necessary to process the task mathematically. The logical connections are hidden under the linguistic structure, i.e., the main and subordinate clauses used in the task have linguistic references. Here, "if-then-connectives" (conditionals) and universal quantifications, respectively, are fundamental logical functions.

3.1.2.2 On the Necessity of Extending the Assessment Scheme

A comparison of the test task as presented in Fig. 3, in which learners are asked to solve a general quadratic equation and the task 3 presented in Fig. 1, shows that they, due to their similarity with regard to the type of mathematical activity required for the respective solution, can both be allocated to the same task type, namely "technical task", although the process of finding a solution in accordance with the requirements of the task is more "difficult".

Find the solutions to the following equation:

$4x^2 + 4 = 8x$

Figure 3. Example for a technical task (Drüke-Noe, 2014, p. 69).

In the scheme presented above, their potential for cognitive activation, unfolding in the course of the solution process, is uniformly captured with congruent property values: 2 (topic area), 3 (curricular knowledge level), 1 (Type of mathematical activity), 0 (Extra-mathematical modeling), 0 (Intra-mathematical modeling), 0 (Processing mathematical texts), 0 (Mathematical argumentation), 0 (Using mathematical representaions) (cf. Table 3).

The analytical category *working technically* ("Technisches Arbeiten") developed by Drüke-Noe amends the scheme by explicating necessary activities and their requirements in terms of technical skills. Since both technical tasks as well as tasks, whose solution process can be described by means of a modeling cycle, require learners to deal with calculi of varying complexity (cf. Fig. 4.4 in Drüke-Noe, 2014, p. 71) it is possible to apply the category to all three types of tasks. Here, the category also captures the requirement level of dealing with calculi, while considering different steps of the solution process (cf. Fig. 4.3 . 70 and Fig. 4.4 p. 71, Drüke-Noe, 2014). In accordance with the classification scheme as presented in Table 2, this dealing with calculi is required in the context of one of the four topic areas (Arithmetic, Algebra, Geometry, Stochastics). Here, Drüke-Noe designates four different levels:

Activity/ Category	Topic area	Level	Properties and examples
Working technically	Arithmetic	0	Not required, e.g. a
		1	Only multiplication and division calculation/ only addition and subtraction calculation, e.g. simple exponentiations (positive base and power)
		2	Simple hierarchical operations (multiplication and division calculation before addition and subtraction calculation, exponentiation before multiplication and division calculation), e.g. multiplying/dividing exponentiations (positive base and power), extracting the root of a given number
		3	Complex hierarchical operations (Exponentiation with multiplication and division calculation/ addition and subtraction calculation)
	Algebra	0	Not required, e.g. Is the mapping. . . a function?
		1	Only multiplication and division calculation/ only addition and subtraction ccalculation, basic multiplication and division calculation before addition and subtraction calculation, e.g. calculating with a linear equation
		2	Simple hierarchical operations (complex multiplication and division calculation before addition and subtraction calculation, exponentiation before multiplication and division calculation) e.g. complex multiplication and division calculation before addition and subtraction calculation, solving quadratic equations, squaring a sum/factorising, calculating with trigonometric functions
		3	Complex hierarchical operations (Exponentiation with multiplication and division calculation/ addition and subtraction calculation), e.g. solving mixed quadratic equations or exponential equations
	Geometry	0	Not required, e.g. what is the measure of a right angle?
		1	Only multiplication and division calculation/ only addition and subtraction ccalculation, e.g. calculating sin(23) , calculating by using the intercept theorem
		2	Simple hierarchical operations (i.a. exponentiation before multiplication and division calculation), e.g. calculating angle size by using the law of sines
		3	Complex hierarchical operations (exponentiation and multiplication and division calculation combined), e.g. calculating the face or volume of solid figures
	Stochastics	0	Not required, e.g. noting all combinations
		1	Only multiplication and division calculation or only addition and subtraction ccalculation, e.g. calculating simple probabilities
		2	Simple hierarchical operations (multiplication and division calculation before addition and subtraction calculation, exponentiation before multiplication and division calculation), e.g. calculating the arithmetic mean
		3	Complex hierarchical operations (exponentiation combined with multiplication and division calculation and addition and subtraction calculation), e.g. calculating probability in the context of the multistep Bernoulli experiment

Table 3. Descriptions of the different levels of working technically (taken from Drüke-Noe, 2014, p. 78/ author's translation).

Exceeding the the limits of the COACTIV approach, a thus operationalised mathematical activity now makes it possible to locate, for example, the requirement level of the two technical tasks presented at the beginning with regard to the "technical work" required by said tasks. Hence, they are both first allocated to the topic area "algebra" and the category "Advanced knowledge at lower secondary level (level III)". By reference to the corresponding, ideal solutions, it becomes clear that, in the first example (cf. Fig. 3), the hierarchical structure of the terms on both sides of the equation can be identified first. Dealing with the mixed quadratic equation requires a) different mathematical operations (general: $ax^2 + bx + c = d$, with a, b, c, d being rational numbers and a$\neq$0) and, irrespective of the solution procedure chosen, b) the understanding use of powers and roots (of sums). In this respect, these are "Complex hierarchical operations of (exponentiation combined with multiplication and division calculation and addition and subtraction calculation) – level 3", since finding the solution requires the use of priority rules, such as "exponentiation before multiplication and division calculation". In the second example (task 3, Fig. 1), the multiplicative-additive structure of the terms on both sides of the equation can be identified first, i.e. before appropriately using sums and products and before finally isolating x by means of dividng two numbers. This designates precisely the technical requirements of level 2.

3.1.3 Coding

The subsequent coding of the ZP10-tasks was done independently by the author (coder 1) and an experienced maths teacher (coder 2), who was thoroughly introduced to the conception of the study beforehand. Since the reliability of the assessment is of major importance, both coders practiced coding with tasks taken from the COACTIV-study. In the context of the COACTIV-project, these taks had already been allocated to COACTIV- assessement categories (cf. Table 2). Moreover, both coders also had the coding of "working technically", done by Drüke-Noe and her colleagues, as a basis.

In a further step, it was of particular interest to first identify and quantify the variability of the *interrater agreements* in order to obtain indications from the agreements as to the stability and reliability of the coding-trial. The Cohen's kappa coefficient k has been applied to measure the quality of agreement. $0.7 < k$ has been defined as the minimum value of reliability and thus for proceeding to the coding of the tasks used in the test conducted at ZP 10. Following Wirtz and Caspar (2002), $0.75 < k$ is understood to be a "very good", $0.6 < k < 0.75$ a "good" and $0.4 < k < 0.6$ an "acceptable" delineation regarding the concordance of the respective assessments. Table 4 shows the

Kappa values for the different assessment categories and their properties used in this study that were determined in the context of the coding-trial.

	Kappa values k ($N = 291$)		
Category/property	Coder 1/ COACTIV-Coder	Coder 2/ COACTIV-Coder	Coder 1/ Coder 2
Arithmetic	0.71	0.73	0.7
Algebra	0.74	0.69	0.71
Geometry	0.89	0.91	0.88
Curricular knowledge level	0.93	0.86	0.83
Type of mathematical activity	0.79	0.77	0.82
Extra-mathematical modeling	0.94	0.93	0.94
Intra-mathematical modeling	0.92	0.95	0.9
Processing mathematical texts	0.84	0.71	0.73
Mathematical argumentation	0.95	0.9	0.93
Using mathematical representaions	0.98	0.99	0.99
Working technically[3]	0.89	0.76	0.82

Table 4. Measure of coding-agreement within the framework of the assessment categories of this study.

The results in the respective columns show that the quality of agreement of the coding-trial ($N = 291$) for all categories/properties can be described as being at least "good". The $k - v$alues for the subjetc areas "Arithmetic" and "Algebra" are good, while the values for "Geometry" suggest a very good quality of agreement. The latter is also true for the allocations in the category "Curricular knowledge level". For both types of mathematical activity – "extra-mathematical modeling" and "intra-mathematical modeling" – the $k - v$alues are consistently very good, while those for the category "Processing mathematical texts" suggest a good quality of agreement. The category "working technically", which extends the assessment sheme, is characterised by a quality of agreement that can be described as "very good" since here, the $k - v$alues are consistently higher than 0.75. The coding of the last two categories of the dimension "Cognitive elements of the mathematical modeling cycle" can even be described as being almost matching. Overall, it can be noted that the coding of the North-Rhine Westphalian set of tasks can sufficiently be connected to the coding of the categories of the COATIV-scheme and the category "working technically".

[3]This ist the measure of agreement between coder 1 and 2, respectively, and the assessments of Drüke-Noe and her colleagues.

4 Results

Guided by the three pivotal questions of this study (cf. 3.), the following chapter presents the results of this study on cognitive activation through North Rhine-Westphalian ZP 10 tasks used between 2007 and 2016 ($N = 1706$).

4.1 Cognitive activation through ZP 10 tasks

First, the topic areas and the curricular level of requirement will be analysed in section 5.1. This analysis is then followed by an examination of the types of mathematical activities and hence the cognitive framework of the tasks (5.2). The chapter is then completed by an in-depth look at the cognitive requirements of the test tasks (5.3).

4.1.1 Content-related activation of ZP 10 tasks

A major goal of central examinations is to gather information about tenth graders' knowledge levels and skills with regard to the standards of the scholastic standards for the *Mittleren Schulabschluss* (General Certificate of Secondary Education (GCSE)). The "cenral ideas" listed within the dimension "content-related competences" (cf. section 2.1) have been captured in terms of the following topic areas: Arithmetic, Algebra, Geometry und Stochastics. First, table 4 shows that the tasks are unevenly distributed across all four topic areas. At percentages between 26.7% and 31.6%, it is thus primarily the topic areas of Arithmetic, Algebra and Geometry that are tested predominantely as well as equally frequent. With 14.9% of the tasks, the topic area of Stochastics constitues – as expected – the smallest percentage.

When analysing these topic areas with regard to the curricular requirement level of the mathematical knowledge tested in said areas, it becomes evident that the intermediate level with $M = 2.43$ is between the "basic knowledge at lower secondary level (Level II)" and the "advanced knowledge at lower secondary level (level III)". In particular, it is furthermore significant that tasks, which exclusively require the elementary knowledge of primary school (with four forms), barely carry any weight. Moreover, the third curricular knowledge level is covered slightly less than the second level. This indicates that – contrary to expectations[4] – mostly tasks whose curricular requirements can be classified as belonging to the periods of teaching before 10^{th} grade are

[4]According to MSW, the ministry in charge for NRW, "the second, more extensive part of the test should include tasks that require competences of forms 9/10 and should be based on major topics relevant for said forms" (https://www.standardsicherung.schulministerium.nrw.de/zp10/fragen-und-antworten/, 21.10.2016, author's translation).

used for the test. Overall, about six out of ten test tasks can be allocated to the category “Basic knowledge of lower secondary education (Level II)”. However, table 4 shows that, in the context of the different topic areas, there are considerable differences as to their curricular requirement level. Here, the results show that tasks that belong to the topic areas Arithmetic ($M = 2.11$, $SD = 0.30$) and Stochastics ($M = 2.29$, $SD = 0.37$) rarely reach a level higher than that of “lower secondary level”, while the tasks that belong to the topic areas Algebra ($M = 2.52$, $SD = 0.57$) and Geometry ($M = 2.51$, $SD = 0.51$) do test knowledge of a wider range of abstraction. Irrespective of the curricular requirement levels, no form of anticipation within the scope of the curriculum could be noted in any of the topic areas. So far, the descriptions show that the average curricular requirement level of the North-Rhine Westphalian test tasks ranks between the low characteristic values of PISA 2003-national and the values achieved for class tests conducted in tenth grade (type of school “Gymnasium”: $M = 2.80$; type of school “not Gymnasium”: $M = 2.67$) in the context of the COACTIV-study (cf. Drüke-Noe, 2014, p. 103). This also shows one of the functions of this instrument within the system of external performance assessment: without a clear focus, the ZP 10 tests the knowledge and skill level of the education period between fifth and tenth grade (= second and third level of curricular knowledge).

Topic area	**Number of items ($N = 1706$)**	**Percen-tage**	**Curricular know-ledge level M**	**Standard devia-tion SD**
Arithmetic	492	31.6	2.11	0.30
Algebra	462	29.7	2.52	0.57
Geometry	416	26.7	2.51	0.51
Stochastics	233	14.9	2.29	0.37
	Sum of allocations with $N = 1706$: 1766[5]	104	2.43	

Table 5. Distribution of the ZP 10 tasks across the four topic areas and the curricular knowledge levels.

[5]This number ist the result of multiple allocations within the different categories.

4.1.2 Type of mathematical activity

When looking at the cognitive processes that characterise the type of mathematical activity in the course of the solution process, it can be noted that procedural modeling tasks constitute a clear focus of testing (82.4%). At 9.0%, conceptual modeling tasks that require learners to find connections/ correlations between/of mathematical concepts only constitute a small proportion. Conceptual-procedural modeling tasks constitute a total of 91.4% of all test tasks. Here, application-oriented, problem-solving contexts constitute a clear majority. Only 134 of the 1706 test tasks require (cf. Table 6) learners to apply algorithms/ numeric methods of calculation to a given approach, i.e. to simply "handle them technically".

	Type of mathematical activity		
Property	Technical task	Procedural/procedural modeling task	Conceptual modeling task
Percentage with $N = 1706$	8.6	82.4	9.0

Table 6. Distribution of tasks across the different types of mathematical activity.

The distribution of the tasks across the different types of tasks indicates the focus on cognitive activation that seemed desirable for the Ministry of Education in the years following the introduction of the ZP 10. As part of centralised examinations, then, an apparent departure from a traditional self-image of mathematics education took place. It was considered to "be focussing to much on the acquisition of calculating skills and on mastering standard procedures (...) while often neglecting the development of flexibly applicable concepts and skills, which are especially required when handling transformation processes in new contexts (*Sachsituationen*)" (Blum et al., 2004, p. 156/ author's translation). Within the scope of the set of tasks analysed here, mathematical terms and procedures are almost exclusively used as "tools" for various, mostly application-oriented context problems. However, the comparison of this distribution of tasks (see Table 5) with that of both PISA tests *(PISA 2000 international (Literacy-Concept)*: technical task: 3.2%; procedural modeling task: 45; conceptual modeling task: 52%/ *PISA 2000 national*: technical task: 26.7%; procedural modeling task: 38.4%; conceptual modeling task: 34.9%) reveals the partial existence of considerable differences in terms of their respective concept (cf. Neubrand et al., 2002, p. 101).

In comparison to the two test concepts, the conceptual modeling tasks are underrepresented while procedural modeling tasks are overrepresented. The proportion of technical tasks is only small when compared to the national PISA tasks. Such a focus on algorithmic-procedural modes of thinking in the time "after PISA", hence, preserves an imbalanced, although currently different, range of thinking skills within this category. Considering the approved three *basic experiences* by Winter (cf. Section 2), which should be the basis of the scholastic standards and thus of all teaching and learning processes, it becomes clear that neither the traditional/process-oriented focus nor the algorithmic-procedural focus are really desirable. Given the strong focus on *Application-Orientation* in the context of the ZP 10 tasks, it can be assumed that the instrument does not sufficiently capture the *basic experiences* in mathematics education

4.1.3 Cognitive Requirements of ZP 10 tasks

The types of mathematical activity considered earlier (see 5.2) show the characteristics of the solution process. However, they do not show the level of cognitive requirement that is necessary to perform said process. In order to present the analytic dimesion in a nuanced manner, both the attained average level of requirement as well as the distribution across the corresponding levels (0-3) are considered for every category. Table 7 shows the distribution of these levels across the mathematical activities within the scope of the North-Rhine Westphalian set of tasks.

	Level of the respective mathematical activity				
Mathematical activity	0	1	2	3	arithmetic mean
Extra-mathematical modeling	24.8	32.5	41.9	0.8	1.187
Intra-mathematical modeling	83.8	4.3	11.2	0.7	0.288
Mathematical argumentation	90.1	5.1	4.8	0.000	0.147
Using mathematical representaions	35.3	56.5	8.2	0.000	0.729
Working technically	19.5	46.3	19.0	15.2	1.299
Processing mathematical texts	21.5	63.0	15.5	0.000	0,94

Table 7. Distribution of the levels of requirement across the mathematical activities.

The table above illustrates that the analysis of transition processes between reality and methematics – i.e. *extra-mathematical modeling* – constitutes the majority (74.4%) of all tasks used for the North-Rhine Westphalian examination ($M = 1.187$). Here, intermediate mathematical modeling tasks constitute a majority (41.9%), while basic as well as challenging extra-mathematical modeling tasks constitute 32.5% and, repsectively, 0.8% of all tasks. Due to the results for the type of task addressed in the previous section - where it became apparent that only 8.86% of the tasks fall under the category *technical task* – the difference of 16.2% is attributable to the category *intra-mathematical modeling* ($M = 0.288$). Analogous to the category "extramathematical modeling" the following is true: If a task requires mathematical modeling, the requirement level is most frequenlty "intermediate" and only rarely "elementary" or "advanced". On this basis, looking at just the cognitive processing of the test tasks (modeling task vs. technical task) while also considering the required levels of cognitive activation, it becomes apparant that even 87.8% of all tasks require either an elementary (36.8%) or an intermediate (53.1%) level of modeling.

Due to the fact that most of the tasks can be categorized as "closed tasks", only 9.9% of all tasks require learners to name closed chains of argumentation or to evaluate already given *argumentations* respectively. The 169 (9.9% of all tasks) tasks are distributed almost evenly across the elementary and the intermediate requirement level (standard and multistep argumentations). In 64.7% of all test tasks, *mathematical representations* are used ($M = 0.729$). They are mostly iconic representations of a low requirement level (56.5%). Here, similar to all the categories considered so far, it can also be noted that the concept does not incorportate cognitively challenging mathematical representations. On average, *working technically* is not required in one out of five tasks ($M = 1.299$). However, if it is needed for the solution process, it most frequently occurs on an elementary/basic level (46.3%) while the intermediate and the low requirement level are allocated to 19.0% and 15.2% of the tasks respectively. With regard to the requirement levels of the category *Processing mathematical texts*, the tasks are distributed similarly. Apart from "extra-mathematical modeling" and "using mathematical representations", they show the highest average in this category ($M = 0,94$). Furthermore, a comparison of the corresponding arithmetic means of the categroy with those of comparable sets of test tasks shows that theese arithmetic means (cf. Drüke-Noe, 2014, p. 160f; Neubrand et al., 2013, p. 128) can be classified as cognitively activating. This is surprising insofar as the tasks used only require, on average in eight ot of nine cases, the "elementary level", which imposes rather low requirements on learners.

5 Discussion and outlook

5.1 Summary and discussion of findings

Overall, the results of this study on cognitive activation show that North-Rhine Westphalian mathematics education of the ISCED level 2 between 2007 and 2016 have a comparatively high potential for cognitive activation, which is, however, relatively inhomogeneous. With regard to curricular characteristics of the test tasks analysed (first research question) it is characteristic that it is mainly the topic areas algebra, arithmetic and geometry that are tested, while stochastics is only considered marginally, as is rue for the centralised examinations in other federal states (cf. Kühn & Drüke-Noe 2013). Against the backdrop of the defined goal of the ZP, the average level of curricular requirement within the realm of the four topic areas can be said to be rather low. The analyses in the context of the second question of this paper, in which the types of mathematical activities in the solution process have been examined, mostly showed that tasks whose solution require modeling processes, predominate considerably. Here, especially procedural and – with certain limitations – conceptual ways of thinking are used as a "tool" for inner-mathematical and especially various extra-mathematical contexts. In contrast, technical tasks, i.e. the task type that is used/employed most frequently in everyday mathematics lessons, are only considered marginally (cf. Jordan et al. 2008). Finally, the analysis of the levels of cognitive requirement in the context of the third research question showed that, with regard to the distribution of relevant test contents and the existing range of mathematical activities (intra-and extra - mathematical modeling, processing mathematical texts, mathematical argumentations, mathematical representations and working technically), the tasks show distinctive characteristics of broadly activating Large Scale Assessment-tasks (of which the international PISA test is one example/ cf. Neubrand et al., 2013).

Against the backdrop of these insights, it seems reasonable to suggest that both the unfolding of effects of any changing mathematics education that is embedded in the scholastic standards/core curriculum – e.g. with regard to cognitive activation (cf. Maier et al., 2011) – as well as the connection of test results with qualification systems concerning the subsequent ISCED 3 level, are a possible explanation for the increased mathematics performance in the PISA test.

However, it must be critically stated that the ascertained focus of the conception of ZP 10 tasks only slightly reflects the fundamental standards designated by the *scholastic standards*, which can be connected to already existing findings of subject-didactic research (cf. Kühn & Drüke-Noe, 2013) and implementation research (cf. Gräsel, 2010). This is, for example, reflected by the fact that the breadth of the three *basic experiences* of mathematics education that broadens general knowledge, as put forward by Winter, is disproportionately embedded in several respects. Considering the high proportion of extra-mathematical modeling, the basic experience of "Application-Orientation" can be described as being excessively implemented. In contrast, both intra-mathematical modeling as well as contextless technical tasks contribute insufficiently to the capture of the basic experience "structure orientation". An analogous limitation of the construct validity becomes apparent in the low proportion of Winter's "Reflexion on one's *Denkhandeln* (*internal acting*)" (1995, p. 41) in the form of mathematical argumentation, which is in turn a substantial aspect of the basic experience of "problem-orientation".

With regard to the implementation of the scholastic standards in the conception of the tasks analysed in this paper, it must be critically stated that the requirement levels of all mathematical avtivities are unevenly distributed across their complete range of requirements. Primarily, this becomes clear when considering the fact that tasks that can be allocated to the highest, reflecting and generaliszing requirement level, are missing. From a diagnostic point of view, this also raises the following question: To what extent does the ZP 10 adequately fulfil its main purpose - i.e. testing skills and knowledge in accordance with the scholastic standards - especially when it comes to learners who show a high level of performance? This obvious lack of veridicality – one of the main catalysing variables for leraning success (cf. Helmke, 2004, p. 94) – thus, seems to encompass the feedback currently provided by the ministry, which reflects teaching success according to set competence standards, shows necessities for affirmative action and helps teachers to deduce measures for individualized education. A question that is both open as well as of great interest is the one concerning the effect of a thus limited requirement distribution in the context of mathematical activities on the controlling of teaching on the meso- and micro level of the North-Rhine Westphalian education system. Here, it seems first of all possible to interpret test tasks themselves, rather than curricular standards, as summative tools used to standardise teaching and that washback effects with regard to lesson design will be identifiable. Clear indications for the effect of this decoupling are, on the one hand, provided by already existing research fidings (cf. Maier et al., 2011) and, on the other hand, by the existence of an already developed, diverse supplement industry, which provides teachers with i.a. workbooks, practice tests as well as material for planning und teaching lessons.

5.2 Outlook

From the perspective of the basic experiences according to Winter as well as from the perspective of a more comprehensive assessment of learners' performance, there emerge possible approaches to an improvement of ZP 10 task culture. These approaches include the establishment of a higher level of cognitive challenges through tasks as well as a form of learning that is directed towards conncetion and comprehension. Implementing these approaches would enable mathematics to contribute profoundly to general education/knowledge

(*Allgemeinbildung*), which should be given expression in the designing of lessons. Because, against the background of the well-documented positive effect of final examinations on the conception of class tests with regard to content-related *and* cognitive characteristics of tasks, the conception of centralised tests according to standards is pivotal (cf. Drüke-Noe 2014; Meier et al. 2011). The adoptions of concrete tasks from cenralised examinations suggest that the amount of tasks that require more challenging activities should also be carefully increased for the ZP10. This conscious effect of control could also be adapted to a consideration of the topic area of stochastics, which has not yet been considered adequaltely in the desinging of ZP 10 until now.

Against this background, continued education and training for teachers in general as well as for eachers designing the tasks for such examinations would be desirable. In the context of cognitive activation, the goal of such continued education and training should be to make teachers aware of different task characteristics as well as the consious consideration of said characteristics in the process of choosing tasks for ZP 10, class tests and the everyday mathematics lesson.

References

Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., Raths, J., Wittrock, M. C.: 2001, *A taxonomy for learning, teaching, and assessing: A revision of Bloom's Taxonomy of Educational Objectives*, New York: Longman.

Blum, W., vom Hofe, R., Jordan, A. Kleine, M.: 2004, Grundvorstellungen als aufgabenanalytisches und diagnostisches Instrument bei PISA, in: *Mathematische Kompetenzen von Schülerinnen und Schülern in Deutschland*, 145–157, Wiesbaden: VS Verlag für Sozialwissenschaften.

Blum, W., Drüke-Noe, C. R., Hartung, C. R., Köller, O.: 2006, *Bildungsstandards Mathematik: konkret. Sekundarstufe I: Aufgabenbeispiele, Unterrichtsanregungen, Fortbildungsideen,* Berlin: Cornelsen Verlag Scriptor.

Büchter, A. Leuders, T.: 2005; *Mathematikaufgaben selbst entwickeln. Lernen fördern-Leistung überprüfen,* Berlin: Cornelsen Scriptor.

Büchter, A., Pallack, A.: 2012, Methodische Überlegungen und empirische Analysen zur impliziten Standardsetzung durch zentrale Prüfungen, *Journal für Mathematikdidaktik* **33**, 59–85.

Drüke-Noe, Ch.: 2014, *Aufgabenkultur in Klassenarbeiten im Fach Mathematik. Empirische Untersuchungen in neunten und zehnten Klassen*, Wiesbaden: Springer Spektrum.

Freudenthal, H.: 1983, *Didactical phenomenology of mathematical structures*, Dordrecht: Reidel.

Hoffart, E.: 2011, *Mathematische Vergleichsarbeiten in der Grundschule: Zum diagnostischen Potential von Aufgaben und deren Bearbeitungen einer landesweiten Vergleichsarbeit für dritte Klassen*, Universität Giessen.

Jordan, A., Ross, N., Krauss, S., Baumert, J., Blum, W., Neubrand, M., Löwen, K., Brunner, M., Kunter, M.: 2006, *Klassifkationsschema für Mathematikaufgaben: Dokumentation der Aufgabenkategorisierung im COACTIV-Projekt*, Berlin: Max-Planck-Institut für Bildungsforschung.

Jordan, A., Krauss, S., Löwen, K., Blum, W., Neubrand, M., Brunner, M., Kunter, M., Baumert, J.: 2008, Aufgaben im COACTIV-Projekt: Zeugnisse des kognitiven Aktivierungspotentials im deutschen Mathematikunterricht, *Journal für Mathematikdidaktik* **29**, 83–107.

Klein, E. D., Krüger, M., Kühn, S., M., Ackeren, I.: 2014, Wirkungen zentraler Abschlussprüfungen im Mehrebenensystem Schule. Eine Zwischenbilanz internationaler und nationaler Befunde und Forschungsdesiderata, *Zeitschrift für Erziehungswissenschaft* **17**, 7–33.

Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruber, H., Prenzel, M., Reiss, K., Riquarts, K., Rost, J., Tenorth, H.-E. Vollmer, H. J: 2003, *Zur Entwicklung nationaler Bildungsstandards. Eine Expertise*, Bonn: Bundesministerium für Bildung und Forschung.

KMK: 2001, *Erste Konsequenzen aus der PISA Studie. Beschluss vom 05./06. 12.2001*, www.kmk.org/aktuell/pm011206.htm#ref1, 01.11.16.

KMK: 2003, *Bildungsstandards im Fach Mathematik für den Mittleren Schulabschluss–Beschluss der Kultusministerkonferenz vom 04.12.2003*, www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2003/2003_12_04-Bildungsstandards-MatheMittleren-SA.pdf, 10.09.2016.

Krüger, M.: 2015, *Aufgabenkultur in zentralen Abschlussprüfungen: Exploration und Deskription naturwissenschaftlicher Aufgabenstellungen im internationalen Vergleich*, Münster: Waxmann.

Kunter, M., Baumert, J., Blum, W., Klusmann, U., Krauss, S., Neubrand, M.: 2013, *Cognitive activation in the mathematics classroom and professional competence of teachers. Results from the COACTIV project*, New York, Springer.

Kühn, S., Drüke-Noe, Ch.: 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.

Kühn, S.: 2013, Vergleichbarkeit beim Mittleren Schulabschluss? Ein Überblick über die Vielfalt schulstrukturell möglicher Bildungswege und Prüfungsverfahren in den deutschen Ländern, *Die Deutsche Schule* **105**, 87–101.

Maier, U., Bohl, T., Metz, K. Kleinknecht, M.: 2011, Einflüsse von Merkmalen des Testsystems und Schulkontextfaktoren auf die Akzeptanz und Rezeption von zentralen Testrückmeldungen durch Lehrkräfte, *Journal of Educational Research Online* **3**, 62–93.

MSW: 2013, *Zentrale Prüfung 2013, Mathematik, Realschule*, Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen.

MSW: 2015, *Zentrale Prüfung 2015, Mathematik, Realschule*, Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen.

Neubrand, M., Klieme, E., Lüdtke, O. Neubrand, J.: 2002, Kompetenzstufen und Schwierigkeitsmodelle für den PISA-Test zur mathematischen Grundbildung, *Unterrichtswissenschaft* **30**, 100–119.

Neubrand, M., Jordan, A., Krauss, S., Blum, W. Löwen, K.: 2011, Aufgaben im COACTIV-Projekt: Einblicke in das Potenzial für kognitive Aktivierung im Mathematikunterricht, in: *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV*, 115–132, Waxmann.

Neubrand, M., Jordan, A. (†), Krauss, S., Blum, W., Löwen, K.: 2013, Task Analysis in COACTIV: Examining the Potential for Cognitive Activation in German Mathematics Classrooms (Chapter 7), in: *Cognitive Activation in the Mathematics Classroom and Professional Competence of Teachers*, 125–144, Springer.

OECD: 2004, *Learning for Tomorrow's World. First Results from PISA 2003*, OECD Publishing.

OECD: 2014, *PISA 2012 Results: What Students Know and Can Do*: *Student Performance in Mathematics, Reading and Science*, OECD Publishing.

Reiss, K.: 2004, Bildungsstandards und die Rolle der Fachdidaktik am Beispiel der Mathematik, *Zeitschrift für Pädagogik* **50**, 635–649.

Resnick, L. Ford, W.: 1981, *The Psychology of Mathematics for Instruction*, Hillsdale: Lawrence Erlbaum Associates.

Rosenshine, B.: 1970, Evaluation of instruction, *Review of Educational Research* **40**, 279–300.

S c h e j a, B.: 2007, *Der Wandel von Bildungssystemen am Beispiel von Deutschland und Polen unter besonderer Berücksichtigung der PISA-Studie,* Dissertation, Universität zu Köln.

T e r h a r t, E., B a u m g a r t, F., M e d e r, N. v o n S y c h o w s k i, G.: 2009, Standardisierte Prüfungsverfahren in der Erziehungswissenschaft: Kontext, Formen, Konsequenzen, *Erziehungswissenschaft* **20**, 9–36.

W i n t e r, H.: 1995, Mathematikunterricht und Allgemeinbildung, *Mitteilungen der Gesellschaft für Didaktik der Mathematik* **61**, 37–46.

Z a b a l a, D., M i n n i c i, A., M c M u r r e r, J., B r i g g s, L.: 2008, *State high school exit exams: A movetoward end-of-course exams*, Washington, D. C., Center on Education Policy (CEP).

Kognitywna aktywizacja przez zadania matematyczne pochodzące z centralnych egzaminów poziomu ISCED 2 na przykładzie Nadrenii Północnej-Westfalii

S t r e s z c z e n i e

Egzaminy zewnętrzne sa powszechnymi elementami monitoringu i sterowania systemem szkolnym. Współczesne badania pokazują wyraźny związek między formą zadań egzaminacyjnych a sposobami pracy i treściami poruszanymi podczas prowadzenia lekcji matematyki. Zentrale Prüfung – ZP 10 – jest centralnym egzaminem przeprowadzonym na koniec nauki na stopniu secondary education w Nadrenii Pomocnej-Westfalii, najbardziej licznym landzie w Niemczech. Wyniki tych egzaminów, jak i tych osiągniętych w badaniach PISA w okresie miedzy 2003 i 2015 pokazują znaczny przyrost umiejętności niemieckich uczniów. Z tego powodu interesujące staja się przeanalizowanie zadań z egzaminów zewnętrznych na poziomie ISCED 2, aby móc określić elementy, których wpływ na takie postępy możemy przyjąć jako znaczny. Artykuł poświęcony jest badaniu kognitywnej aktywizacji przez zadania matematyczne pochodzące z ZP 10. Przez kognitywną aktywizację rozumiane są kategorie programowe, treściowe (content-related) i kognitywne, które potencjalnie mogą być zidentyfikowane w procesie rozwiązania zadania. Te analizę poprzedza przedstawienie standardów programowych i dyskusja dotycząca funkcji zadań matematycznych. Pomiar kognitywnej aktywizacji został przeprowadzony poprzez analizę pewnego zbioru zadań egzaminacyjnych, użyte we wszystkich typach szkol Nadrenii Północnej-Westfalii na poziomie ISCED 2 w okresie pomiędzy 2007 – (rok wprowadzenia narzędzia) i 2016 ($N = 1706$).

University of Cologne
Institut für Mathematikdidaktik
Gronewaldstraße 2
50931 Köln GERMANY
e-mail: *schejab@uni-koeln.de*

2.3 Publikation III

Scheja, B. (2017). The changing cognitive requirement of test tasks in mathematics – a longitudinal study of the Polish middel school exam. *Didactica Mathematicae*, 39, 101 – 130.

Abstract

Due to the remarkably positive results of Polish students in Mathematics in the PISA survey it is of interest from a didactic point of view to focus on instruments which can be regarded to significantly enhance the impact on quality. Therefore, a current survey investigates the assessment tasks of the Polish middle school examinations which were used between 2002 and 2015 with the concept of cognitive activation. By analyzing this particular period the effect of the 2012 reforms in curricula and organization of the middle school examinations and their impact on the concept of the assessment tasks are difficult to grasp, thus the results can only be categorized as indistinct. The present study expands on this approach by adding a development perspective and consequently it is compiled as a longitudinal study. In doing so, the long-term change of the cognitive requirements of the assessment tasks of the Polish middle school examinations between 2002 and 2015 have been recorded and analyzed. The term "cognitive requirements" refers to various expressions within a broad range of cognitive categories that can be activated throughout the process of solving a task. The investigation starts by shedding light on the curricular framework of the middle school examination in mathematics, whose directives form an important basis for the development of assessment tasks. The longitudinal change in cognitive requirement is assessed by analyzing the problems used during the initial years of the middle school examinations (2002 – 2005; N=65) and in the years following their reform (2012 – 2015; N=92). This shows that the distribution of mathematical activities and, with some restrictions, the respective difficulty levels, has become arguably more balanced after the reforms, and that tasks tend to be more difficult than before. This development is particularly evident regarding tasks that require mathematical argumentation and an increased amount of task context processing, and task contexts now involve both intra-mathematical and extra-mathematical matters alike.

DIDACTICA MATHEMATICAE
Vol. 39, 2017, p. 101–130

Bruno Scheja (Köln, Germany)

The changing cognitive requirement of test tasks in mathematics – A longitudinal study of the Polish middle school examinations

Abstract: Due to the remarkably positive results of Polish students in Mathematics in the PISA survey it is of interest from a didactic point of view to focus on instruments which can be regarded to significantly enhance the impact on quality. Therefore a current survey investigates the assessment tasks of the Polish middle school examinations which were used between 2002 and 2015 with the concept of cognitive activation. By analyzing this particular period the effect of the 2012 reforms in curricula and organization of the middle school examinations and their impact on the concept of the assessment tasks are difficult to grasp, thus the results can only be categorized as indistinct. The present study expands on this approach by adding a development perspective and consequently it is compiled as longitudinal study. In doing so the long-term change of the cognitive requirements of the assessment tasks of the Polish middle school examinations between 2002 and 2015 have been recorded and analyzed. The term "cognitive requirements" refers to various expressions within a broad range of cognitive categories that can be activated throughout the process of solving a task. The investigation starts by shedding light on the curricular framework of the middle school examination in mathematics, whose directives form an important basis for the development of assessment tasks. The longitudinal change in cognitive requirement is assessed by analyzing the problems used during the initial years of the middle school examinations (2002–2005; $N = 65$) and in the years following their reform (2012–2015; $N = 92$). This shows that the distribution of mathematical activities and, with some restrictions, the respective difficulty levels, has become arguably more balanced after the reforms,

Key words: secondary education, cognitive processes, curriculum research, task design.

and that tasks tend to be more difficult than before. This development is particularly evident regarding tasks that require mathematical argumentation and an increased amount of task context processing, and task contexts now involve both intra-mathematical and extra-mathematical matters alike.

1 Introduction

Intense political debates about the defective state of the Polish educational system led to extensive restructuring in the late 1990s: The reforms, which came into effect in 1999, introduced a new school structure, central exams, exam committees, and core curricula (cf. Denek, 2005; Konarzewski, 2004; Kupisiewicz, 2006; Zahorska, 2002). Considering the positive development in mathematical performance that followed, and in light of the new *educational reforms of 2018*, it seems highly relevant for didactics of mathematics to investigate control mechanisms that promise a particularly large impact on quality (cf. Bialecki, Haman, 2001; Kunter, Voss 2013; MEN, 2013). Current empirical findings primarily assign such impact to instruments that not only express *desired effects* from educational politics, but that show actual *observable effects* (cf. Maier et al., 2011)[1]. Unlike system monitoring studies with evaluations based on a multi-matrix design, the middle school examination introduced in 2002 is primarily relevant on an individual level *(primary function)* and can be described as highly effective in terms of control, especially regarding the ISCED 2 level[2]. Here, observable effects within mathematics as an exam subject can be viewed as problem formats or difficulty-increasing features of problems and their orchestration (cf. Ibd.; Jordan et al., 2006). This research aim is addressed by Scheja (2017) with the concept of *cognitive activation* (cf. Baumert et al., 2010; Kunter et al., 2013; Neubrand et al., 2013), which is applied in the analysis and assessment of the entire set of tasks for the middle school examination between 2002 and 2015. The choice of this period ignores the effects of the 2012 reforms to the curriculum and organization of the middle school examinations on the development of assessment tasks, so that the results of this study must be regarded as indistinct. The present study expands on this approach by adding a development perspective. It is a longitudinal study that analyzes the tasks of the middle school examination

[1]For instance, German studies show that problems posed in centralized final exams only form an unbalanced representation of the respective core curriculum. Büchter and Pallack (2012) describe this as implicit standardization, i.e., standardization anchored in the exam questions (cf. Kühn, Drüke-Noe, 2013; Neubrand, Neubrand, 2010; Scheja, 2017).

[2]International Standard Classification of Education, lower Secondary

for the periods of 2002–2005 and 2012–2015. By investigating this change over time, any developments in terms of cognitive requirements between 2002 and 2015 can be shown. This forms the basis for a *study of implementation* (cf. Gräsel, Parchmann, 2004; Karpinski et al. 2013) that assesses whether and to what extent the curricular guidelines of the two periods are represented in the middle school examination tasks.

The theory section below outlines the curricular and organizational framework of the middle school examination. Next, the question whether and to what extent the concept of cognitive requirements has evolved in recent years and is suitable for analyzing the assessment tasks is addressed. The methodology section introduces the underlying scheme of categories, which draws upon established research on cognitive requirements. The results chapter compares the tasks used during the periods when the examination was first introduced (2002–2005) and after the exam reforms (2012–2015). Particular interest is placed on differences between the assessment tasks in question. This focus aims at creating a testing profile for each period of investigation.

2 Theoretical Background

2.1 Mathematics Tasks of Polish Middle School Examinations

Over the past three decades, the instrument of external performance assessment has gained in significance across Europe, as shown by van Ackeren (2003) in a study of selected countries that participate in PISA. Poland's emphasis on implementing this instrument in its educational policies is evidenced by its explication as one of the five elements of reform. Kupisiewicz (2006, p. 80) and others welcome this decision; Brozek (2005, p. 240) discusses introducing a "system of external performance assessment" in the context of this instrument, which shows that he considers this element of reform as an organized complex of individual elements that unites institutional and conceptual aspects of educational reform. The term is based on the following elements that constitute the system of external performance assessment:

1. Boards of examiners that design, organize, and to some extent perform external examinations
2. External performance assessments at the end of the 2^{nd}, 3^{rd}, and 4^{th} educational stages
3. Directives for performance assessments in the form of standards

External performance assessments are distributed vertically on the ISCED 1 and ISCED 2 levels, and horizontally on the ISCED 3 level. Figure one illustrates the timing of the assessments and how they are embedded in the new school structure.

School years	Type of school	
10–12	Stage IV: advanced ISCED 3 school types	← Centralized A-levels/ Professional exams
7–9	Stage III: middle school	← Middle school
4–6	Stage II of primary school	← Primary
- - - - - - - - -	- - - - - - - - - - - - - - - - - -	
1–3	Stage I of primary school	

Figure 1. Embedding of centralized examinations into the school structure (Art. 9, MEN, July 25th, 1998).

At the end of the third and final year of middle school (ninth grade overall), which is just before the end of mandatory education, the responsible *Regional Boards of Examiners* (Abbr.: OKE) work together with the individual schools to perform the second centrally designed examination. This is the nationally standardized middle school exam (cf. Fig. 1), which is mandatory for *all* Polish students aged between 15 and 16 years since it was conducted for the first time in 2002.

In a step towards the central research question of this study, the next subsections outline the following aspects of the mathematics section of the exam:

- Which curricular focus was set in the middle school examinations during the two periods under review here and
- On which evaluation concept the exam is based.

2.1.1 Curricular bases for Mathematics Tasks in 2002–2005 and 2012–2015

After the middle school examination was introduced in 2002, the exam consisted of two parts: The mathematics and science part of the exam was conducted on the second day of examinations and focused on subjects such as mathematics, biology, geography, physics, astronomy, and chemistry. One essential component of this part of the exam was mathematical problems, which were the same across the entire country and accounted for 40–50% of the exam each year (cf. Scheja, 2017). The time allotted for answering the 35 mathematical and scientific questions was 120 minutes. In terms of content, the questions were based on nation-wide directives which included the core curriculum (MEN, 1999), which was also introduced as part of the reforms, and

Field of Standards	Abilities and Skills of the Field	
I. Correct use of terms, formulas, and procedures in mathematical and scientific subjects that are relevant for everyday life and professional education	The student uses mathematical and scientific terms and procedures.	The student... a) ...reads and understands texts that contain mathematical and scientific terms and concepts, e.g., from books or newspapers. b) ...chooses proper terms and concepts to describe phenomena, properties, and behaviors of objects and organisms. c) ...applies rational concepts with regards to interaction with the environment.
	The student performs calculations in various practical situations.	a) ...applies calculations in practice. b) ...uses percentages. c) ...uses approximations. d) ...uses units of measurement.
	The student utilizes the properties of shapes.	a) ...recognizes geometrical shapes in his/her surroundings. b) ...calculates the dimensions of planar and spatial shapes. c) ...utilizes the properties of units of measurement.
II. Finding and utilizing information	..	..
III. Recording and describing facts, systematic connections and dependencies, especially in causal, functional, spatial, and temporal terms	..	..
IV. Applying integral knowledge and abilities to the process of problem solving	..	..

Table 1. Exemplary specification of requirement standards in mathematical and natural science subjects (cf. CKE, 1999).

the *Standards for Requirements of Mathematical and Natural Science Subjects* (cf. CKE, 1999), which stem from the same period. The latter are divided into four standard areas and feature three levels of abstraction (level 1: standards, levels 2 and 3; abilities and skills). Table 1 shows the implementation of these specifications for the standards level (level 1). When the Central Board of Examiners (Abbr.: CKE) developed the exam questions, they required that the scheme of standards is properly operationalized in the tasks (cf. MEN, 1998b).
The nation-wide middle school examination underwent extensive reforms in 2012, primarily as a result of great conceptional needs for improvement in terms of the correlation between the curricular basis and the examination standards shown in table 1 (cf. Konarzewski, 1998; Konarzewski, 2004, p. 127). Mathematics now forms a *separate* exam. This exam consists of 23 tasks, is designed for a duration of 90 minutes, and forms one out of two components in the mathematical and scientific examination day ("Mathematics", 90 minutes, and "Scientific Subjects", 60 minutes). The content is now based exclusively on the nationwide core curriculum, which was thoroughly redesigned in 2008. With regards to the requirements, it defines two types of categories: five *general requirements* and eleven *content requirements*. The first category is characterized below.

1. Using and creating information

   The student interprets and creates a text of a mathematical nature, uses mathematical language to describe the reasoning and obtained.

2. Using and interpreting representations

   The student uses simple, well-known mathematical objects, interprets mathematical concepts, and makes use of mathematical objects.

3. Mathematical modelling

   The student assigns a mathematical model to a simple situation, creates a mathematical model of a situation.

4. Using and creating strategies

   The student makes use of the strategy clearly defined in the task, creates problems-solving strategies.

5. Reasoning

   The student provides simple reasoning, justifies its validity (cf. MEN, 2008, p. 35).

Comparing these requirements to the general requirements of the German *Educational Standards for Middle School Exams* shows great similarities, and only the K5[3] category has no equivalent in the core curriculum. This conceptional similarity is surprising, since neither of the standards has any explicit foundation in specific learning theories (cf. MEN, 2009; Reiss, 2004). At the same time, the overlap between the two systems reflects an equal influence of Heinrich Winter's *Three Basic Experiences in Mathematics Education* (cf. 1995). Their intended implementation in the general requirements is reflected – though not completely isolated – in the following definitions:

- *Structure-Orientation*: using and interpreting representations, mathematical modeling and using and creating strategies
- *Application-Orientation*: using and creating strategies, using and interpreting representations and using and creating information
- *Problem-Orientation*: using and creating strategies and reasoning.

If we look at the pre-reform standards with a view to achieving a similar alignment of one basic experience and individual requirements (Tab 1.; CKE, 1999), it turns out that almost only the basic experiences of application- and problem-orientation are regarded as major goals of educational processes.

At the same time, the aforementioned shift in orientation regarding the required abilities corresponds to the vaguely outlined task concept of their designers. For instance, in breaking with the problem design from the early days of the middle school examination, a smaller portion of tasks is now included to test knowledge and use of algorithms in their typical application. Instead, tasks that require an understanding of mathematical concepts and the ability to select the right mathematical strategies for solving unfamiliar types of problems are to account for a larger portion of the exams (CKE, 2010). Samborska argues in favor of these new directives, invoking "a deeper understanding of mathematics" (2015, p. 49) as their main goal.

The two requirement categories of the newly devised core curriculum are essentially scales for standardization and comparison, which serve the intradisciplinary advancement of schools and lessons by specifying the educational mission of middle schools in the subject of mathematics on two distinct levels of abstraction (*orienting function*). Their second purpose is to provide the Central Board of Examiners with a basis for devising the assessment tasks each year – just like the exam standards did between 2002 and 2011. The responsible CKE bodies have the goal of ensuring that the performance of students is assessed with respect to the requirement standards (*evaluating function*; Art. 9 MEN, 1998b).

[3]K5: dealing with symbolic, formal, and technical elements of mathematics (KMK, 2003, p. 8).

Much to the dismay of some educators, the specific implementation of this process, i.e., which conceptional aspects are involved regarding the existence of distinct requirement levels or even their relation to the defined standards, is not published (cf. Konarzewski, 2004, p. 126.). In his interviews on the new core curriculum, Konarzewski shows that the responsible CKE employees remained silent when faced with questions about especially the design of assessment tasks, invoking official confidentiality. Since strict secrecy is a core principle of developing assessment tasks, the actual problems used in the middle school examinations are not tested empirically in advance – much like in the centralized examination administered in North-Rhine Westphalia, and unlike system monitoring studies (cf. Büchter, Pallack, 2012; Konarzewski, 2004).

2.1.2 Evaluation of Mathematics Tasks in the Middle School Exam

The middle school examinations are evaluated by external examiners from the Regional Board of Examiners (OKE; Art. 9 MEN, 1998b). The assessment tasks are officially divided into open ended and closed tasks, depending on their individual characteristics. *Closed tasks* are multiple choice, and the evaluation sheet distinguishes between correct and incorrect solved states (1 or 0 points, respectively) (CKE, 2010 p. 61; Greefrath, 2004, p. 17.). *Open ended tasks* typically involve ambiguous transformations and unclear target states. They are assigned more points (between 2 and 5 points each), and they can have up to six distinct levels of solution (cf. CKE 2010, p. 61). In the periods under review, these tasks accounted for 7–13 % of all tasks (cf. Scheja, 2017). Much like the tasks of the centralized examination administered in North-Rhine Westphalia, the weighting of the solution levels does not follow any established educational concept. We can therefore assume that the mapping of solution levels and points in the expectation specifications is based on the "extent of processing" (cf. CKE 2010, p. 61; Büchter, Pallack, 2012). The seven levels of solution for open ended tasks are defined in general and then applied to each exam problem in the expectation specifications.

The numbers of points are then mapped to percentages and percentiles, which are handed out with the middle school diplomas. And although these numbers do not influence the grades, the fact that the examinations are mandatory for all students if they wish to graduate from middle school and factor into the recruitment scores for admission to the (subsequent) ISCED 3 schools, the middle school exams also serve a purpose of selection in addition to their informative purpose.

2.2 Mathematical Tasks and Cognitive Requirements

2.2.1 Conceptual Development in the Cognitive Requirement of Mathematical Tasks

Cognitively demanding mathematical education is currently operationalized in different ways. According to Leuders and Holzäpfel (2011, p. 213), this can be done either by means of observable teaching methods, or – as it is done in this study – by means of the tasks used. Following this second approach, Neubrand et al. (2002, p. 105.) identified properties that make tasks more difficult by analyzing 117 tasks used in PISA-2000 (PISA international: 31 tasks; PISA national: 86 tasks) from a cognitive science point of view. These properties include:

1. the curricular knowledge level,
2. the context given for a task,
3. and the complexity level of the required modeling.

Furthermore, it turns out that the following sub-processes within the modeling process contribute to the cognitive requirements of tasks:

3.1 Openness of the modeling process,

3.2 Degree of processing,

3.3 and forming arguments.

Neubrand et al. conclude that the cognitive complexity of mathematical tasks, and therefore their empirical difficulty, increase with a growing number of cognitive processes that must be completed on the path to a solution (cf. ibid.). More generally speaking, and focusing on the quality of mathematical instruction, Klieme et al. (2001, p. 51) and others demand complex tasks and chains of argument as a means of facilitating comprehensive learning processes. This kind of learning is characterized in particular by the way it provokes contemplation, argumentation, communication, invention, and decision making in the students. Finally, the amount of pre-existing knowledge activated during a learning process contributes to cognitive activation as well, according to Leuders and Holzäpfel (2011, p. 124).

Based on work by Neubrand et al. (2002), the COACTIV project defines cognitive activation based on the potential inherent in the relevant tasks. Tasks that have this kind of potential are supposed to offer students opportunities for learning during lessons, encourage them to engage in goal-oriented cognitive activities relevant to the subject, prompt them to think more deeply, and get them to actively engage their minds. Tasks that fulfill these criteria are particularly suitable for changing or expanding existing knowledge structures, create new ones, and form connections within them (Neubrand et al. 2013, p. 129). The COACTIV project assesses the degree of cognitive activation, the *level of cognitive requirement*, based on four types of mathematical activities required for solving tasks: extra-mathematical modeling, intra-mathematical modeling, argumentation, and using mathematical representations (Jordan et al., 2006, p. 31.). This approach to the concept forms an instrument for describing the level of cognitive requirement that can describe the potential of mathematical tasks drawing only upon properties of those tasks. In a study of task culture, Drüke-Noe also shows that the four COACTIV assessment categories require an addition (2014, p. 66.), because they do not differentiate between levels of processing calculus; for instance, the use of numbers and symbols required for solving tasks can vary a lot in terms of cognitive complexity, resulting in varying levels of cognitive requirement in the respective tasks. Approaches to operationalizing the processing of calculus can be found in Cohors-Fresenborg et al. (2004) or in the PISA category *mathematical* s*ymbols and formalism competency* (Turner et al., 2013). These approaches take evaluation elements, specifically the "degree of processing" (Neubrand et al., 2002, p. 111) into account within the required calculations. However, unlike Drüke-Noe's technical work, they do not fully meet the COACTIV criteria for an assessment category because, among other things, they are operationalized using a different number of requirement levels (cf. ibid.).

It remains to be clarified whether the use of analytic categories for assessing the cognitive requirements of *learning* tasks can be adopted as a valid instrument for analyzing centrally devised *examination* tasks, and vice versa. *Learning* tasks are tasks that are used in learning situations, for instance as introductory tasks or homework assignments (cf. Jordan et al., 2006). They are an interface for activities and communication between teachers and students, which makes them a key instrument for controlling lessons and therefore also students' mental activities. Within this process, tasks become the main conveyors of mathematical content, determining its scope. We can therefore consider them the primary control medium for building competence. The requirements demanded of and experienced by students can in turn be identified by analyzing the tasks used for teaching. This makes learning tasks a vehicle for a wide variety of learners' cognitive activities in the process of knowledge building. These activities can be classified by using the previously mentioned five categories (ibid.). Unlike learning tasks, *assessment tasks* are designed with the goal of identifying which skills and knowledge were successfully taught in the educational process (cf. Terhart et al., 2009, p. 24). They can be designed internally – e.g., for regular exams – or externally – e.g., for centralized exams. Contrary to expectations, there is strong evidence based on extensive analyses of structures and requirements which suggests that learning and assessment tasks share mostly identical cognitive requirement characteristics (Drüke-Noe 2014, p. 9). The only differences are in the treatment and evaluation of solutions (cf. Büchter, Leuders, 2005). Regarding the purpose of assessment tasks, note that the cognitive activities that learners are potentially expected to perform during the solution process can be interpreted as cognitive *requirement expectation.* To conclude, the kind of assessment tasks under investigation in this study can be viewed and analyzed in the same way as learning tasks as bearers of cognitive requirement.

2.2.2 Findings on the Cognitive Requirement of Mathematical Tasks

The number of studies focusing on the identification of cognitive requirement levels in teaching and assessment tasks has grown steadily, especially in recent years. With regards to assessment tasks, a study by Neubrand et al. (2002), which investigates the tasks used in national and international PISA tests, is particularly noteworthy. The requirement level of the national PISA tests turns out to be relatively high, as almost three quarters of the tasks require modeling operations that primarily belong into the categories of reproduction and connection (95% in total). Ultimately, around 77% of the tasks have some connection to some intra- or extra-mathematical context, while at the same time being highly open ended; in turn, a relatively small amount of argumentation is required (ibid.).

Apart from German learning and assessment tasks (introductory tasks, homework, and exam tasks) analyzed as part of the COACTIV study, which were found to be overall homogeneous and have a very low level of cognitive requirement (cf. Neubrand et al. 2013), this paper draws heavily on results of studies investigating the written tasks of centralized (final) exams and the expected cognitive requirement. For instance, Drüke-Noe (2014) and later Scheja (2016) report on the cognitive requirement level of exam questions in two German federal states: Hesse (H; 2007–2009) and North-Rhine Westphalia (NRW; 2007–2016). Both studies are based on the five (or six) mathematical activities (extra- and intra-mathematical modeling, argumentation, use of mathematical representations, working technically, and processing mathematical texts). One major insight of both studies is that the distribution of activities and requirement levels is fairly even; the assessment tasks have an overall high level of cognitive requirement and feature much less technical work than, for instance, COACTIV exam tasks (cf. Drüke-Noe, 2014). Modeling tasks account for 92.5% (H) and 91.4% (NRW) respectively, and they are mostly embedded in extra-mathematical contexts (H: 80%; NRW: 75.2%). Tasks that require extra-mathematical modeling mostly do so on low (H: 56.1%; NRW: 32.5%) or medium (H: 20.2%; NRW: 41.9%) requirement levels; more demanding intra- or extra-mathematical modeling tasks are used sparingly (<2% in both cases). Much like the "processing context", the using mathematical representations and working technically contribute significantly to the level of cognitive requirement in Hesse's and North-Rhine Westphalia's assessment tasks, while the overall share of argumentation tasks (H: 16%; NRW: 9.9%) is fairly low.

The *second* point of interest, investigated in Scheja's (2017) study of cognitive requirement levels in the Polish middle school examinations, lies in a centralized, also recently reformed education system. It shows that assessment tasks from the 2002–2015 period have a generally rather high level of cognitive requirement. This is a result of relatively demanding representations and technical work embedded in complex modeling processes that often involve several steps (modeling tasks: 98.6%; mid-level intra- and extra-mathematical modeling tasks: 64.5% in total). Argumentation, which is merely required by 4.5% of assessment tasks, has only a minor impact on the overall level of cognitive requirement. Scheja sees a major difference between the requirement profiles in NRW and Poland in the amount and requirement levels of intra-mathematical modeling tasks included (ibd.). However, statements about the middle school exam questions and their contextual embedding and argumentation elements are rather *indistinct* due to the monolithically analyzed period under review (2002–2015). This criticism is probably accurate, as the 2012 reform of the middle school exam presumably led to longitudinal shifts in task design. For instance, comparing the conceptual guidelines before and after the reforms (cf. 2.1) clearly shows that there was a shift from extra- to intra-mathematical problem contexts, and that argumentation has become more important. If these guidelines were implemented as expected, the tasks from the 2012–2015 period would tend to diverge more with regards to different contexts and the required level of argumentation would become more similar between the two task profiles[4]. So far, there has yet to be a confirming analysis that goes beyond a qualitative description of the status quo and takes into account the development of requirement by means of a longitudinal study. This paper is an attempt to start closing that gap by taking on a development perspective.

[4]This was confirmed for longitudinal development findings by Kühn and Drüke-Noe with respect to the NRW task pool. They find the German task design to be conceptually stable (2013, p. 925 f.).

In order to answer the central question of this paper, we shall now categorize the tasks of the Polish middle school exam in such a way that their contribution to the cognitive requirement becomes evident. The categories used for this purpose are empirically tried and tested, and they are based on pertinent didactic and pedagogical/psychological literature. The central question of this paper is:

Do the mathematical tasks of the Polish middle school exam from different periods of examination (2002–2005, after the introduction, and 2012–2015, after the reform) exhibit any tendencies for change regarding the level of cognitive requirement, and which possible explanations can be found for this?

3 Method

The present study investigates the cognitive complexity of assessment tasks using the concept of cognitive requirement. For this purpose, the author draws upon methods for categorizing tasks first proposed by Neubrand et al. (cf. 2002) and expanded conceptually by Drüke-Noe (cf. 2014). After presenting the two datasets in 3.1, the conceptual foundations and the method of task categorization are explained (3.2). The last subsection outlines the process of coding (3.3).

3.1 Sample and Dataset

The two sets of data analyzed in this study are, *on the one hand*, the pool of tasks for Polish middle school exams between 2002 (when the exam was introduced) and 2005. During this period, the mathematics section was an integral part of the mathematics and science exam (cf. 2.1). The set comprises 65 task instructions (assessment tasks). *On the other hand*, there are 92 tasks from the period after the mathematical section of the test was separated from the mathematics and science part of the middle school exam (2012–2015). The following conditions ensure comparability between the two periods:

- Tasks from the main test date were used for both periods under comparison.
- Both datasets comprise tasks from periods of the same total duration (4 years each).
- Both sets of tasks stem exclusively from the "pool of tasks for learners without impairments and learners with specific learning impairments".

The much smaller number of tasks from the introductory phase of the middle school exam results in greater relative effects from possible errors in coding. As a result, statements can only be generalized within certain limitations.

3.2 Categorization Scheme for Determining Cognitive Requirement

To determine the degree of cognitive activation, i.e. the level of cognitive requirement, for the assessment tasks, the author uses a selection of tried and tested categories. These are taken from the conceptual work done by Neubrand et al. (2002), which was expanded upon in the COACTIV project and by Drüke-Noe, and applied on a large scale in analyses of the cognitive activation achieved by learning and assessment tasks in German schools (Drüke-Noe, 2014; Jordan et al., 2006; Neubrand et al., 2013). The study is based on a total of six categories for analyzing sources of cognitive requirement. Classifications must be regarded as highly inferential (cf. table 2). Further details can be found in the categorization scheme below, and in Drüke-Noe (2014, p. 71.) and Jordan et al. (2006).

Category	**Properties**
Extra-mathematical modeling	0 - not required, 1 - standard modeling, 2 - multistep modeling, 3 - reflection on a model, development and validation of complex models
Intra-mathematical modeling	0 - not required, 1 - standard modeling, 2 - multistep modeling, 3 - reflection on a model, validation, strategy development
Mathematical argumentation	0 - not required, 1 - standard reasoning, 2 - multistep argumentation, 3 - development of complex argumentation, proofs, evaluation of arguments
Using mathematical representaions	0 - not required, 1 - standard represenataions, 2 - switching between different representations, 3 - evaluating representations
Working technically	0 - not requiered, 1 - standard technic, 2- simple hierarchical operations, 3 - complex hierarchical operations
Processing mathematical texts	0 - not required, 1 - direct text comprehension, 2 - text comprehension with reorganisation, 3 - comprehension of logically complex texts

Table 2. Pivotal assessment criteria of this study (cf. Neubrand et al., 2013, p. 134; Drüke-Noe, 2014, p. 72).

The underlying categories for sources of requirement operationalize mathematical activities[5] on four cognitive requirement levels each. These levels express whether an activity is required on a high (3), medium (2), or low level (1), or not at all (0) for solving a task. The degree of cognitive complexity involved in an activity is expressed by these levels.

$$\begin{array}{ccc} \text{SIT} & \longrightarrow & \text{MOD} \\ \uparrow & & \downarrow \\ \text{KON} & \longleftarrow & \text{RES} \end{array}$$

Figure 2. Simplified cycle of modeling in mathematics (cf. Jordan et al. 2006).

Neubrand et al. (2013, p. 130) assume that modeling tasks involve a model-like solution process that can be based on either an intra- or an extra-mathematical context,[6] which is completed in a "structurally equivalent" (ibid.) way for both contexts. Figure 2 is a schematic showing the cyclical view of modeling in mathematics. This cycle idealizes the multi-step process of solving modeling tasks, focusing on reading and understanding a problematic situation (SIT), transferring it into a model for processing (MOD), and applying the results (RES) and the consequences (KON) to the original situation (SIT) (Jordan et al. 2006, p. 31). In addition to the cognitive requirement for translating between contexts, the level of argumentation required for completing the modeling process is assessed as well. Jordan et al. (2006, p. 40) define *mathematical argumentation* as the ability to present a valid chain of reasoning or to understand and evaluate various forms of mathematical arguments. This extends to all kinds of reasoning, including the ability to reflect on a mathematical model. Just like the modeling process, this cognitive activity relates to both intra- and extra-mathematical facts (say, as a proof or explanation). The category *using mathematical representation* – based on work by Brunner (1972) – primarily distinguishes between *iconic* levels of representation with regards to cognitive complexity. In *working technically*, individual elements of technically processing calculus in an idealized problem solution process are identified. Throughout these classifications, the levels from table 2 are assessed with regards to one of the four subject areas (arithmetic, algebra, geometry, and stochastics) based on the relevant term structures and applicable operations and any rules regarding their order. For the field of arithmetic, the four levels are implemented as follows:

[5] The category "Processing mathematical texts" is not strictly speaking a mathematical activity (cf. Neubrand et al. 2013). However, this category is included as a mathematical activity due to its significance for international comparative studies and within the Polish core curriculum and the standards (cf. Scheja, 2017).

[6] If a task merely requires the context-free use of abilities or factual knowledge with given parameters, no processing skills such as modeling or structuring are involved in the solution process (level 0). Neubrand et al. (2002, p.101) call this type of task *technical.*

Level 0: Not required, e.g. a^0.

Level 1: Only multiplication and division calculation/ only addition and subtraction calculation, e.g. simple exponentiations (positive base and power).

Level 2: Simple hierarchical operations (multiplication and division calculation before addition and subtraction calculation, exponentiation before multiplication and division calculation), e.g. multiplying/dividing exponentiations (positive base and power), extracting the root of a given number.

Level 3: Complex hierarchical operations (exponentiation with multiplication and division calculation/addition and subtraction calculation) (cf. Drüke-Noe, 2014, p. 78).

Finally, the tasks are classified according to the skill level in *processing mathematical texts* required for filtering information on the structure of the solution process from the task text in a structured manner. Here, tasks are particularly difficult (level 3) if the order of sentences or phrases does not match the order of mathematical processing. Logical connections are only present in a covert way; the main clauses and subclauses refer back to each other; important logical operators are multiple conditionals (if-then-connections) or all-quantifiers (Jordan et al., 2006, p. 53).

As an example of the categorization scheme from table 2, selected assessment tasks from the middle school examinations of the two periods under review are presented below.

The following table shows an overview of the ratings of the assessment tasks shown in table 3 with regards to the levels of requirement in the different categories described above.

In what follows, the coding process is explained using tasks 3 and 4 as examples (cf. tab 3; tab 4). Task 3 is open-ended (cf. 2.1.2), and its empirically most probably solution approach requires procedural modeling. The complexity of its linguistic logic is of a medium level, since the sentences and clauses do not directly match the steps of mathematical processing. There are several main clauses and sub-clauses, as well as references; the values given in the text (volume of the two pot sizes) cannot simply be copied into the calculation. Extra-mathematical modeling is also required on a medium level, because it takes the right translation processes between reality and mathematics to calculate the volume of a small and large pot, respectively (MOD I: system of linear equations with two variables). The given numbers of pots which Wojtek wants to fill allow multiplying these values (MOD II) and giving a complete answer. The required argumentation is of a medium level, since it requires a closed chain of argument involving several steps. The visual representation merely provides redundant information and is therefore not necessary, and neither is any intra-mathematical modeling (both rated 0). The requirements regarding calculus processing are rated 2; solving the linear system involves the application of simple hierarchies (e.g., order of operations). We can conclude that with the exceptions of intra-mathematical modeling and using mathematical representations, all mathematical activities contribute substantially to the cognitive requirement level of this task.

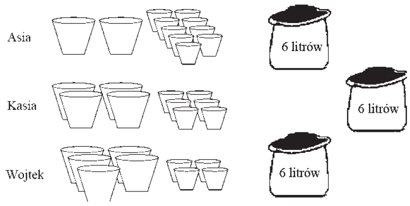
<table>
<tr><th>Example tasks from the 2002–2005 period</th><th>Example tasks from the 2012–2015 period</th></tr>
<tr><td>1) During a bicycle tour, Jacek and Marta met in the middle between their homes, which are 8 km apart. Marta's average speed was 16 km/h, Jacek's 20 km/h. Marta left home at 14:00. When did Jacek leave home if he arrived at the meeting point at the same time as Marta?
A. 13:53 B. 13:57 C. 14:03 D. 14:12</td><td>3) Asia, Kasia, and Wojtek are re-potting their flowers. Each of them has a 6-liter bag of soil and several pots of different sizes. Asia has used up all her soil by filling 2 large and 9 small pots. Kasia has used up all her soil to fill 4 large and 6 small pots. Wojtek wants to fill 5 large and 4 small pots. Does he have enough soil in his bag for this? Explain your answer.

</td></tr>
<tr><td>2) The table shows admission prices for a swimming pool according to the number of visits and their individual duration.
<table><tr><td>Ticket type</td><td>I</td><td>II</td><td>III</td><td>IV</td></tr><tr><td>Swimming times</td><td>10×1 h</td><td>8×1.5 h</td><td>20×1 h</td><td>15×1 h</td></tr><tr><td>Price</td><td>50 zł</td><td>50 zł</td><td>80 zł</td><td>70 zł</td></tr></table>One hour of swimming is the cheapest using ticket type
A. I B. II C. III D. IV</td><td>4) If a, b, and c are the lengths of the sides of a triangle and c is the longest side, that triangle is:
– a right triangle if $a^2 + b^2 = c^2$,
– an obtuse triangle if $a^2 + b^2 < c^2$,
– an acute triangle if $a^2 + b^2 > c^2$.
Complete this sentence using the correct answer.
The lengths $2\sqrt{3}$, $3\sqrt{2}$, $\sqrt{3}$.
A. do not form a triangle.
B. form a right triangle.
C. form an obtuse triangle.
D. form an acute triangle.</td></tr>
</table>

Table 3. Example tasks from 2002–2005 (tasks 1 and 2) and 2012–2015 (tasks 3 and 4).

	Requirement categories and their ratings					
Assessment task	Extra-math. modeling	Intra-math. modeling	Mathematical argumentation	Using mathematical representations	Working technically	Processing mathematical texts
1)	2	0	0	0	1	2
2)	2	0	0	1	1	1
3)	2	0	2	0	2	2
4)	0	2	0	0	3	2

Table 4. Requirement ratings for the tasks from tab. 3 in the categories from tab. 2.

Take assessment task no. 4 from 2012 (cf. tab. 3; tab. 4) as an example. It requires a medium understanding of mathematical text; logical connections are only given underneath the external linguistic form (multiple "if-then-connectives"), and the given values are not immediately applicable for the required calculations, because the specific lengths are not openly assigned to their individual designations. The subsequent intra-mathematical modeling is also on a medium requirement level (extra-mathematical modeling is not required and rated 0). Solving the task requires establishing connections – using, for instance, power laws – that go beyond the object mentioned in the task itself; therefore, the correct approach to solving it is not obvious from the information provided but requires several steps of modeling (cf. Jordan et al., 2006, p. 36). Since neither mathematical argumentation nor the use of mathematical representations are required, both are rated 0. Lastly, the requirements regarding the processing of calculus, i.e., working technically, are high, because complex hierarchical operations (powers, multiplication/division, and addition/subtraction) are involved in the solution process. Overall, we can note that intra-mathematical modeling, working technically, and processing texts are the exclusive sources contributing to task no. 4's level of cognitive requirement.

3.3 Coding the Tasks

The rating training and final rating of the 157 assessment tasks included in this (sub-)study form an integral part of the overarching study by Scheja (cf. 2017), which rated all tasks of the Polish middle school examination for the period of 2002 to 2015 ($N = 249$). The rating was done independently by the author (evaluator 1) and an experienced mathematics teacher (evaluator 2), who was thoroughly instructed about the details of the study in advance. Since the reliability of the rating is extremely important due to the highly inferential categories, both evaluators trained the rating in advance using COACTIV tasks that had already been assigned ratings in all six categories. A minimum reliability value before switching from the training tasks to rating the assessment tasks was set to $0.7 < k$ (cf. Wirtz, Caspar, 2002). This value corresponds to an excellent consistency of ratings. Comparing the two evaluators' ratings to those of the COACTIV evaluators (i.e., Drüke-Noe and her team) showed that the training ratings ($N = 291$) had a match rate of no less than "good" ($k \geqslant 0.7$) for all categories and levels (cf. ibid.). Finally it was determined that the ratings for the entire set of Polish tasks was sufficiently compatible to the rating levels of the selected categories (cf. Scheja, 2017).

4 Results

This section presents descriptive, statistical results and analyzes the two task sets with a view to specific temporal variations. In light of the main question of this paper, the discussion focuses on requirement level distributions across the six requirement-generating categories. Thanks to the developmental perspective, the findings presented here extend beyond previous results on the cognitive requirement of mathematical tasks in the Polish middle school examination (cf. 2.2.2; Scheja, 2017).

The first set of tasks under examination comprises tasks from the introductory phase of the middle school examinations, during which the mathematics section was one component of the mathematics and sciences test section (A1: exam years 2002–2005). The other set of tasks was chosen to show potential effects of the mathematics section's separation in 2012 and the relevant conceptual goals of the CKE on the tasks – also with a view to future developments in Polish students' performance (A2: 2012–2015) (cf. table 5).

		Middle school examinations: Distribution of tasks from 2002–2005 and 2012–2015 with regards to the levels in each category and their levels									
		Set 1 (A1: N= 65)					**Set 2 (A2: N= 92)**				
Category	**Level**		1	2	3	x_{A1}		1	2	3	x_{A2}
Extra-mathematical modeling	Number	5	23	35	2	1.523	50	8	32	2	0.848
	Share in %	7.7	35.4	53.8	3.1		54.3	8.6	34.8	2.2	
Intra-mathematical modeling	Number	61	1	1	2	0.138	44	2	39	7	1.098
	Share in %	93.9	1.5	1.5	3.1		47.8	2.2	42.4	7.6	
Mathematical argumentation	Number	64	0	1	0	0.030	82	1	9	0	0.207
	Share in %	98.5	0	1.5	0.0		89.1	1.1	9.8	0.0	
Using mathematical representations	Number	29	34	2	0	0.585	45	26	20	1	0.751
	Share in %	44.6	52.3	3.1	0.0		48.9	28.3	21.7	1.1	
Working technically	Number	7	39	12	7	1.293	12	44	24	12	1.390
	Share in %	10.8	60.0	18.5	10.8		13.0	47.8	26.1	13	
Processing mathematical texts	Number	6	41	18	0	1.185	4	59	25	4	1.314
	Share in %	9.2	63.1	27.7	0		4.3	64.1	27.2	4.3	

Table 5. Distribution of tasks from the Polish middle school examinations from 2002–2005 (A1) and 2012–2015 (A2) with regards to requirement-generating categories and their levels.

As expected, the modeling process accounts for the vast majority of mathematical work in both sets (A1: 98.4% and A2: 98.9%). However, there are significant differences in the distribution of modeling tasks with regards to their contextual embedding. For instance, assessment tasks from the 2002–2005 period require a much larger amount of extra-mathematical modeling (A1: 92.3%) than those from the 2012–2015 period (A2: 45.6%). Within this category, there is a very high amount of "multi-step modeling" required in A2, whereas simple and complex extra-mathematical modeling only accounts for 9.8%. The A1 tasks are more balanced with regards to their distribution across low (35.4%) and especially medium (53.8%) requirement levels of modeling, which also means they have a lower level of cognitive requirement overall.

Intra-mathematical modeling is only required by four tasks from the first set (6.1%). Compared to the entire collection of tasks from 2002 to 2015 (22.1%) on the one hand (cf. Scheja, 2017) and A2 (52.2%) on the other hand, this points to a drastic conceptual shift within modeling tasks towards intra-mathematical problem contexts with mostly medium (42.4%) and high

(7.6%) cognitive requirement levels. The focus on the first two activities can be accounted for primarily by a respective shift in the underlying exam directives. From 2002 to 2011 – unlike the following ones – they feature only a few isolated intra-mathematical problem contexts. Furthermore, they emphasize particularly those abilities that correspond to extra-mathematical application and problem contexts (cf. 2.1.1). With regards to the medium requirement levels of the first two mathematical activities in A1, it is of particular interest that their share seems to match the high requirement levels of the PISA 2003 tasks (extra-math. modeling: 1.52; intra-math. modeling: 0.20; cf. Neubrand et al., 2013, p. 139). Corresponding effects of exam directives only provide limited explanations for the distribution of mathematical argumentation; it is only required by one task in A1, but has gained significantly in importance in A2 (10.9%) in accordance with the underlying general requirements (cf. 2.1.1). Here, it is typically required by open-ended tasks on low (1.1%) and especially medium (9.8%) requirement levels, while the average requirement level is relatively high (cf. ibid.). On the other hand, the average requirement level for argumentation in the 2002–2005 period is far below that of the other categories as well as that of the PISA tasks (argumentation: 0.15; cf. ibid.). This shows that, contrary to expectations, conceptual directives for assessment tasks that include this ability within the requirement spectrum – if only implicitly – were not properly implemented (ability: “the student identifies correct processes”; cf. CKE 1999). The average level of requirement for the use of mathematical representations has increased slightly in A2, but this value is generally high in both sets (cf. Neubrand et al. 2013, p. 139f.). Among A1 tasks, low-requirement representations are dominant (52.3%), while A2 tasks are distributed almost evenly across the low (28.3%) and medium (21.7%) levels. The distribution across requirement levels in A1 is surprising in how the underlying standards of mathematical and scientific subjects include “finding and utilizing information” as one of the four main standards (cf. 2.1.1), which would presumably lead to a greater differentiation of representations regarding their cognitive complexity.

The average level of requirement for working technically is high in relation to the other activities in both sets; comparing the distributions within A1 and A2 shows minimal deviations, and low- to medium-level calculus is predominant in both sets. The requirement regarding this mathematical activity has therefore contributed to the requirement level of the middle school examinations in a constant fashion over the long term. The average requirement levels of both periods roughly match that of the centralized exam ZP 10, a comparable instrument of external performance assessment from North-Rhine Westphalia wchich are taken at the end of grade 10 in all scholl forms (cf.

Scheja, 2017). Processing mathematical texts is required on a relatively high level (cf. Neubrand et al. 2013, p. 139), which is only comparable with some restrictions. These restrictions arise because the low (A1: 63.1%; A2: 64.1%) and medium language requirement levels (A1: 27.7%; A2: 27.2%) have almost the same distribution, which means that the slightly higher average requirement level in A2 results from a distribution shift in the other two levels (level 0: 9.2% to 4.3%; level 3: 0.0% to 4.3%). As a result, the overall distributions of activities and requirement levels can only be compared between the two sets to a limited extent. Shared task characteristics include the fairly high level of cognitive requirement, which is generated by a fairly broad spectrum of required activities and their average requirement levels. A difference that must be considered a fundamental one since the 2012 reform of the middle school examination can be seen in the amount and requirement level of tasks that involve intra-mathematical modeling and argumentation.

5 Discussion and outlook

5.1 Summary and discussion of findings

This paper presents the findings of a comparative analysis of assessment tasks from the Polish middle school examination (N=157) in the periods after its introduction (2002–2005, N=65) and after its organizational and conceptual reform (2012–2015, N=92). An empirical examination based on a system of categories investigated the extent to which tendencies for change regarding the level of cognitive requirement are present. Based on longitudinal studies of the distribution across requirement levels for mathematical activities, previous findings have been expanded by showing that (and to what extent) the results from Scheja (2017) must be regarded as *indistinct*. First, it turns out that both periods under review are characterized by a relatively high level of cognitive requirement, which is generated by a fairly broad spectrum of required activities and rather high average requirement levels. These properties of mathematical tasks for centralized (final) exams make the findings presented here compatible to those already known in the relevant didactics discourse (Drüke-Noe, 2014; Scheja, 2016; Scheja, 2017).

However, the detailed analysis showed that splitting off the mathematics exam from the mathematics and sciences exam section in 2012 involved extensive conceptual shifts in the assessment tasks. *One* major area of differences can be found in the representation of intra- and extra-mathematical modeling activities; here, we see a strong focus on extra-mathematical context in the introductory period, whereas post-reform tasks are distributed evenly across the two categories. Taking into account the category of technical tasks, which both sets of tasks require only occasionally (cf. Scheja, 2017), the two types of contextual embedding can be regarded as a profile-forming characteristic, which would seem to appropriately represent the profiles demanded by the examination directives of both periods. This shift in focus of the middle school examination after 2012, as well as the observed recent longitudinal shifts in German centralized final exams, confirm the presumed divergence of the two task profiles (cf. Kühn, Drüke-Noe, 2013; Scheja, 2017).

Another difference was shown between the two sets of tasks regarding the contribution of mathematical argumentation to the overall level of requirement: Against all expectations based on the exam directives, which greatly emphasize argumentative skills, this activity has only been a significant factor in terms of cognitive requirement in the exam after the reform. Contrary to the activities discussed above, the requirements in working technically and processing mathematical representations and texts exhibit comparable levels (cf. Neubrand et al. 2013, p. 139).

Just like the core curriculum, the post-reform assessment tasks demand an overall broad spectrum of mathematical activities which contribute to their level of cognitive requirement. Bearing Winter (1995) in mind, this characteristic points strongly to a) a balanced foundation in the three basic experiences of mathematics education, and b) a movement away from a limited requirement profile which highly resembles that of PISA tasks. Within the required spectrum of mathematical activities, there is also a higher average requirement level for all activities with the exception of extra-mathematical modeling, which results in a higher overall level of cognitive requirement in assessment tasks after 2012.

A point of criticism regarding the observed development of the middle school examination is that the mathematical activities are now broadly represented – as demanded in the relevant literature – whereas the requirement levels for argumentation as well as extra- and intra-mathematical modeling are not distributed very well across the entire requirement spectrum. This primarily becomes evident in the portion of tasks with low or high requirement levels due to the fact that these activities are underrepresented in relation to medium-level tasks. In light of the fact that calculus-oriented technical tasks are hardly included at all, this raises the question to what extent the reformed middle school examination can be said to fulfill its primary purpose of assessing the knowledge and abilities visà-vis the core curriculum– especially with regards to lower-performing students. A very interesting question that remains to be answered is in which ways such a low level of requirement control affects work done on meso and micro levels of the Polish education system. It seems mostly plausible that (assessment) tasks themselves can be increasingly considered as a *summative instrument of standardization*, which will likely lead to *washback effects* with regards to teaching practices (cf. Büchter, Pallack, 2012). Evidence for this effect would include, on the one hand, a noticeable focus of lessons on assessment tasks, and on the other hand a steadily growing support industry supplying teaching aids like training books, preparation exams, and lesson plans and teaching materials (cf. Klein et al., 2014; Maier et al., 2011; Smith, 1991). Studies investigating such mechanisms in Polish mathematics education have yet to be conducted, although their existence has at least been suggested (Konarzewski, 2008). However, against the background of the observed broad spectrum of requirements – i.e., relatively high average requirement levels for the sic mathematical activities – paired with areas of suboptimal distribution within at least the assessment tasks' medium level of requirement, it must be assumed that this *type* of effect is only present to a minor *extent* (cf. Skowronek, 1989).

5.2 Outlook

Against the background of these empirical results, normative suggestions for the design of assessment tasks for future centralized examinations, such as the middle school examination and the *eighth-grade examination* that will be introduced with a new set of reforms in 2019 are given below.

Because the assessment tasks from the periods studied here represent their respective directives to a sufficient extent, it seems like a sensible *further* step, to operationalize levels of cognitive requirement with respect to the general requirements. This should aim to leverage the cognitive complexity of the general requirements in such a way as to produce an increasing empirical level of difficulty from one level to another. Those involved in education, including the people who design the assessment tasks, would thus receive an effective tool for designing tasks with specific requirements in mind and to cover Winter's basic experiences of mathematics education not only across their spectrum, but also along the qualitative axis, i.e., vertically (*conceptual* function). A more varied spectrum of mathematical tasks would help the results of the new eighth-grade examination provide a greater range of indicators for student competence. To this end, the design of the tasks used would have to be done in such a way as to require as many mathematical activities on different levels as possible. Fine-tuning the requirement levels to such an effect would enable fine-grained performance evaluation, and centralized examinations with this kind of diagnostic advantage could provide more useful feedback to members of CKE and OKE for lesson planning and quality development (*analytical* function).

References

B a n a c h, C.: 1999, *Ku dobrej edukacji*, Torun: Adam Marszałek.

B a u m e r t, J., K u n t e r, M., B l u m, W., B r u n n e r, M., V o s s, T., J o r d a n, A., K l u s m a n n, U., K r a u s s, S., N e u b r a n d, M., T s a i, Y.: 2010, 'Teachers' mathematical knowledge, cognitive activation in the classroom and student progress', *American Educational Research Journal* **47(1)**, 133–80.

B i a łe c k i, I., H a m a n, J.: 2001, *Program Miedzynarodowej Oceny Umiejetnosci Uczniów OECD/PISA*, Warszawa: Wydawnictwo Naukowe Scholar.

B r o ż e k, A.: 2005, Wyniki sprawdzianów i egzaminów zewnętrznych a podnoszenie jakości pracy szkoły. *in: Kompetencje zawodowe nauczycieli i jakość kształcenia w dobie przemian edukacyjnych,* Wydawnictwo Uniwersytetu Opolskiego, Opole, 239–245.

B ü c h t e r, A., P a l l a c k, A.: 2012, Methodische Überlegungen und empirische Analysen zur impliziten Standardsetzung durch zentrale Prüfungen, *Journal für Mathematikdidaktik* **33**, 59–85.

C K E: 1999, Standardy wymagań będące podstawa przeprowadzenia egzaminu w ostatnim roku nauki w gimnazjum, Warszawa: Centralna Komisja Egzaminacyjna.

C K E: 2010, Informator o egzaminie gimnazjalnym od roku szkolnego 2011/12. http://www.cke.edu.pl/images/files/Gimnazjum_2011_2012/Informator_G1.pdf, Gesehen 07. 10. 15.

C K E: 2011, Ramowy opis polskiego systemu egzaminów zewnętrznych, www.bip.cke.edu.pl/bip_download.php?id=1293, Gesehen 20. 10. 15.

D e n e k, K.: 2005, *Ku dobrej edukacji*, Torun: AKAPIT.

G r ä s e l, C., P a r c h m a n n, I.: 2004, Die Entwicklung und Implementation von Konzepten situierten und selbstgesteuerten Lernens, *Zeitschrift fur Erziehungswissenschaft*, Sonderheft 3: PISA und die Folgen, 169–182.

K a r p i ń s k i, M., G r u d n i e w s k a, M., Z a m b r o w s k a, M.: 2013, *Nauczanie matematyki w gimnazjum – Raport z badania*, Warszawa: IBE.

K o n a r z e w s k i, K.: 2004, *Reforma oświaty. Podstawa programowa i warunki kształcenia*, Warszawa: ISP.

K o n a r z e w s k i, K.: 2008, *Przygotowanie uczniów do egzaminu: pokusa łatwego zysku. Raport badawczy*, Warszawa: ISP.

K u n t e r, M., B a u m e r t, J., B l u m, W., K l u s m a n n, U., K r a u s s, S. N e u b r a n d, M.: 2013, *Cognitive activation in the* *mathematics classroom and professional competence of teachers. Results from the COACTIV project*, New York, NY: Springer.

K u n t e r, M., V o s s, T.: 2013, The Model of Instructional Quality in COACTIV: A Multicriterial Analysis, (Chapter 6), in: *Cognitive Activation in the Mathematics Classroom and Professional Competence of Teachers*, 97–124, Springer.

K u p i s i e w i c z, C z.: 2006, *Projekty reform edukacyjnych w Polsce*, Warszawa: PWN.

K ü h n, S. M., D r ü k e - N o e, C.: 2013, Qualität und Vergleichbarkeit durch Bildungsstandards und zentrale Prüfungen? – Ein bundesweiter Vergleich von Prüfungsanforderungen im Fach Mathematik zum Erwerb des Mittleren Schulabschlusses. Erscheint in *Zeitschrift für Pädagogik*, **6**, 912–932.

M a i e r, U., B o h l, T., M e t z, K., K l e i n k n e c h t, M.: 2011, Einflüsse von Merkmalen des Testsystems und Schulkontextfaktoren auf die Akzeptanz und Rezeption von zentralen Testrückmeldungen durch Lehrkräfte, *Journal of Educational Research Online* **3**, 62–93.

M E N: 1998a, *Reforma systemu edukacji – projekt*, Warszawa: WSiP.

M E N: 1998b, *Ustawa z dnia 25 lipca 1998 roku o zmianie ustawy o systemie oświaty*, Warszawa: Dz.U. Z 1998 r. Nr 117, poz. 759.

M E N: 1999, *Rozporządzenie Ministra Edukacji Narodowej w sprawie podstawy programowej*, Warszawa: MENIS.

M E N: 2008, *Podstawa programowa z komentarzami. Tom 6: Edukacja matematyczna i techniczna w szkole podstawowej, gimnazjum i liceum*, Warszawa: MEN.

M E N: 2013, Programme for international student assessment. Wyniki badania 2012 w Polsce, www.ifispan.waw.pl, Gesehen 14. 12. 2016.

M S W: 2004, *Kernlehrplan für die Realschule in Nordrhein-Westfalen. Mathematik*, Frechen: Ritterbach.

M o l e d a, A., P i e s y k, Z.: 1993, Przegląd zmian programów nauczania matematyki w szkole podstawowej w latach 1963-1990, *Folia Mathematica* **6**, 25–57.

N e u b r a n d, M., K l i e m e, E., L ü d t k e, O., N e u b r a n d, J.: 2002, Kompetenzstufen und Schwierigkeitsmodelle für den PISA-Test zur mathematischen Grundbildung, *Unterrichtswissenschaft* **30**, 100–119.

N e u b r a n d, M.: 2004, *Mathematische Kompetenzen von Schülerinnen und Schülern in Deutschland. Vertiefende Analysen im Rahmen von PISA 2000.* Wiesbaden: Verlag für Sozialwissenschaften.

N e u b r a n d, J., N e u b r a n d, 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 und Oldenburg.

N e u b r a n d, M., J o r d a n, A., K r a u s s, S., B l u m, W., L ö w e n, K.: 2013, Task Analysis in COACTIV: Examining the Potential for Cognitive Activation in German Mathematics Classrooms (Chapter 7), in: *Cognitive Activation in the Mathematics Classroom and Professional Competence of Teachers*, 125–144, Springer.

R e i s s, K.: 2004, Bildungsstandards und die Rolle der Fachdidaktik am Beispiel der Mathematik, *Zeitschrift für Pädagogik* **50**, 635–649.

S a m b o r s k a, M.: 2015, Using Student-Friendly Tasks to promote mathematical understanding in a middle school classroom, *Didacitica Mathematicae* **37**, 47–74.

S c h e j a, B.: 2016, Cognitive activation through mathematics tasks in the context of centralised examinations on ISCED 2 Level, using the example of North-Rhine Westphalia, *Didacitica Mathematicae* **38**, 175–202.

S c h e j a, B.: 2017, Kognitive Aktivierung durch Mathematikaufgaben zentraler Abschlussprüfungen- Eine Vergleichsanalyse der polnischen Mittelschulprüfung und der Zentralen Prüfung aus Nordrhein-Westfalen, *Journal für Mathematikdidaktik* **38(2)**, 291–322 .

S k o w r o n e k, H.: 1989, Denken – Denkerziehung, in: C. Wulf (Hrsg.), *Wörterbuch der Erziehung*, 132–135, München: R. Piper & Co. Verlag.

S m i t h M. L.: 1991, Meanings of test preparation, *American Educational Research Journal*, **28**, 521–542.

T u r n e r, R., D o s s e y, J., B l u m, W., N i s s, M.: 2013, Using mathematical competencies to predict item difficulty in PISA: A MEG Study, in: M. Prenzel, M. Kobarg, K. Schöps, S. Rönnebeck, *Research on PISA. Research Outcomes of the PISA Research Conference 2009*, 23–37, Dordrecht: Springer.

W i n t e r, H.: 1995, Mathematikunterricht und Allgemeinbildung, *Mitteilungen der Gesellschaft für Didaktik der Mathematik*, **61**, 37–46.

W i r t z, M., C a s p a r, F.: 2002, *Beurteilerübereinstimmung und Beurteilerreliabilität. Methoden zur Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Kategoriensystemen und Ratingskale*, Göttingen: Hogrefe.

Z a h o r s k a, M.: 2002, *Szkola miedzy panstwem, spoleczenstwem a rynkiem*, Warszawa: Zak.

Zmiany kognitywnych wymagań zadań egzaminacyjnych z matematyki – Wyniki przekrojowych badań egzaminu gimnazjalnego

S t r e s z c z e n i e

Intensywne dyskusje w obszarze polityki edukacyjnej dotyczące złego funkcjonowania polskiego systemu szkolnego doprowadziły pod koniec lat 90-tych do głębokich jego zmian. Głównymi elementami sukcesywnie wprowadzanych reform było wprowadzenie w 1999 nowej struktury szkolnictwa, co skutkowało stworzeniem nowej podstawy programowej, organizacji centralnych egzaminów i powołaniem komisji egzaminacyjnych. Badania PISA, realizowane w okresie pomiędzy 2000 i 2015 wykazały, że wprowadzone zmiany poskutkowały pozytywnymi wynikami w odniesieniu do nauczania matematyki. W obliczu aktualnie wprowadzanych reform, z dydaktycznego punktu widzenia staje się interesujące zbadanie tych instrumentów sterowania, co do których można przyjąć, że mają pozytywny wpływ na osiągane wyniki. Z badań empirycznych prowadzonych w Niemczech wynika, że na przebieg procesu nauczania wpływ instrumentów, w których manifestują sie *oczekiwane efekty* (np. Curriculum, Standardy itp.) jak i tych, w których odbijają się *obserwowalne efekty* (np. egzaminy) jest duży. Z powodów formalnych można przyjąć, że dla poziomu *lower secondary education* (ISCED 2) takim instrumentem jest egzamin gimnazjalny. Dla przedmiotu matematyka na efekty typu *obserwowalne* wpływ mogą mieć np. struktura zadań egzaminacyjnych jak i cechy generujące ich poziom trudności. Ten to aspekt badawczy był brany pod uwagę w artykule Scheja (2017). Tam koncepcja *kognitywnej aktywizacji* została użyta do analizy i oceny pełnego zestawu zadań z egzaminów gimnazjalnych z okresu pomiędzy 2002 i 2015. Jednak we wspomnianym artykule nie został wzięty pod uwagę wpływ reformy egzaminów, dokonany w roku 2012 ze względu na wprowadzone wtedy zmiany w podstawach programowych, dlatego też wyniki przeprowadzonych tam analiz powinny być raczej oceniane jako niespójne.

Obecny artykuł poszerza to podejście o perspektywę rozwojową. Badania przekrojowe dotyczą zadań matematycznych z egzaminów gimnazjalnych z lat 2002–2005 ($N = 65$) i 2012–2015 ($N = 92$). Został opisany i zanalizowany zawarty w nich poziom kognitywnych wymagań. Jest to poziom kognitywnej aktywizacji, czyli wymiar złożoności odnoszony do kategorii kognitywnych, które potencjalnie mogą być zidentyfikowane w procesie rozwiązania zadania. Dla analizy zadań przyjęto 6 kategorii, w których rozróżniane są cztery poziomy kognitywnej kompleksowości: modelowanie praktyczne, modelowanie matematyczne, argumentacja, wykorzystanie matematycznych reprezentacji,

sprawność wykonywania działań matematycznych i posługiwanie się matematycznym tekstem.

Celem badań jest prześledzenie pod tym względem czasowych zmian obserwowanych w strukturze i stopniu złożoności zadań egzaminacyjnych. Na tej podstawie możliwe jest – poszerzając perpektywę badawczą – szukanie odpowiedzi na pytania dotyczące stopnia powiązania zadań egzaminacyjnych z podstawą programową.

Z przeprowadzonych badań wynika między innymi, że:

- Centralnym typem matematycznych zagadnień w dwóch seriach zadań jest spodziewane *modelowanie*. Duże różnice wystepują jednak w proponowanym kontekście tych zadań. W okresie 2002–2005 mocno dominują zadania umieszczone w kontekście praktycznym, głównie o średnim poziomie kognitywnej kompeksowości. W latach 2012–2015 można stwierdzić koncepcyjne przesunięcie w obrębie modelowania w kierunku matematycznego modelowania o średniej i wysokiej kompleksowości kognitywnej. W tych dwóch kognitywnych kategoriach zaznacza się duży związek zadań egzaminacyjnych z podstawami programowymi.
- W przeciwieństwie do modelowania, potrzeba *argumentacji* jest wymieniona tylko w jednym zadaniu z okresu 2002–2005, zaś pomiędzy 2012 i 2015 w około 11% wszystkich zadań. Tutaj analogiczne – skutkowanie podstawy programowej na poziom tej kognitywnej umiejętności, w szczególności w latach wprowadzania egzaminu gimnazjalnego, jest bardzo ograniczone.
- Średni poziom kognitywny *wykorzystania matematycznych reprezentacji* jest a) porównywalnie wysoki i b) wzrasta lekko w zadaniach po reformie egzaminu w 2012. Zadania z lat 2002–2005 cechują się jednak – wbrew oczekiwaniom – brakiem szerokiego rozłożenia kompeksowości tej umiejętności na wszystkie cztery stopnie.
- Średni poziom wymagań *sprawności w wykonywaniu matematycznych obliczeń,* w porównaniu do pozostałych kognitywnych umiejętności może być oceniony w obu okresach jako wysoki. Odpowiada on zarazem poziomowi tej umiejetności w Zentrale Prüfung ZP 10, porównywalnemu instrumentowi zewnętrznego pomiaru osiągnięć w Nadrenii Północnej-Westfalii (Niemcy) po dziesiątej klasie.
- Średni poziom obu zestawów zadań w *posługiwaniu się matematycznym tekstem* jest a) w porównaniu do badania PISA wysoki, zarazem b) w obydwu zestawach zadań raczej ograniczenie porównywalny. Podpunkt b) jest spowodowany przede wszystkim częściowym przemieszcze-

niem w latach 2012-2015 zadań z poziomu „0" na poziom wysoki w tej kategorii.

Można wieęc stwierdzić, że rozkład zadań na kognitywne kategorie i ich stopnie kompeksowości w zadaniach egzaminu gimnazjalnego jest w dwóch badanych zestawach tylko ograniczenie porównywalny. Do wspólnej charakterystyki zadań zaliczyć można względnie wysoki poziom kognitywnych umiejętności, który jest generowany dość szerokim zakresem aktywności i ich raczej wysokim poziomem. Fundamentalna różnica w zadaniach matematycznych po reformie egzaminu gimnazjalnego w roku 2012 w stosunku do zadań wcześniejszych polega na mierze i poziomie wymagań dotyczących modelowania oraz argumentowania.

University of Cologne
Institut für Mathematikdidaktik
Gronewaldstraße 2
50931 Köln
Germany
e-mail: *schejab@uni-koeln.de*

2.4 Publikation IV

Scheja, B. & Castelli, S. (2018). Developing teacher competence regarding the cognitive requirement of mathematical tasks – a video-based intervention. *Didactica Mathematicae*, 40, 65 – 95.

Abstract

To date, there has not been much research into the practice of planning math lessons and how teachers make didactic decisions regarding the basic dimension of cognitive activation. The present study shows how video-based trainings can be used to help teachers (N=48) expand their categorical knowledge regarding the cognitive requirement of mathematical tasks and use that knowledge in a flexible manner. This is meaningful insofar as it can show whether and to what extent teachers can address such challenges more effectively when planning their lessons. For example, such challenges might arise in the field of individualized learning processes. Finally, implications for optimizing teacher training and for further research requirements are discussed.

DIDACTICA MATHEMATICAE
Vol. 40, 2018, p. 65–95

Bruno Scheja (Köln, Germany)
Sabine Castelli (Bielefeld, Germany)

Developing teacher competence regarding the cognitive requirement of mathematical tasks – a video-based intervention

Abstract: To date, there has not been much research into the practice of planning math lessons and how teachers make didactic decisions regarding the basic dimension of cognitive activation. The present study shows how video-based trainings can be used to help teachers ($N = 48$) expand their categorical knowledge in terms of cognitive requirement of mathematical tasks and use that knowledge in a flexible manner. This is meaningful insofar as it can show whether and to what extent teachers can address such challenges more effectively when planning their lessons. For example, such challenges might arise in the field of individualized learning processes. Finally, implications for optimizing teacher training and for further research requirements are discussed.

1 Introduction

As teaching and learning depend on the relevant domains, there is currently consensus in the relevant disciplines that a firm foundation of specialized didactic knowledge is essential to the instructional quality in any subject (cf. e.g., Ball et al., 2001; Kunter et al., 2013; Kunter, Voss, 2013; Grossman, Schoenfeld, 2005; Mewborn, 2003; Schoenfeld, 1998; Stigler et al., 1999). This knowledge determines the extent to which teachers succeed in creating the

Key words: professional development, secondary education, cognitive processes, task design, teacher education.

right kind of structures, which allow students to experience learning processes that enhance understanding. These structures provide an appropriate level of motivation in math lessons and include what Oser and Baeriswyl (2001) call sight structures and deep structures of teaching. With regard to deep structures, studies show that especially the *potential for cognitive activation* in teaching processes and, accordingly, the selection of tasks which feature cognitive activation, are beneficial to the performance development of students (Kunter, Voss, 2013). Against this background, the following two points seem surprising: *First*, current empirical studies on the tasks used in math lessons continue to identify a low level of cognitive requirement (cf. Neubrand et al., 2013; Drüke-Noe, 2014; Karpinski et al., 2013). For the obviously necessary improvement of mathematical instruction, we can ,therefore, assume that teachers would benefit from a conscious familiarization with various features that contribute to the cognitive requirements of mathematical tasks. *Second*, and against all expectations, it seems that the criteria for designing tasks are still hardly featured in teacher training and that any properties of tasks are hardly known beyond the level of content (cf. Drüke-Noe, 2014, p. 250).

This research gap is addressed by the present study, which first assesses the categorical knowledge about requirements, and then expands it by means of a video-based training that comprises several steps in order to make it usable when selecting, modifying, or designing mathematical tasks. This teacher training concept would be a prudent addition to existing training programs for teachers in North-Rhine Westphalia across all levels (individual, micro, and meso level). Based on existing findings (cf. Kunter, Voss, 2013), we forgo an accompanying evaluation study that would monitor the development of student competence as a means of assessing the effectiveness of the program regarding teacher performance.

2 Theoretical background

2.1 Assumptions regarding the nature of effective teacher training programs

Over the last 20 years, more and more training concepts have been developed and empirically examined. Their effects can be assessed according to several aspects (cf. Guskey, 2000), which Lipowsky (cf. 2010) arranges into *four levels*:

Level 1: Immediate reactions and opinions shared by teachers who participated in the training

Level 2: Expanded teacher cognition after the training

Level 3: Changes in teaching practice of the participants

Level 4: Effects of teacher training on the students development of performance or motivation

Across all of these levels, the development of teachers' cognition is influenced by several success factors of the trainings. According to Lipowsky (ibid.), these factors can be divided into two main categories: *Structural* aspects (cf. Garet et al., 2001; Yoon et al., 2007) and *methodical* or *didactic/content-related* aspects. Dreher et al. (2017) list three relevant structural aspects: format (conferences, workshops or networks), duration (a minimum of contact hours increases the long-term effectiveness of a training)[1] and collective participation (e.g., several teachers of one faculty taking part together) (cf. Garet et al., 2001; Yoon et al., 2007).

With regards to the methods used in a training, studies show that considering cooperation and professional learning groups has a positive impact on the effectiveness of trainings. Especially regarding active participation, Maldonado (2002) talks about "collaborative grouping" and "establishing learning communities", which foster a deeper reflection about lessons and the learning processes of students. Effective teacher trainings should also take matters of content and didactics into account. Against this background, trainings turn out to be more effective if they are domain-specific and limited to or focused on a narrow range of content (cf. Maldonado 2002). "Successful programs should not address 'generic' learning, but instead focus on the learning of particular mathematical ideas" (ibid., p. 8). As a result, it is assumed that – if the content is focused on one specific subject – teachers can apply what they have learned in their lessons: "Context-specific approaches promote teaching practices that are consistent with the principles of effective teaching but also assist teachers to translate those principles into locally adapted applications. By developing this kind of knowledge teachers can better solve identified issues about student outcomes in their particular teaching situations" (Timperley, 2008, p. 10).

Lipowsky and Rzejak conducted a study about the success of a training, which showed a variety of influencing factors. The research question was, which kind of characteristics influence the effectiveness of trainings. The teacher's characteristics, such as expectations, goals, attitude towards own learning and student's learning, prior knowledge, specific motivation towards attending this training and the teacher's willingness to use the learned content play a major role for the individual success of a training. The contextual conditions like the

[1] Yoon et al. (2007) cite a training duration of at least 14 hours for lasting effects on level 4.

concept and the design of the training and the integration and support of the school may influence the success as well. The effectiveness of a training can be measured by the immediate reaction of the participants, the learning progress of the participants, the changes in teaching and the impact on students (cf. Lipowsky, Rzejak, 2012, p. 2ff). However, not all participants benefit of a training to the same extent, because of individual motivations, prior knowledge and self-efficacy expectations (cf. ibid., p. 11).

Furthermore, Lipowsky and Rzejak investigated a variety of criteria to decide, which characteristics distinguish an effective teacher training. To achieve an effective teacher training, the pedagogical content knowledge and diagnostical knowledge of the participant have to be expanded. An effective training is also the realization that teaching can be changed, and these changes will have impact on students, just as a permanent expansion of professional teaching and giving space to deepen conceptual comprehension, gain new knowledge, changing behavior patterns and reflecting with other participants about that (cf. Lipowsky, Rzejak, 2012, p. 5ff).

While the present study focuses on effects of teacher training on teachers' cognition (level 2), its conclusions can also be considered relevant for their teaching practices on the respective level. The aforementioned success factors for teacher training are essential to the design of the video-based teacher trainings outlined in this study (cf. sections 3.1–3.5).

2.2 Impact and effectiveness of video-based trainings

There are many studies on the impact and effectiveness of video usage in teacher education and training of (mathematics) teachers (e.g. Blomberg et al. 2014; Mitchell, Marine, 2015; Sherin, van Es, 2009; Star et al. 2012; Stockero, 2008). Those studies consistently show the positive impact of the meaningful use of video sequences at certain times of education for the competencies of reflection of participants as well as for the professionalism of teachers in general. For example, Blomberg et al. (cf. 2014, p. 443) resume that videos can be a valuable tool in teacher training, but only under the condition of fully integrating it into teaching syllabuses. Mitchell et al. (cf. 2015, p. 551) found out that participants of video-based training could appreciate important aspects of mathematics in general and in specific. For this study in particular, the study of Sherin and van Es is interesting. The authors investigate, whether teachers will develop a professional perspective and establish certain competencies, if they participate in a video-based training. Their results show that the participation will affect the professional vision of teachers (cf. Sherin, van Es, 2009, p. 20). The study of Star et al. (cf. 2011, p. 117) shows that the observational abilities of teachers will increase after a video-based action. Particularly, the abilities to recognize characteristics of the teaching environment and the mathematical content will improve. Stockero (cf. 2008, p. 373) has indicated similar results. She concludes that participating teachers will show significant changes on their reflection competences.

Brouwer (cf. 2014, p. 180) conducted a literature review of empirical studies published since 2000 and analysed the effects, processes and conditions of video usage in the teacher education and training. He took 388 empirical studies into account, mainly from the USA, Germany and the Netherlands. Brouwer only included studies, whose research question, data and results refer to the professional learning of teacher with digital videos.

Brouwers results are significant: more than 60% of the studies measure the effect of video usage on the thinking and teaching of the participating teachers (cf. Brouwer 2014, p. 182f). One can name different effects on altered thinking: the altered recognition of effective behavioural patterns, the strengthened capability to analyse the interdependencies of teaching and learning of students as well as the increased efforts at the lesson planning (cf. Brouwer, p. 184).

At those video-based interventions, in which changes in teaching could be registered, are a variety of influential characteristics including (amongst others) the discussion of videotaped lessons with colleagues and/or a coach as well as the rehearsal of specific skills (cf. Brouwer 2014, p. 183).

Brouwer sees a central element of the teacher education and training in the connection of professional thinking and action. He criticizes that teachers are expected to just apply the didactical concepts they've learned (cf. Brouwer 2014, p. 177). By increasing usage of videos in teacher education and training, one can observe a different approach to bridge the gap between theory and praxis. The digital tool „video“ is suitable, because four characteristics are present: video usage in teacher education and training can channel the interest to specific aspects; statements are given a unique concreteness by videos and thereby domain specificity; videos are creating a representative experience and an emotional compassion for the viewer; one can repeat the analysis of content from different perspective without pressure to immediate action (cf. Brouwer 2014, p. 177).

Vondorvá (cf. 2016, p. 699) examined the effect of video-based interventions on the knowledge-based thinking of future mathematics teachers. She was interested in the impact of video-based interventions on the thinking as part of a professional vision. Also based on other studies, one result is that the perception of teachers (here primarily: teachers in preparation) is being influenced in a major way by video-based interventions (cf. ibid., p. 712).

2.3 Cognitive Activation in the Context of Instructional Quality

One influential approach to simplifying the notion of quality while at the same time enabling fine-grained analyses is attributed to Donebedian (1966). This approach introduces the distinction between quality of structures, processes, and results as the main dimensions of quality.

Bloom also worked on a classification scheme of learning domains in consideration of demanded cognitive processes in 1956 (cf. Bloom 1994). Up until now his six-stage taxonomy of learning domains in the cognitive domain (knowledge, comprehension, application, analysis, synthesis and evaluation) and modified versions are being used and considered necessary.

The consideration of learning domains within the meaning of his taxonomy emphasizes the necessity to integrate learning domains, which require a higher level of cognitive abilities, which lead to a deeper learning and transfer of knowledge and abilities. This leads to a bigger variety of tasks and contexts. One aim of using the taxonomy in teacher education is to encourage teachers to gradually realize ambitious activities, therefore to initiate and realize more complex cognitive processes in class. An orientation along the hierarchy of learning domains enables a reflexion towards the level of difficulty of a task and their imparted content. By considering the stages of the learning domain, necessary competencies to manage the task, can be described in more detail. Over decades, Bloom has provided an important contribution to the professionalization of teachers and he has also influenced working with tasks.

Turning towards the question which factors contribute the most to the quality of results in teaching, Oser and Baeriswyl (2001) point to sight structures and deep structures. The perspective of describing lessons using sight structures draws from their overarching organizational features, which according to Kunter and Voss (2013, p. 99) include framework conditions, observable instructional arrangements and teaching methods. The individual properties of deep structures can vary independently and relate to the processes of interaction between teachers and students, among students, or between students and the material being taught. Current studies in empirical teaching research show that sight structures have much less explanatory power than deep structures when it comes to the learning progress of students (cf. Hattie, 2012).

In the German-speaking world, a threefold distinction of basic dimensions has become prevalent for the systematic description of deep structures: potential for cognitive activation, classroom management, and individual learning support (Kunter, Voss, 2013; Riecke-Baulecke, 2017). By now, there are several studies that demonstrate how these three dimensions affect specific aspects of learning. For instance, cognitive activation has been shown to have the greatest influence on the performance of students (effect strength: 0.32*) (Kunter, Voss, 2013, p. 115). The degree of cognitive activation in turn depends on the extent to which students are challenged to deal with a topic in depth and on their own, changing and expanding their own knowledge structures. In math lessons, the tasks used are the primary mediators of this effect (Jordan et al., 2006). They form a link between the normative curricular structures and the professional actions of teachers on the one side and the individual learning processes of students on the other side (Neubrand et al., 2013). They structure the learning opportunities on the level of mathematical work, which are then specified and worked out in more detail during the teaching discourse. As a result, the tasks form the framework for generating learning processes and define the basic structure of potential learning opportunities in their role as "*bearers of cognitive activities*" (Jordan et al., 2006). Learning activities can be described as „cognitive activating", if all students can be stimulated to an active discussion about content on their own individual level. So, if the teacher stimulates the learners on a high level, and if he is picking up on the prior knowledge and if he gives the opportunity to explain individual ideas, concepts and solutions while being flexible, the lesson is cognitive activating (cf. Leuders, Holzäpfel 2011, p.212).

In research cognitive activating teaching can be operationalized differently by characteristics of tasks, by visible characteristics of teaching or by the perception of teachers and students (cf. Leuders, Holzäpfel 2011, p. 214). Against this background, we can conclude that mathematical tasks are an effective indicator for the basic dimension of cognitive activation and can therefore serve as a source of indirect information about instructional quality.

2.4 Mathematical Tasks as a Control Device for Teachers

Neubrand defines a task as a teacher's instruction to work on a defined mathematical topic with a specified goal (2002, p. 16). The complexity of the relations involved can range all the way up to general relations (cf. e.g., Jordan et al., 2006), as these also require dealing with a specific mathematical situation. Regardless of the quality of the required activities, tasks can be functionally regarded as a crucial link between the curricular framework and the professional knowledge of teachers on the one hand and the learning processes within a group of students on the other hand (cf. Neubrand et al., 2013, p. 128). They serve to control the level of motivation as well as the creation and development of students' knowledge networks. Luthiger calls tasks that are embedded in learning situations like this *learning tasks* (2012, p. 5), which on the highest level of abstraction include introductory and practice tasks (cf. Abrah, Müller, 2009, p. 6). Learning tasks form an interface of activity and communication between teachers and students, which makes them an essential tool of lesson design.

In a study about the reflection of creating tasks as one way of improving teachers´ professionalism, Hošpesová and Tichá have shown that teachers ability to design tasks (i.e. designing new tasks, pre-formulating specific tasks and changigng parameters) ist part of teachers´ set of professional competences and also contributes to the development of a key element for the professional development of teachers trainees: qualified pedagogical reflection (cf. 2010, p. 122 f). Future teachers should evolve this. The authors are accentuating the necessity of actions, which are focused on building tasks, as inevitable in teacher education. One indication of the competence to build tasks is the competence to detect mathematics. The competence to build tasks is a way to improve teacher competence altogether (cf. Hošpesová, Tichá, 2010, p. 123). In their study of conducting teacher competences they found out that participants worked and argument on building and assessing tasks homogenous regarding situational contexts, quality of environment and interpretation. They are often incapable to use representations and transfer between them (cf. Hošpesová, Tichá, 2010, p. 124).

According to Drüke-Noe et al. (2017, p. 209), tasks are intended to build, consolidate, and deepen or practice knowledge and skills. This requires appropriate tasks, which can be used as "flexible, broadly applicable, and actively configurable content-related and didactic elements that serve to structure mathematics instruction" (Neubrand et al. 2013, p. 126) in order to serve as bearers of cognitive activities. The intended purpose of a task ultimately determines the broad spectrum of its possible uses, ranging from types of tasks with homogeneous solutions, which only need to be adapted to a certain specified case, all the way to completely open-ended tasks. Depending on the current phase of teaching, this spectrum of tasks is a formative element of introducing, practicing, and deepening mathematical contents and abilities (cf. Luthiger 2012, p. 5).

Unlike learning tasks, *assessment tasks* serve the purpose of certifying which level of skill has been attained during the educational process (Terhart et al. 2009, p. 24). Assessment tasks for centralized exams are designed externally. Depending on the functionality of individual tasks, task design is supposed to operationalize the curricular goals in different ways (cf. Büchter, Pallack, 2012). The (assessment) tasks used in local exams, in turn, are the key instrument of performance evaluation in a particular school. When designing or selecting tasks for these exams, teachers tend to act according to their own lessons and the curricular goals, favoring tasks which they expect to provide an adequate representation of the level of achievement in their class (cf. Blum, Neubrand, 1998; Lenné, 1969). As a result, tasks in local exams reflect the types of tasks used in the lessons and can provide information about the normative expectations of a teacher with regards to what their students should have learned. This is adequately reflected by what Jordan et al. (2006, p. 11) call the "summative core of a lesson" in this context. Regarding the purpose of assessment tasks, note that the cognitive activities that learners may experience during the solution process can be interpreted as "expected lesson results".

Contrary to expectations, there is strong evidence based on extensive analyses of structures and requirements which suggests that learning and assessment tasks share mostly identical cognitive requirement characteristics (Drüke-Noe 2014, p. 8). The only differences are in the treatment and evaluation of solutions (cf. Büchter, Leuders 2005).

2.5 Cognitive Requirement in Mathematical Tasks – Conceptual Developments

The current debate about cognitive activation in mathematics lessons is quite extensive (cf. Hiebert et al., 2003; Jordan et al., 2006; Neubrand et al., 2013; Stigler, Hiebert, 1999; Stigler et al., 1999). Due to different operationalizations, the findings are only comparable within certain limitations according to Leuders and Holzäpfel (2011). For instance, Hugener et al. (2007) consider a lesson to feature cognitive activation

- if the teacher lets students explain their own ideas, concepts, and solutions and treats this output in a flexible and evolving manner,
- and if they use tasks to foster the students' thinking on a high cognitive level.

Math lessons with good cognitive activation can thus be characterized either by using observable features, which according to Oser and Baeriswyl (cf. 2001) are primarily found insight structures, or by using features of the tasks employed, which are part of the deep structures. Following this second approach, Neubrand et al. (2002, p. 105 ff) identified properties that make tasks more difficult by analyzing 117 tasks used in PISA-2000 (PISA international: 31 tasks; PISA national: 86 tasks) from a cognitive science point of view. These properties include the curricular knowledge level, the context given for a task, and the complexity level of the required modeling. In addition, it has also been shown that the openness of the modeling process, the necessary extent of processing, and elements of argumentation within the modeling process contribute to the cognitive requirement of a task.

Neubrand et al. conclude that the cognitive complexity of mathematical tasks, and therefore their empirical difficulty, increases with a growing number of cognitive processes that must be completed on the path to a solution (cf. ibid.).

Based on work by Neubrand et al. (2002), the COACTIV project defines cognitive activation via the potential inherent to the relevant tasks. Tasks that have this kind of potential are supposed to offer students opportunities for learning during lessons, encourage them to engage in goal-oriented cognitive activities relevant to the subject, prompt them to think more deeply, and get them to actively engage their minds (Neubrand et al. 2013, p. 129 f). Tasks that fulfill these criteria are particularly suitable for changing or expanding existing knowledge structures, create new ones, and form connections within them. The COACTIV project assesses the degree of cognitive activation, i.e., the *level of cognitive requirement,* according to four mathematical activities required for solving a task:

- extra-mathematical modeling,
- intra-mathematical modeling,
- argumentation, and
- using mathematical representations (cf. Jordan et al., 2006).

This approach forms an instrument for describing the level of cognitive requirement that can describe the potential of mathematical tasks drawing only upon properties of those tasks. In a study of task culture in German-language math instruction, Drüke-Noe also shows that the four COACTIV assessment categories require an extension (2014, p. 66 ff), because they do not differentiate between levels of processing calculus; for instance, the use of numbers and symbols required for solving tasks can vary a lot in terms of cognitive complexity, resulting in varying levels of cognitive requirement in the respective tasks. Among others, Cohors-Fresenborg et al. (2004) provide approaches to operationalizing the use of calculus. These approaches take evaluation elements, specifically the "degree of processing" into account. However, unlike Drüke-Noe's technical work (cf. 2014), they do not fully meet the COACTIV criteria for an assessment category, because, among other things, they are operationalized using a different number of requirement levels.

The system of categories discussed above, which is highly inferential in parts, has proven quite useful and is suitable for describing the requirement level of tasks by means of identifying certain task properties. Unsurprisingly, a reliable application of this system requires appropriate training for the respective teachers (cf. Jordan et al., 2006). The question whether and to what extent it is already used in teaching with regards to the breadth and complexity of cognition – say, in terms of Winter's basic experiences of general education (cf. 1995) – is addressed in the following subsection.

2.6 Findings on the Cognitive Requirement of Teaching Tasks

The number of studies focusing on the identification of cognitive requirement levels in teaching and assessment tasks has grown steadily, especially in recent years. Primary examples regarding externally designed assessment tasks include the work of Neubrand et al. (cf. 2002, 2013), Drüke-Noe (2014), as well as Scheja (2016, 2017a, 2017b). Unlike the admittedly vast canon of existing empirical research into externally designed assessment tasks, only a few studies have investigated the level of cognitive requirement of learning tasks and assessment tasks conceived within individual schools. So far, the findings in this area are in agreement. With regards to the type of mathematical activity (cf. Neubrand et al, 2002), they show that nearly half the assessment tasks used in local tenth grade exams are technical tasks which only require factual knowledge or skills. The remaining tasks require modeling (43 % procedural and 8 % conceptual), which only require argumentation in 1/4 and inner-mathematical modeling in 1/3 of cases. Without going into detail about the categories of cognitive requirement, Kunter et al. (2006) conclude that the overall requirement level is low and that context-free calculus is highly significant.

Further evidence of an overall very low requirement level can be found in a study of homework tasks, introductory tasks, and class tests from two periods (more than 45,000 tasks from 2003 and 2004) from various school forms with regards to four activities (extra-mathematical modeling, inner-mathematical modeling, argumentation and using mathematical representations) conducted by Neubrand et al. (2013, p. 137 f). It becomes apparent that the aforementioned tasks, and therefore the German system of math instruction, hold very little potential for cognitive activation. For instance, the tasks do not require argumentation at all, while extra- and intra-mathematical modeling turn out to only be required on a very low level. In terms of using mathematical representations and processing mathematical texts, the situation is equally dire, with very little differentiation across requirement levels. Furthermore, Neubrand et al. show that the three types of tasks under examination barely differ in terms of their level of cognitive requirement (ibid.).

Maier et al. (cf. 2010, p. 84) point out that tasks are a central element for planning lessons and are significant to the capability of assessing the cognitive activating potential. The competence to assess tasks suitable by its level of difficulty is significant and has an impact on the success of teaching (cf. Maier et a. 2010, p. 94). The estimation of the cognitive potential of tasks is also a central characteristic of diagnostical competence of teachers (cf. Maier et al. 2010, p 94). More than 60% of the study is limited to tasks, which require only skills of reproduction. The evaluated learning and audit tasks mostly had low cognitive requirements, so (future) teachers should be able to refine tasks and use them on its periphery (cf. Maier et al. 2010, p. 94). Knowledge of benefits of a category system of cognitive requirement may help with that.

The widespread lack of tasks that involve cognitive activities on medium or high requirement levels, which would actually foster active integration of content, raises the question of why there is so little variation. Highlighting possible causes is of interest insofar as it has also been shown that the *use of tasks with high levels of cognitive requirement is beneficial to the performance of students* (cf. Kunter, Voss, 2013). Drüke-Noe argues that one possible reason for the current use of tasks could be that the literature mentions criteria-based design of learning and assessment tasks only rarely (2014, p. 249). This would suggest that teachers are hardly, or perhaps not at all, trained to use these methods, which would lead to an unsatisfactory level of awareness for task properties beyond the level of content. Current studies with classifications across subjects would seem to support this claim (cf. Drüke-Noe et al. 2017). With regard to mathematics, the question whether teachers know, understand, and are able to use such classification structures for planning and evaluating their lessons, remains unanswered. This research gap is addressed in the present study based on the following three research questions:

- *Which properties of cognitive requirement in mathematical tasks do teachers know?*
- *To what extent are teachers able to actively adjust these properties of cognitive requirement with regards to cognitive complexity?*
- *To what extent is the knowledge which teachers hold about the properties of cognitive requirement extended and can it be used in a flexible manner?*

3 Method

A concept for a video-based teacher training was devised based on the theoretical and empirical findings outlined above. The aim of this training, which is described below, is to use a system of categories to help teachers recognize the level of cognitive requirement of mathematical tasks and to expand their categorical knowledge. To fully explain the training concept, the following subsections present the categories of cognitive requirement (3.1), outline the design and organizational structure (3.2 and 3.3) and describe its implementation (3.4 and 3.5).

3.1 Categories for Assessing the Cognitive Requirement of Mathematical Tasks

We will be basing our efforts to assess and promote the level of cognitive requirement in mathematical tasks on a selection of empirically proven classification categories, which were originally developed as part of the PISA and COACTIV projects (cf. OECD, 1999; Jordan et al., 2006) and further expanded by Drüke-Noe (cf. 2014). Building on the preliminary work by Neubrand et al. (2002), the latter two studies included large-scale analyses of cognitive activation through teaching and assessment tasks used in the ninth and tenth grades at German schools. The assessment categories chosen for this study are further distinguished into four levels, which describe the respective degrees of cognitive complexity within each category (cf. Tab. 1)[2]. The categories themselves range from low (e.g., using mathematical representations) to medium level inferential classifications (mathematical modeling), while highly inferential categories are excluded due to a complex overarching didactic concept.[3]

The underlying categories for sources of requirement operationalize mathematical activities[4] on four cognitive requirement levels each. These levels express whether an activity is required on a high (3), medium (2), or low level (1), or not at all (0) for solving a task. The degree of cognitive complexity involved in an activity is expressed by these levels.

A detailed description of the five assessment categories, the relevant theoretical background, and numerous example tasks can be found in the existing literature (Jordan et al., 2006; Neubrand et al. 2013; OECD, 1999).

[2]These categories are conceptually equivalent to four of the general mathematical competences referenced in the education standards for mathematics (cf. KMK, 2004).

[3]For instance, the basic concepts category used in both PISA and COACTIV is not applied here.

[4]The category "using mathematical representations" is not strictly speaking a mathematical activity, because several regression analyses have not shown any contribution to the difficulty level of a task. However, this category is included as a mathematical activity due to its significance for international comparative studies.

Cognitive category	Level of cognitive activity
Mathematical modeling	0 – not required, 1 – reproduction, standard modeling, 2 – connection, multistep modeling, 3 – generalization, reflection on a model, development and validation of complex model
Argumentation	0 – not required, 1 – standard reasoning, 2 – multistep argumentation, 3 – development of complex argumentation, proofs, evaluation of arguments
Using mathematical representations	0 – not required, 1 – standard representations, 2 – switching between different representations, 3 – evaluating representations
Working technically	0 – not required, 1 – standard technic, 2 – simple hierarchical operations, 3 – complex hierarchical operations
Processing mathematical texts	0 – not required, 1 – direct text comprehension, 2 – text comprehension with reorganisation, 3 – comprehension of logically complex texts

Table 1. Pivotal categories of this study (cf. OECD, 1999; Neubrand et al., 2013; Drüke-Noe, 2014.

3.2 Designing and Organizing a Video-Based Teacher Training

The aim of the video-based training described here is to use a system of categories to help teachers recognize the level of cognitive requirement of mathematical tasks and to expand their categorical knowledge. It is intended to expand teachers' professional knowledge by the core competence of selecting, modifying, or even designing adequate learning and assessment tasks based on the criteria of features which contribute difficulty. The system of categories shown in table 1 comprises cognitive activities for describing these kinds of task characteristics in order to conceptually comprehend the level of cognitive requirement. All in all, the study is based on research by Drüke-Noe et al. (2017, p. 209 ff). However, it does not apply subject-independent categories, but investigates cognitive activities specific to the subject of mathematics. The study was conducted between February 2017 and July 2017 and comprises three main phases: *preliminary research* (February and March), a *four-step video-based training* (April to June) and *concluding research* (July).

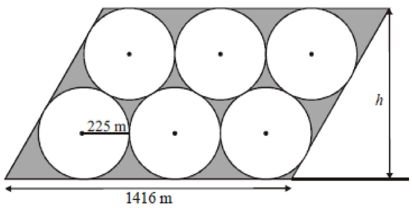
3.3 Preliminary Research

Neubrand et al. (2013) and Drüke-Noe (2017, p. 213) identified research gaps regarding the question whether teachers know about structures for describing and analysing tasks and whether they are able and willing to use these structures when planning or reviewing their lessons. This is where this study continues by assessing teachers' cognition with regard to tasks. It is expected that the participating teachers can name some task characteristics and apply them to the tasks presented based on their familiarity with the categorization from the standards of education (KMK, 2004), while not applying *all* available characteristics and failing to employ the breadth of the categories discussed in this study.

The first question, which serves to establish the prior knowledge of difficulty-generating task characteristics, is: *"Name characteristics according to which you determine the difficulty of a mathematical task."* Next, the participating teachers are asked to *apply* the features of cognitive requirement they listed to mathematical tasks by modifying them to adjust the difficulty level.

Task 1

During periods of drought in the summertime, many American farms use irrigation systems. They are firmly attached at the center, while long "arms" with water pipes on large wheels ride over the fields in circles. The image shows an arrangement of such irrigation devices.



Calculate the percentage of field area that is not irrigated (marked in grey). Assume that $h = 840$ m.

Task 2

Julian and Anja are wondering about the maximum number of intersections between seven straight lines. Julian says: "It's simple – I can even tell you, how many intersections there can be between 100 straight lines." How did Julian figure out the solution?

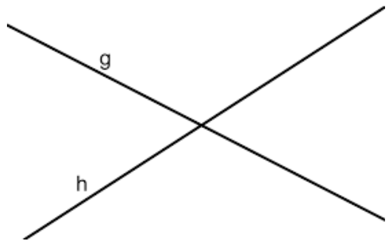


Figure 1. Tasks from the preliminary research for this study (task one taken from MSW, 2008.

The participants are offered two tasks (cf. Fig. 1); one task that requires extra-mathematical modeling, and one that requires intra-mathematical modeling (cf. Neubrandt et al. 2013). The teachers are told to alter the intra-mathematical task in a way that lowers its requirement level, and to alter the extra-mathematical task both in a way that lowers and in a way that increases its requirement level. This is to be done by increasing or reducing the cognitive complexity according to one of the previously discussed categories. The task is phrased as follows: *"Using the following task, explain how you would alter the task characteristics you described before to make it easier/harder."*

3.4 Video-Based Intervention

The idea of a video-based training of teachers is based on the firm belief that a system of categories can help increase awareness for different types of tasks and to effectively employ tasks that belong to the categories (cf. Drüke-Noe et al., 2017, p. 219). Kleinknecht et al. (2014) have conducted a study about the necessity that teachers should be able to verify and change tasks individually, but criteria are missing. These criteria are important for the diagnoses, the support and reviewing the learning processes. In addition to Kleinknecht et al. (cf. ibd. 2014, p. 137f), this study investigates the criteria that teachers use to evaluate the difficulty of tasks and how teacher training about criteria for a content-pedagogical analysis can be supported to a better targeted assessment of tasks. But the system of catalogue of task characteristics, which is interdisciplinary and pedagogy in general, as well as the training, which has a classical setting and is focusing on working on tasks, varies.

Kleinknecht et al. (cf. 2014, p. 140) are viewing tasks as significant indicators of the difficulties of teaching. Therefore, the task analysis is enabling to determine the level of available learning opportunities in the class by standardized tools.

Even though the focus is on the analysis of the potential of tasks, it is important to the quality of class, how tasks are used in the class and how students use the range of tasks (cf. ibd. 2014, p. 140f).

Teachers may simply require the right concepts to analyse and describe tasks according to the processes of cognition and learning they bring about in students or to accurately assess the potential for cognitive activation of a task.

Assuming that – as suggested by previous studies – a catalog of task characteristics can serve to determine the difficulty levels of tasks, there needs to be specific training to show teachers how to employ this system of differentiating categories. Training teachers to use the system of categories is already necessary for conveying unambiguous definitions of the concepts involved and then applying them in an explanatory way to sample tasks.

Starting with the classification categories and the four levels outlined in table 1, the videos should demonstrate tasks that highlight what each category entails and how it can be utilized to identify or adjust the level of cognitive requirement in any given task. To this end, examples of tasks from various levels in each category are explained and rated. Teachers should also be trained in applying their categorical knowledge to select, design, or modify appropriate learning and assessment tasks. They should be able to apply their professional knowledge when planning, conducting, and evaluating their lessons, for instance by means of internal differentiation.

The authors decided to settle on video-based training, because of the effects presented in chapter 2.2. The video-based training is divided into four sequences, which are made available to the participants on a portable medium as well as in a wiki once the preliminary research is concluded. In order to provide the best possible support, manuals with further explanations and additional material for sequences two through four are provided as well. The *first sequence* (duration: approx. 29 minutes) introduces and explains the cognitive activities of the system of categories and the four levels. Alongside increasing transparency regarding the training, it also involves clarification and standardization of terms (e.g. “cognitive requirement”: impacts within the rage cognitive categories of different states, which are potentially activating at solving processes of tasks), the necessity and importance of cognitive activation, the presentation of the COACTIV study and the detailed analysis and explanation of the category system. In the *second sequence* (duration: approx. 38 minutes, incl. accompanying print materials), the categories are used to assess the level of cognitive requirement for fitting tasks by showing how to rate them. For each category of registration and for each category of requirement is a coded sample task. All terms and the whole process are being explained. The *third sequence* (duration: approx. 46 minutes, incl. accompanying print materials) is about modifying existing tasks. Teachers learn to alter specific characteristics in order to increase or decrease the difficulty of a task. There are specifically chosen tasks, which are changing in many states of requirements. The *fourth sequence* (duration: approx. 22 minutes, incl. accompanying print materials) is about using the system of categories for preparing everyday lessons in school.

A video-based approach was chosen for the training because this format does not require physical attendance and allows the greatest planning flexibility for each individual participant. Another advantage of educational videos is that the participants can select specific video sequences to re-watch as often as they like in order to deepen their understanding of the subject matter.

3.5 Concluding Research

We expect that teachers can expand their existing knowledge identified in the preliminary stage to include the system of categories by taking part in the video-based training. By dealing with the categories in depth, the participants should have gained or expanded their ability to describe and modify the cognitive requirement of tasks in order to use them for their respective groups of students.

In the concluding research phase, teachers are asked to describe their prior knowledge about difficulty-generating features of tasks, just like in the preliminary phase. The respective task is phrased as follows: *"Name characteristics according to which you determine the difficulty of a mathematical task."* Next, the participants are asked to once again apply the features of cognitive requirement they listed to the two mathematical tasks by modifying them to adjust the difficulty level up or down. We expect that teachers list a greater number of more differentiated task properties than in the preliminary phase, and that they know how to employ these properties.

3.6 Study Design and Sample Selection

Comparing the results from the preliminary and concluding stages provide first insights into the development of the participants' way of thinking about tasks. The researchers involved in the project hope to gain insights into how the teachers' understanding of tasks and their difficulty has changed (cf. Kleinknecht et al., 2014, Drüke-Noe et al., 2017). To this end, the answers of the teachers are put into context with the system of categories. In the first step of evaluation, the categories are applied *deductively* to sort the characteristics named by the participants. Any responses that cannot be sorted into any of the categories of the system require the *inductive* development of new categories. The training concept outlined above was tested once in advance. The 48 participants are fully trained teachers working at schools in North-Rhine Westphalia, which all belong to the ISCED 2[5] level. The participating teachers aren't inexperienced. Besides Mathematics. they all have a second teaching subject. The participants (male and female) are between 30-60 years old. To utilize the established structural properties of trainings ("collective participation"; cf. 2.1) for increased benefits to the participants, and their respective schools, three to five teachers from each of the 11 participating schools were included. Due to the *fundamentally exploratory character* of the study, the results were empirically evaluated without further differentiation by the teachers' personality traits or their type of school.

[5]International Standard Classification of Education, lower secondary

4 Results

The results regarding the *first research question* show that the participating teachers do have some prior knowledge of task characteristics ($M = 3.38$), though only some of these characteristics match the categories of the system outlined above. Each participant listed between two and four characteristics. The four most common responses were the number of cognitive steps or procedures required (deductively assigned → mathematical modeling, 39 times), the complexity of tasks/texts (deductively assigned → processing mathematical texts, 37 times), curricular/thematic breadth (inductively developed, 25 times), and the openness of the task (inductively developed, 21 times) (cf. table 2). Overall, it seems noteworthy that three of the categories on which this study is based are barely known and were only mentioned by less than 10% of participants (argumentation, using mathematical representations, and working technically); this is surprising as these categories are also represented in the obligatory system put forward by the standards of education. However, some categories were developed or identified inductively as well. While these categories have been studied and discussed to some extent, they have merely been classified in terms of cognitive complexity but not yet operationalized with regards to generating cognitive requirement due to lacking evidence. This points to some conceptual ambiguity among the participants, especially when it comes to the two inductively developed categories, an openness of the task (cf. Greefrath, 2004) and manifestation in the curriculum (cf. Jordan et al., 2006).

Furthermore, the degree of familiarity with a task type, the use of supporting resources, the orientation of the lessons, and the time allowance were regarded as significant, even though all of these are not task properties *per se* but rather characteristics of the lessons and cannot be assigned to any of the categories (see table 2).

With regard to the *second research question,* the results show that the previously mentioned aspects of the tasks provided were only specifically manipulated in three cases. This is mainly due to the fact that changing the requirement level was not discussed with regard to all of the previously introduced characteristics. Explications about how to increase or decrease the difficulty of a task were limited to one characteristic in 29 out of 48 cases. Moreover, the suggestions not always matched the characteristics discussed beforehand, which strongly limits any statements about how these characteristics are actually employed. Overall, the teachers gave a fairly wide variety of different suggestions how to alter the difficulty of the two tasks (cf. Fig. 1). Suggestions for simplifying task 1 were primarily geared towards simplifying the required modeling by including parts of the area to be calculated in the text or drawing. Suggestions for increasing the difficulty of this task involved reducing the amount of available information ("don't show h in the drawing", "don't specify r", "rearrange the circles"), using different units of measurement, or moving the problem into a more realistic context ("photo of the real setup → justified estimations").

Category	**Number of items ($N = 162$)**	**Percentage with $N = 162$**	**Deductive assigned cat. (d)/ inductive developed cat. (i)**
Mathematical Modeling	39	24,07	d
Argumentation	11	6,79	d
Using math. Representations	8	4,93	d
Working technically	9	5,56	d
Processing mathematical texts	37	22,84	d
Openness of the Task	21	12,96	i
Manifestation in the Curricul.	25	15,43	i
Other Items	12	7,41	Without categorization
	Sum: 162 ($M = 3{,}38$)	Sum: 100	

Table 2. Results of the preliminary studies on difficulty characteristics of mathematical tasks.

In order to simplify task 2, participants put forward suggestions that, with the exception of excluding "special cases," focused almost exclusively on different modes of representation that would allow students to manipulate them more actively. These suggestions included a possible visual representation of the first four cases or listing them in a table showing the individual cases.

Overall, the results indicate that the participating teachers were only familiar with a limited set of categories for describing and analyzing the level of cognitive requirement in mathematical tasks at the beginning of the video-based training. They were then barely able to apply these categories to the tasks in a purposeful, differentiated, and flexible manner.

With regards to the *third research question*, the results of the follow-up study clearly indicate that participating teachers expanded their knowledge to include features which contribute to the cognitive requirement of mathematical tasks. First, this is evident because the categories that were at the focus of the video-based training could now be named more reliably (M =4.4). Second, categories that were previously described in primarily associative ways and which do not contribute to the cognitive requirement of a task (e.g., manifestation in the curriculum, degree of familiarity, time allowance, or use of resources) were only considered to be part of this construct in a few isolated cases. Regarding the question whether this knowledge can be applied in an effective and flexible manner, several tendencies can be identified. Comparing the results of the preliminary and concluding surveys indicates that the requirement categories can not only be employed with greater precision, but also in a more differentiated way, as most participants now identified the requirement levels of each category correctly – albeit implicitly – and then utilized them more effectively to alter the overall requirement level of a task as well as the required level in a specific category as needed. These observations can be made across the entire spectrum of the five categories but are particularly apparent with regards to the awareness for and utilization of argumentation, using mathematical representations, and working technically.

Thanks to the design of the study, these results can be carefully interpreted to mean that the teachers who participated in the training now possess a much broader range of categories for assessing and analyzing the level of cognitive requirement in mathematical tasks. Furthermore, they are able to implicitly identify the categorical requirement level and alter it as needed in most cases.

Those results confirm the study of Kleinknecht et al. The aim of their study was to develop a manual to analyse tasks. This manual focusses on an interdisciplinary design and reflection of tasks in every-day lessons. It aims to empirically proof the quality of measurement and to offer a training, in which teachers can learn to analyse and change tasks oriented to learning processes (cf. Kleinknecht et al. 2014, p.147). By using an evaluation in the beginning and at the end, they delivered the following results: some teachers only name a few criteria (pedagogical and cognitive characteristics) to assess tasks. They often named a high amount of pedagogical criteria, even though they were asked to name criteria to assess the difficulty of tasks. It was noticeable, so the authors that teachers put an emphasis on educational standards and more overtly on operators (cf. Kleinknecht et al. 2014, p. 156).

Other criteria being named were aimed at the amount/time to solve, prior knowledge, relations to the surrounding world, connection to other topics/subjects. Only half of the named criteria aiming at the cognitive potential of tasks (cf. Kleinknecht et al. 2014, p. 156). The authors put an emphasis on the necessity to apply the analysis of tasks within the planning and follow-up of lessons as a result, but they also see an intensive preparation to theuse of (pedagogical in general) criteria of analysis and a discussion of relevant content-pedagogical terms (cf. Kleinknecht et al. 2014, p. 156).

In order to professionalize teachers, it raises a challenge at analysing tasks to sensitize teachers to connect general pedagogical and content-pedagogical knowledge and to support them at applying this knowledge on the analysis of different sample task (cf. ibid., p. 157).

5 Discussion

The aim of this exploratory study was to examine the effects of a video-based teacher training ($N = 48$) on the participants' development of task competence regarding the cognitive requirement of mathematical tasks. The concept of cognitive requirement was defined by means of five cognitive categories: mathematical modeling, argumentation, use of mathematical representations, working technically, and processing mathematical texts. Compared to previous, large-scale empirical studies (Komorek, 2006; Jordan et al., 2008; Neubrand et al., 2013, Drüke-Noe, 2014), this study stands out because teachers' cognitive processes and their changes were investigated "directly", i.e. on level 2 according to Lipowsky's (2010) model.

The results of the study indicate that the teachers' prior knowledge about the range of requirement-generating categories falls short of expectations. They are initially only able to identify categorical requirement levels and apply them to tasks or the challenge of altering tasks in a few cases. Given this limited potential, it seems questionable whether and to what extent the teachers will be able to anchor, for instance, Winter's (1995) three basic experiences in their lessons by selecting, designing, and modifying their mathematical tasks accordingly. This reinforces existing assumptions from subject-independent research, which identify the low potential for cognitive activation in learning and internal assessment tasks (cf. Karpinski et al., 2013; Neubrand et al., 2013) by showing possible causes of these shortcomings. Furthermore, the aforementioned results seem to make Drüke-Noe's assumption (2014, p. 249) more plausible that tasks are only occasionally designed based on criteria beyond the level of content, since we have found that other systems of categories are not known to a sufficient degree *either*, first and foremost the mandatory *standards of education for mathematics* (cf. KMK, 2004).

First results of the four-stage training point towards a successful expansion of teachers' knowledge regarding requirement levels based on the system of categories. In particular, the results of the surveys indicate that significantly more task characteristics are taken into account, and that this is done in a more differentiated way within each category. This tendency towards a better conceptual grasp of "cognitive requirement" is also apparent in the participants' reduced use of task characteristics whose intensity has not been empirically linked to task difficulty so far. Therefore, we can assume that the training benefits an effective variation of the level of cognitive requirement for mathematical tasks, which in turn supports expectations of improved teaching practices (level 3) (cf. Lipowsky, 2010). Since this study also takes into account the structural feature of collective participation (cf. Garet et al., 2001), it can be further assumed that this potentially opens up room for discourse within a faculty, which can even the path towards category-based task design and lesson planning within the mathematical faculty. Such processes ultimately hold a potential for moving towards *individualized learning processes*, which better addresses the needs of individual students and their often unequal levels of ability, both on the individual and the classroom level.

However, it became apparent during the study that the narrow focus of the training and lack of expert support can lead to delayed or difficult progress for some participants, which can only partially be attributed to the increased study workload. Especially learning and applying the requirement levels gave rise to conceptual difficulties, which could not be compensated entirely by the intended collective participation. In terms of overall effectiveness, it would seem like a good idea to add on-site training phases with opportunities for professional communication to other video-based trainings with a similar narrow content focus (cf. Maldonado, 2002). One possible alternative method for effectively addressing these input difficulties might be the concept of an *inverted classroom*, where the on-site phases are used primarily for asking questions and having discussions (cf. Strayer, 2012).

The results of the present study cannot explain why teachers have such limited prior knowledge of requirement-generating categories and their requirement levels, as one should expect subject-specific category systems that have already been implemented to be sufficiently well known. But it shows, that the application of knowledge (even if it is low) about requirement-generating categories needs support and preparation in practice. Eventually, terms (especially those of pedagogy in general) have to be explained and translated to content-specific terminology, which can be a fundamental problem (cf. Kleinknecht et a. 2014, p. 156).

Another question that remains unanswered is to what extent the system of categories can be beneficial to teaching practices in the long run, and how it actually influences the design and reception of mathematical tasks. It is necessary to examine the question whether the used system of categories is selective in itself and its specifications critically. A task-based approach is problematic because one can conclude from cognitive potential of a task to the cognitive quality of doing tasks (cf. Leuders, Holzäpfel 2011, p. 214).

Lastly, the effectiveness of combining a video-based training with accompanying print materials, and to what extent this combination is necessary to rise a significant increase in the participating teachers' knowledge, remains unknown.

Teachers should be sensitized, in order to achieve teacher competencies and the competence to stimulate thinking of student on a high cognitive level on tasks. A joint analysis of sample tasks can be a good start for that.

References

Abraham, U., Müller, A.: 2009, Aus Leistungsaufgaben lernen, *Praxis Deutsch* **36(214)**, 4–12.

Artelt, C., Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., chiefele, U., Schneider, W., Schümer, G., Stanat, P., Tillmann, K.-J., Weiß, M. (Eds.): 2001, Pisa 2000. Zusammenfassung zentraler Befunde, Berlin: Max Planck-Institut für Bildungsforschung.

Ball, D. L., Cohen, D. K.: 2001, Making change. Instruction and its improvement, *Kappan* **83(1)**, 73–77.

Blomberg, G. et all: 2014, Understanding video as a tool for teacher education. Invastigating instructional strategies to promote reflection, *Instructional Science* **42 (3)**, 443–463.

Bloom, B. S.: 1994, Reflections on the development and use of the taxanomy, in: Rehage, K.J. et all, *Bloom's taxonomy. A forty-year retrospective*, Yearbook of the National Society of the Study of Education 93, Chicago.

Blum, W., Neubrand, M. (Eds.): 1998, TIMSS und der Mathematikunterricht. Informationen, Analysen, Konsequenzen Hannover: Schroedel.

Brouwer, N.: 2014, Was lernen Lehrpersonen durch die Arbeit mit Videos? Ergebnisse eines Dezenniums empirischer Forschung, *Beiträge zur Lehrerinnen – und Lehrerbildung* **32(2014) 2**, 176-195 – URN:urn:nbn:de:0111-pedocs-138644.

Büchter, A., Leuders, T.: 2005, Mathematikaufgaben selbst entwickeln: Lernen fördern – Leistung überprüfen Berlin: Cornelsen Scriptor.

Büchter, A., Pallack, A.: 2012, Zur impliziten Standardsetzung durch zentrale Prüfungen. Methodische Überlegungen und empirische Analysen, *Journal für Mathematik-Didaktik* **33**, 59–85.

Cohors-Fresenborg, E., Sjuts, J., Sommer, N.: 2004, Komplexität von Denkvorgängen und Formalisierung von Wissen, in: Neubrand,

M. (Eds..), *Mathematische Kompetenzen von Schülerinnen und Schülern in Deutschland: Vertiefende Analysen im Rahmen von PISA 2000*, 109-144, Wiesbaden: VS Verlag für Sozialwissenschaften.

Donabedian, A.: 1966, Evaluating the quality of medical care, *Milbank Q* **83(4) 2005**, 691–729.

Dreher, U., Holzäpfel, L., Leuders, T., Stahnke, R.: 2018, Problemlösen lehren lernen – Effekte einer Lehrerfortbildung auf die prozessbezogenen mathematischen Kompetenzen von Schülerinnen und Schülern, *Journal für Mathematikdidaktik* **39**, 227–256

Drüke-Noe, C.: 2014, Aufgabenkultur in Klassenarbeiten im Fach Mathematik. Empirische Untersuchungen in neunten und zehnten Klassen Wiesbaden: Springer Spektrum.

Drüke-Noe, C., Maier, U., Kleinknecht, M., Metz, K., Hoppe, H., Bohl, T.,: 2017, Lehrkräfte bei der Auswahl und Gestaltung von Aufgaben professionalisie- ren, in: Wernke, S., Zierer, K. (Eds.): 2017, *Die Unterrichtsplanung: ein in Vergessenheit geratener Kompetenzbereich?! Satus Quo und Perspektiven aus Sicht der empirischen Forschung* 208–224, Bad Heilbrunn: Verlag Julius Klinkhardt.

Garet, M., Porter, A. C., Desimone, L., Birman, B. F., Yoon, K. S.: 2001, What makes professional development effective? Results from a national sample of teachers, *American Educational Research Journal* **38(4)**, 915–945.

Greefrath, G.: 2004, Offene Aufgaben mit Realitätsbezug. Eine Übersicht mit Beispielen und erste Ergebnisse aus Fallstudien, *mathematica didactica* **27(2)**, 16–38.

Grossman, P. L., Schoenfeld, A.: 2005, Teaching subject matter, in: Darling-Hammond, L., Bransford, J., LePage, P., Hammerness, K., Duffy, H. (Eds.): 2005, *Preparing teachers for a changing world. What teachers should learn and be able to do*, 201–231, San Francisco: Jossey-Bass.

Guskey, T. R.: 2000, Evaluating professional development, Thousand Oaks: CA Corwin Press.

Hattie, J.: 2011, Visible Learning for Teachers. Maximizing impact on Learning, Abington: Routledge

Hiebert, J., Gallimore, R., Garnier, H., Givvin, K. B., Hollingsworth, H., Jacobs, J., Chui, A. M., Wearne, D., Smith, M., Kersting, N., Manaster, A., Tseng, E., Etterbeek, W., Manaster, C., Gonzales, P., Stigler, J. W.: 2003, Teaching mathematics in seven countries: results from the TIMSS 1999 Video Study (NCES 2003–013), US Department of Education, Washington, DC: National Centre for Education Statistics.

H o š p e s o v á, A., T i c h á, M.: 2010, Reflexion der Aufgabenbildung als Weg zur Erhöhung Der Lehrerprofessionalität, in: *Mathematik im Denken der Kinder. Anregungen zur mathematik-didaktischen Reflexion*, 122–126, Seelze: Klett/Kalmeyer.

H u g e n e r, I., P a u l i, C., R e u s s e r, K.: 2007, Inszenierungsmuster, kognitive Aktivierung und Leistung im Mathematikunterricht. Analysen aus der schweizerisch-deutschen Videostudie in: Lemmermöhle, D., Tothgangel, M., Bögeholz, S., Hasselhorn, M., Watermann R. (Eds.): 2007, *Professionell Lehren – Erfolgreich Lernen*, 109–212,

J o r d a n, A., R o s s, N., K r a u s s, S., B a u m e r t, J., B l u m, W., N e u b r a n d, M., L ö w e n, K., B r u n n e r, M., K u n t e r, M.: 2006, Klassifikationsschema für Mathematik-aufgaben: *Dokumentation der Aufgabenkategorisierung im COACTIV-Projekt* Berlin: MPI

K a r p i n s k i, M., G r u d n i e w s k a, M., Z a m b r o w s k a, M.: 2013, Nauczanie matematyki w gimnazjum- Raport z badania Warszawa: IBE

K l e i n k n e c h t, M., O t t i n g e r, S., R i c h t e r, D.: 2014, Aufgabenanalyse erlernen – Empirische Forschung zum Einsatz eines allgemeindidaktischen Kategoriensystems in der Lehrerfortbildung, in: Pfitzner, M. (Eds.), Aufgabenkultur im Sportunterricht, Bildung und Sport 5, 137–159, Wiesbaden: Springer.

K M K 2004, Beschlüsse der Kultusministerkonferenz. Bildungsstandards im Fach Mathematik für den Mittleren Schulabschluss. München: Wolters Kluwer.

K o m o r e k, E.: 2006, *Mit Hausaufgaben Problemlösen und eigenverantwortliches Lernen in der Sekundarstufe I fördern. Ein Ausbildungsprogramm für zukünftige Mathematiklehrer*, Berlin: Logos.

K u n t e r, M., D u b b e r k e, T., B a u m e r t, J., B l u m, W., B r u n n e r, M., J o r d a n, A., K l u s m a n n, U., K r a u s s, S., L ö w e n, K., N e u b r a n d, M., T s a i, Y.-M.: 2006, Mathematics instruction in the PISA 2004 classes: conditions, forms, and teaching and learning processes, in: PISA-Konsortium Deutschland (Eds.), PISA 2003. *Untersuchungen zur Kompetenzentwicklung im Verlauf eines Schuljahres*, 61–194, Münster: Waxmann.

K u n t e r, M., V o s s , T.: 2013, The Model of Instructional Quality in COACTIV. A Multicriterial Analysis. (Chapter 6) in: *Cognitive Activation in the Mathematics Classroom and Professional Competence of Teachers*, 97–124, Springer.

L e n n é, H.: 1969, Analyse der Mathematikdidaktik in Deutschland, Stuttgart: Klett.

L e u d e r s, T., H o l z ä p f e l, L.: 2011, Kognitive Aktivierung im Mathematikunterricht, *Unterrichtswissenschaft* **39(2011)**, 213–230.

L i p o w s k y, F.: 2010, Lernen im Beruf. Empirische Befunde zur Wirksamkeit von Lehrerfortbildungen, in: Müller, F.H., Eichenberger, A., Lüders, M., Mayr, J., (Eds.), *Lehrerinnen und Lehrer lernen. Konzepte und Befunde zur Lehrerfortbildung*, 51–70, Münster: Waxmann.

L i p o w s k y, F., R z e j a k, D.: 2012, Lehrerinnen und Lehrer als Lerner – Wann gelingt der Rollentausch? Merkmale und Wirkungen wirksamer Lehrerfortbildungen, *Reform der Lehrerbildung* **5(2012)3**, 1–17.

L u t h i g e r, H.: 2012, Lern – und Leistungsaufgaben in einem kompetenzorientierten Unterricht. *Haushalt in Bildung & Forschung* **1(3)**, 3–14.

M a i e r, U. et al.: 2010, Ein allgemeindidaktisches Kategoriensystem zur Analyse des kognitiven Potenzials von Aufgaben, *Beiträge zur Lehrerinnen – und Lehrerbildung* **28 (2010) 1**, 84-96 – URN: urn:nbn:de:01111-pedocs-137347.

M a l d o n a d o, L.: 2002, *Effective professional development. Findings from research*, New York: The College Board.

M e w b o r n, D.: 2003, Teachers, teaching, and their professional development, in: Kilpatrick, J., Martin, W.G., Schifter, D., (Eds.), *A research companion to principles and standards for school mathematics*, 45–52, Reston, VA: National Council of Teachers of Mathematics.

L i n d m e i e r, A.: 2011, *Modeling and measuring knowledge and competencies of teachers: A threefold domain-specific structure model for mathematics*, Munster: Waxmann.

M i t c h e l l, R. N., M a r i n, K. A.: 2015, Examining the use of a structured analysis framework to support prospective teacher noticing, *Journal of Mathematics Teacher Education* **18(6)**, 551–575.

M S W 2008, Zentrale Prüfungen am Ende der Klasse 10. Gymnasium Mathematik 2008. Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen.

N e u b r a n d, J.: 2002, Eine Klassifikation mathematischer Aufgaben zur Analyse von Unterrichtssituationen: *Selbsttätiges Arbeiten in Schülerarbeitsphasen in den Stunden der TIMSS-Video-Studie* Hildesheim: Verlag Franzbecker.

N e u b r a n d, M., K l i e m e, E., L ü d t k e, O., N e u b r a n d, J.: 2002, Kompetenzstufen und Schwierigkeitsmodelle für den PISA-Test zur mathematischen Grundbildung, *Unterrichtswissenschaft* **30(2002) 2**, 100–119.

N e u b r a n d, M., K u n t e r, M., B a u m e r t, J., B l u m, W., K l u s m a n, U., K r a u s s, S. (Eds.): 2013, Cognitive Activation in the Mathematics Classroom and Professional Competence of Teachers. Results from the COACTIV Project, *Mathematics Teacher Education* **8**, Springer US.

O E C D – Organisation for Economic Co-operation and Development: 1999, *Measuring student knowledge and skills* A new framework for assessment, Paris: OECD.

O s e r, F. K., B a e r e s w y l, F. J.: 2001, Choreographics of teaching. Bridging instruction to learning, in: Richardson, V. (ds.), *Handbook of Research of Teaching*, 4 ed., 1031–1065, Washington D.C.

R i e c k e-B a u l e c k e, T.: 2017, Unterrichtsqualität, in: Abshaben, M. et al (Eds.), *Basiswissen Lehrerbildung, Mathematik unterrichten*, 149–166, Seelze: Klett.

S c h e j a, B.: 2016, Cognitive activation through mathematics tasks in the context of centralised examinations on ISCED 2 Level, using the example of North-Rhine Westphalia, *Didacitica Mathematicae* **38**, 175–202

S c h e j a, B.: 2017a, Kognitive Aktivierung durch Mathematikaufgaben zentraler Abschlussprüfungen – Eine Vergleichsanalyse der polnischen Mittelschulprüfung und der Zentralen Prüfung aus Nordrhein-Westfalen, *Journal für Mathematikdidaktik* **38(2)** 291–322.

S c h e j a, B.: 2017b, The changing cognitive requirement of test tasks in mathematics – A longitudinal study of the Polish middle school examinations, *Didacitica Mathematicae* **39**, 175–204.

S h e r i n, M. G., v a n E s, E.A.: 2009, Effects of video club participation on teachers' professional vision, *Journal of Teacher Education* **60(1)**, 20–37.

S c h o e n f e l d, A. H.: 1998, Toward a theory of teaching-in-context, *Issues in Education* **4(1)**, 1–94.

S t a r, J. R. et all: 2011, Using video to improve preservice mathematics teachers' abilities to attend to classroom features. A replication study, in: Sherin, M.G. et all, *Mathematics teacher noticing. Seeing through teachers' eyes*, 117–133, London.

S t i g l e r, J., H i e b e r t, J.: 1999, The teaching gap, New York: Free Press.

S t i g l e r, J., W., G o n z a l e s, P., K a w a n a k a, T., K n o l l, S., S e r r a n o, A.: 1999, *The TIMSS Videotape Classroom Study.* Methods and Findings from an Exploratory Research Project on Eighth-Grade Mathematics Instruction in Germany, Japan, and the United States, Washington, DC: U.S. Department of Education, National Center for Education Statistics.

S t o c k e r o, S. L.: 2008, Using a video-based curriculum to develop a reflective stance in prospective mathematics teachers, *Journal of Mathematics Teacher Education* **11(5)**, 373–394.

S t r a y e r, J. F.: 2012, How learning in an inverted classroom influences cooperation, innovation and task orientation. 15:171–193 DOI 10.1007/s10984-012-9108-4

T e r h a r t, E., B a u m g a r t, F., M e d e r, N., v o n S y c h o w s k i, G.: 2009, Standardisierte Prüfungsverfahren in der Erziehungswissenschaft: Kontext, Formen, Konsequenzen, *Erziehungswissenschaft* **20**, 9–36.
The Didactical Contract: The Teacher, the Student and the Milieu 2002, in: Brousseau, G., Theory of Didactical Situations in Mathematics. *Mathematics Education Library*, **vol 19**, 226–249, Dordrecht: Springer.
T i m p e r l e y, H.: 2008, Teacher professional learning and development, Brussels, Belgium: International Academy of Education/International Bureau of Education.
V o n d r o v á, N.: 2016, The Effect of a Video-Based Intervention on the Knowledge-Based Reasoning of Future Mathematics Teachers, in: *Invited Lectures from the 13th International Congress on Mathematical Education*, 699–717, Hamburg: Springer.
W i n t e r, H.: 1995, Mathematikunterricht und Allgemeinbildung, *Mitteilungen der Gesellschaft für Didaktik der Mathematik* **61**, 37–46.
Y o o n, K., D u n c a n, T., L e e, S.W.-Y., S c a r l o s s, B., S h a p l e y, K.: 2007, Reviewing the evidence on how teacher professional development affects student achievement. *Issues & answers report*, **REL 2007 – no. 033** Washington, DC: U.S. Department of Education

Wykorzystanie nagrań wideo dla rozwijania kompetencji nauczycieli w zakresie kognitywnych wymogów zadań matematycznych

S t r e s z c z e n i e

Szerokie debaty w obszarze polityki edukacyjnej dotyczące złego funkcjonowania systemu szkolnego doprowadziły w latach 80-tych i 90-tych w wielu krajach do poważnych zmian systemowo-organizacyjnych. Głównymi elementami tego nowego podejścia było między innymi wprowadzenie nowych struktur w szkolnictwie, opracowanie nowych podstaw programowych dla nauczania różnych przedmiotów, wprowadzenie egzaminów centralnych. Innym istotnym elementem zmian było preferowanie nowej koncepcji uczenia się i nauczania, zorientowanej bardziej na założenia konstruktywizmu.

W literaturze dydaktycznej nie ma zbyt wielu publikacji badawczych związanych z praktyką planowania lekcji matematyki, których celem jest przebadanie jak nauczyciele podejmują decyzje dydaktyczne dotyczące kognitywnej aktywacji, jako podstawowego wymiaru mającego wpływ na jakość „nowego

nauczania" (Oser & Besriswyl, 2001; Kunter & Voss 2013). W tym artykule przedstawiamy w jaki sposób szkolenia wideo mogą być wykorzystywane dla podnoszenia jakości pracy nauczycieli w kierunku poszerzenia ich wiedzy odnośnie kognitywnych wymagań zadań matematycznych i sposobów korzystania z tej wiedzy w elastyczny sposób. Jest to istotne o ile może wykazać, czy i w jakim zakresie nauczyciele są w stanie skuteczniej radzić sobie z aktualnymi wyzwaniami podczas planowania lekcji. Takie wyzwania mogą pojawić się na przykład w trakcie prowadzenia zindywidualizowanego procesu uczenia i nauczania. Badanie zostało przeprowadzone na $N = 48$ nauczycielach ze szkół poziomu ISCED 2 z Nadrenii Północnej-Westfalii. W koncepcji szkolenia brano pod uwagę 5 kognitywnych kategorii: modelowanie matematyczne, argumentacja, wykorzystanie matematycznych reprezentacji, sprawność wykonywania działań matematycznych i posługiwanie się matematycznym tekstem. W ramach tych kategorii wyróżniono cztery poziomy kognitywnej kompleksowości (0–3).

W artykule zostały omówione wnioski z przeprowadzonych badań:

- Wstępne wyniki badania wskazują, że wiedza nauczycieli dotycząca kategorii generujących kognitywną kompleksowość jest niższa niż oczekiwano. Ponadto tylko niekiedy są oni w stanie zidentyfikować poziomy wymagań (związane z określonymi kategoriami) i zastosować je w zadaniach lub w trakcie modyfikacji zadań. Biorąc pod uwagę ten ograniczony potencjał, wydaje się wątpliwe, czy i w jakim stopniu nauczyciele są w stanie stosować trzy podstawowe elementy (wybieranie, projektowanie i modyfikowanie zadań matematycznych zgodnie z (Winter 1995)) na swoich lekcjach, wybierając odpowiednie zadania czy też modyfikując swoje zadania matematyczne. Te wyniki są zgodne z innymi badaniami (cf. Karpinski et al., 2013; Neubrand et al., 2013), które określają niski potencjał aktywacji kognitywnej lekcji matematyki, związany z doborem zadań wykorzystywanych na lekcjach.

- Pierwsze wyniki czterostopniowego szkolenia wskazują na znaczący rozwój wiedzy nauczycieli na temat kognitywnych wymagań opartych na wybranym systemie kategorii. W szczególności wyniki ankiet wskazują na to, że biorą oni pod uwagę znacznie więcej cech zadań i że odbywa się to w sposób bardziej zróżnicowany w ramach każdej kategorii. Ta tendencja do głębszego pojęciowego pojmowania „wymagań kognitywnych" jest również widoczna w węższym stosowaniu przez uczestników zadań o cechach, których intensywność nie była dotychczas empirycznie powiązana z trudnościami zadania. W związku z tym możemy założyć, że szkolenie przynosi wymierne zróżnicowanie poziomu wymagań kognitywnych dla zadań matematycznych, co z kolei wspiera oczekiwania związane z poprawą praktyki pedagogicznej.

Na koniec zostały omówione wnioski w kierunku optymalizacji szkolenia nauczycieli oraz wskazane zostały kierunki dalszych badań.

University of Cologne
Institut für Mathematikdidaktik
Gronewaldstraße 2
50931 Köln
Germany
e-mail: *schejab@uni-koeln.de*

Institut für Didaktik der Mathematik
Universität Bielefeld
Universitätsstraße 25
D-33615 Bielefeld
e-mail: *sabine.castelli@uni-bielefeld.de*

3 Zusammenfassung

Im Folgenden werden die Ergebnisse der zwei Studien herangezogen, um die zentralen Forschungsanliegen dieser Arbeit zu beleuchten. Hierzu werden zunächst die empirischen Ergebnisse in Bezug auf die jeweils zugrunde liegenden Kategoriensysteme aus qualitätsorientierter, vergleichsorientierter und entwicklungsorientierter Perspektive wiedergegeben (3.1) und anschließend vor dem Hintergrund einer normativen Perspektive diskutiert (3.2). Im abschließenden Abschnitt der Arbeit werden offene und weiterführende Fragen thematisiert (3.3).

3.1 Ergebnisse

Die folgende Darstellung der Ergebnisse erfolgt zunächst auf der Grundlage von Publikationen I – III, die in ihrer Zusammensetzung die *erste Studie* bilden, und unter den drei Forschungsperspektiven auf eine differenzierte Beantwortung des ersten zentralen Anliegens dieser Arbeit zielen (Forschungsgegenstand: zentrale Testaufgaben); aufgrund des im Bereich des Lehrerprofessionswissens verorteten zweiten Forschungsgegenstandes werden die Befunde der *zweiten Studie* (Publikation IV), in der die zweiten Forschungsfrage beleuchtet wird, davon getrennt diskutiert.

Erste Studie

Die zur *ersten Studie* gehörende *Publikation I* stellt die Befunde einer Analyse von Mathematikaufgaben der polnischen Mittelschulprüfung des Zeitraums 2002 – 2015 dar. Dabei wurde auf der Grundlage eines im Wesentlichen auf die COACTIV-Studie und dessen Erweiterungen zurückführbaren Kategoriensystems empirisch erfasst, welches kognitive Aktivierungspotential diese Testaufgaben aufweisen (vgl. Scheja 2017a).

Aus der Perspektive der rationalen Aufgabenanalyse zeigt sich, dass die polnischen Mittelschulprüfungs-Aufgaben dieses Zeitraums sowohl im Vergleich zu PISA- als auch zu den deutschen Unterrichts-Aufgaben ein verhältnismäßig hohes kognitives Aktivierungspotential aufweisen (vgl. Drüke-Noe 2014; Neubrand et al. 2011). So wird der Typ mathematischen Arbeitens in den Testaufgaben annähernd ausschließlich durch begriffliche und vor allem rechnerische Modellierungen bestimmt. Das Vorkommen der beiden Aufgabenklassen und die allenfalls punktuell berücksichtigten technischen Aufgaben gehen folglich mit einer großen Spannweite mathematischen Denkens in den Testaufgaben einher, die in den Lösungsprozessen verlangt wird (Aktivierungsbreite). Dabei kommt den erfassten mathematischen Tätigkeiten außer- und innermathematisches Modellieren, Gebrauch mathematischer Darstellungen, technisches Arbeiten und dem Umgang mit mathematischen Texten zum Teil eine hohe Bedeutung zu. Eine differenzierte Betrachtung zeigt dabei, dass insbesondere vergleichsweise anspruchsvolle mathematische Texte und Darstellungen sowie das technische Arbeiten, das in anspruchsvollen, da zumeist mehrschrittigen Modellierungsprozessen eingebettet ist, das Anspruchsniveau der polnischen Testaufgaben prägt, wohingegen das Argumentieren eher in geringerem Maße

dazu beiträgt.
Dass die Befunde zu den polnischen Mittelschulprüfungsaufgaben aus qualitätsorientierter Perspektive nicht international verallgemeinerbar sind und eher mit Einschränkungen in der Konzeption von Testaufgaben anderer zentraler Abschlussprüfungen vorfindbar sind, wird aus dem Vergleich der Aufgaben der polnischen Mittelschulprüfung und der nordrhein-westfälischen ZP 10 deutlich. So zeigen die Ergebnisse von *Publikation II* (vgl. Scheja 2016), dass die Mathematikaufgaben der ZP 10 im Zeitraum zwischen 2007 und 2016 ein vergleichsweise hohes kognitives Aktivierungspotential aufweisen (vgl. Drüke-Noe 2014; Neubrand et al. 2011), welches jedoch eher Schwerpunkte im Spektrum der Anforderungen aufweist. Kennzeichnend ist hierbei zunächst, dass der Typ des mathematischen Arbeitens, der im Lösungsprozess der Testaufgaben verlangt wird, überwiegend als Modellierungsprozess charakterisierbar ist. Vor allem das prozedurale und – mit erheblichen Einschränkungen auch – das begriffliche Denken findet hier als „Werkzeug" (Neubrand et al. 2002, S. 102) in inner- und vor allem vielfältigen außermathematischen Kontexten Anwendung. Die nachweislich den deutschen Unterricht dominierende Aufgabenklasse der kontextlosen technischen Aufgaben (vgl. Drüke-Noe 2014, S. 107) wird folglich im gesamten Untersuchungszeitraum der ZP 10 nur marginal und somit entgegen der konsentierten Forderungen (vgl. KMK 2003) eher unausgewogen berücksichtigt.
Ein gemeinsames Merkmal des polnischen und des nordrhein-westfälischen Aufgabensatzes ist, dass sich die jeweiligen Testaufgaben vergleichsweise ausgewogen auf die mathematischen Tätigkeiten und mit erheblichen Einschränkungen ihre Niveaus verteilen (vgl. Drüke-Noe 2014, S. 160 f.; Neubrand et al. 2011). Damit sind die Ergebnisse hinsichtlich des kognitiven Aufgabenanspruchs des ZP 10-Aufgabensatzes anschlussfähig an die wenigen bislang vorliegenden Forschungsergebnisse zu innerdeutschen MSA-Prüfungen (vgl. Neubrand & Neubrand 2010; Drüke-Noe 2014, S. 160 ff.). Wesentliche Unterschiede zeigen sich zunächst in der Berücksichtigung der Tätigkeiten des innermathematischen und des außermathematischen Modellierens; hier greifen die nordrhein-westfälischen Aufgaben häufiger außermathematische Kontexte auf, wohingegen die Aufgaben der polnischen Mittelschulprüfung eher ausgewogener über die beiden Tätigkeiten verteilt sind. Lässt man in dieser Betrachtung die beiderseits geringfügig berücksichtigte Aufgabenklasse der technischen Aufgaben einfließen, so können die beiden Kontexte tendenziell gar insgesamt als profilprägend angesehen werden. Und dennoch: Eine differenzierende Betrachtung hinsichtlich Aufgabenverteilung auf die drei Anspruchsniveaus zeigt zugleich, dass das kognitive Anspruchsniveau dieser beiden Tätigkeiten bei der ZP 10 niedriger als bei der polnischen Mittelschulprüfung ausfällt, was unter anderem darauf zurückzuführen ist, dass hier kognitiv anspruchsvolle „Modellreflexionen, -validierungen oder -eigenentwicklungen" allenfalls punktuell verlangt werden. Zu den Gemeinsamkeiten der beiden Aufgabensätze zählt einerseits ein vergleichsweise geringer Beitrag mathematischer Argumentationen zum kognitiven

Anspruchsniveau und andererseits ein relativ hoher kognitiver Anspruch des technischen Arbeitens und des Umgangs mit mathematischen Darstellungen bzw. Texten; dass der Befund zum Argumentieren gegenüber anderen Tätigkeiten dennoch relativiert werden kann und auch sollte, zeigt beispielsweise ein Vergleich mit dem als hoch eingeschätzten Kategorieergebnis von PISA (vgl. Neubrand et al. 2011, S. 128).

Die *dritte Publikation* stellt die Befunde einer entwicklungsorientierten Analyse von Testaufgaben der polnischen Mittelschulprüfung *innerhalb* des Zeitraums von 2002 bis 2015 dar (s. Scheja 2017a). Um mögliche Veränderungstendenzen hinsichtlich des kognitiven Anspruchsniveaus in diesem Zeitverlauf aufzuzeigen, werden dabei jeweils vier Jahre umfassende Zeiträume, der der Einführung der Mittelschulprüfung (2002 – 2005) und der nach ihrer organisatorisch-konzeptionellen Reform im Schuljahr 2011/12 (2012 – 2015), betrachtet (vgl. Scheja 2017b). Auf der Grundlage einer derart angelegten längsschnittlichen Untersuchungen der Verteilungen über die Anspruchsniveaus mathematischer Tätigkeiten konnte – die obigen Erkenntnisse erweiternd – aufgezeigt werden, dass und inwiefern die Ergebnisse von Scheja (2017a) hinsichtlich ihres Deutungspotentials als „diffus“ (ebd., S. 317) zu charakterisieren sind. Dabei zeigte sich zunächst, dass die Aufgaben beider Untersuchungszeiträume ein relativ hohes kognitives Anspruchsniveau aufweisen, das über eine vergleichsweise breite Spanne an eingeforderten Tätigkeiten wie auch zum Teil recht hohe mittlere Anspruchsniveaus dieser generiert wird. Mit diesen Merkmalen von Mathematikaufgaben zentraler (Abschluss-)Prüfungen sind die vorliegenden Ergebnisse anschlussfähig an bereits vorliegende Befunde aus der fachdidaktischen Forschung (Drüke-Noe 2014; Scheja 2016; Scheja 2017a). In der Detailanalyse wird jedoch deutlich, dass die im Schuljahr 2011/12 im Zuge der Mittelschulprüfungs-Reform vollzogene Abspaltung des Mathematiktests aus dem bis dahin fächerverbindenden mathematisch-naturwissenschaftlichen Prüfungsteil (bis 2010/11 abgeprüfte Fächer dieses Prüfungsteils: Mathematik, Biologie, Geografie, Physik, Astronomie und Chemie) mit grundlegenden konzeptionellen Veränderungen der polnischen Testaufgaben einherging. Wesentliche Unterschiede zeigen sich *zum einen* in der Berücksichtigung der Tätigkeiten des innermathematischen und des außermathematischen Modellierens (innermathematisch: vgl. 2.3, Tab. 3, Aufgabe 4; Abb. 2, Aufgabe 1; außermathematisch: vgl. 2.3, Tab. 3, Aufgaben 1 – 3; Abb. 2, Aufgabe 2 – 4); hier greifen die Aufgaben des Einführungszeitraums der polnischen Mittelschulprüfung mit deutlichem Schwerpunkt außermathematische Kontexte auf, wohingegen die Aufgaben nach der Reform ausgewogen über die beiden Tätigkeiten verteilt sind. Lässt man in dieser Betrachtung die beiderseits punktuell berücksichtigte Aufgabenklasse der technischen Aufgaben einfließen, so können die derart gewichteten kontextuellen Einbettungen gar insgesamt als profilprägend angesehen werden, womit die vorfindbaren Profile die Prüfungsvorgaben beider Untersuchungszeiträume zugleich angemessen abzubilden scheinen. Die derart veränderte kontextuelle Schwerpunktverlagerung der polnischen Mittelschulprüfungsaufgaben nach der Reform, wie auch die aktuellen längsschnittlichen

Entwicklungsbefunde der deutschen zentralen Abschlussprüfungen bestätigen zugleich die von Scheja (2017a, S. 317) bisher für diesen Zeitraum vermutete kontextuelle Divergenz beider Testaufgabenprofile empirisch. So zeigt die Studie von Kühn und Drüke-Noe, dass deutsche MSA-Aufgaben flächendeckend eine „erhebliche Konstanz in der Gestaltung” (2013, S. 925) aufweisen, was folglich auf ein langfristig konstantes Maß an vorwiegend außermathematischen Kontexten schließen läßt. Die Konzeption der polnischen Mittelschulprüfungsaufgaben hat dagegen nach der Reform offenbar eine substantielle Verschiebung von einer überwiegenden Anwendungs- in Richtung der Strukturorientierung im Profil der kontextuellen Ansprüche erfahren.

Ein weiterer Unterschied der untersuchten polnischen Aufgabensätze zeigt sich *zum anderen* im Beitrag des mathematischen Argumentierens zur Anspruchsbildung, denn: Entgegen der auf den Prüfungsvorgaben gründenden Erwartung, die für beide Zeiträume argumentative Fähigkeiten besonders herausstellen, wird diese erst seit der Reform nennenswert verlangt und wirkt damit nur hier anspruchsgenerierend. Im Gegensatz zu den bisher betrachteten Tätigkeiten verlangen das technische Arbeiten und der Umgang mit mathematischen Darstellungen bzw. Texten innerhalb der verglichenen Zeiträume ein vergleichbares Anspruchsniveau.

Damit wird in den polnischen Testaufgaben nach der Prüfungsreform insgesamt ein breiteres Spektrum mathematischer Tätigkeiten eingefordert, die innerhalb der beiden Vergleichszeiträume zum Anspruchsniveau der polnischen Mittelschulprüfung beitragen. Gerade der fehlende Schwerpunkt innerhalb der beiden Kontexte (inner- bzw. außermathematisch) weist dabei deutlich auf eine bildungspolitische Abkehr von dem Anforderungsprofil des Einführungszeitraums hin, das dem der internationalen PISA-Aufgaben und somit auch dem Literacy-Konzept in auffällig hohem Maße ähnelt (vgl. Neubrand et al. 2002; Neubrand et al. 2011). Vor diesem Erkenntnishintergrund scheint es durchaus plausibel, die Ursachen der auffälligen PISA-Leistungszunahme polnischer Schülerinnen und Schüler im Zeitraum zwischen 2000 und 2012 beispielsweise auf dem Feld einer sich reformbedingt gleichgerichtet wandelnden Unterrichtskultur zu verorten, zu der unter anderem die Unterrichtssteuerung mittels offenbar PISA-naher Test- wie auch Beispielaufgaben zählt. Während erste Studien bereits punktuelle Hinweise auf einen sich derart wandelnden Mathematikunterricht und seine Rahmenbedingungen liefern (vgl. Karpinski et al. 2013; Maciejak et al. 2012; Konarzewski 2008; Pieronkiewicz 2017), konnte im Rahmen der ersten Studie (vgl. Scheja 2017a, S. 306) eine „Top-down-Aufgabensteuerung“ mittels frei verfügbarer Beispielaufgaben allenfalls als ein nicht intendierter Effekt identifiziert werden. So zeigte eine gegenüberstellende Analyse der gesamten polnischen Test- und Beispielaufgaben, dass eine Duplizierung in (jeweils) nur einer Aufgabenstellung vorfindbar ist. Aufgrund des hohen Maßes an verfahrensbedingter Unabhängigkeit in beiden Aufgabensätzen kann damit den Beispielaufgaben keine *intendierte* Steuerung seitens der für die Aufgabenkonzeption verantwortlichen Zentralen Prüfungskommission (CKE) auf den Unterricht zugeschrieben werden. Inwiefern jedoch

Test- und Beispielaufgaben der Mittelschulprüfung *tatsächlich* auf die Unterrichtsgestaltung Einfluss nehmen, bleibt weiterhin offen.
Innerhalb der eingeforderten Breite der mathematischen Tätigkeiten wird mit Ausnahme des außermathematischen Modellierens zudem ein höheres mittleres Anspruchsniveau aller mathematischen Tätigkeiten eingefordert, womit die polnischen Testaufgaben nach 2012 insgesamt ein höheres kognitives Anspruchsniveau aufweisen. Kritisch ist bei der festgestellten Entwicklung der polnischen Testaufgaben jedoch anzumerken, dass die mathematischen Tätigkeiten nun – wie in der Literatur gefordert (vgl. Winter 1995) – vielseitig abgebildet werden, die Anforderungsniveaus des Argumentierens, des außermathematischen und des innermathematischen Modellierens und damit hinsichtlich zentraler mathematischer Kompetenzen jedoch tendentiell unzureichend über deren gesamtes Anforderungsspektrum streuen. Dies zeigt sich in erster Linie daran, dass hier der Anteil der Aufgaben, die dem einfachen bzw. hohen Anforderungsniveau zuzuordenen sind, im Vergleich zum mittleren Niveau unverhältnismäßig klein ausfällt. Insofern stellt sich vor dem Hintergrund dessen, dass auch kalkülorientierte technische Aufgaben praktisch unberücksichtigt bleiben, die Frage, inwieweit die reformierte Mittelschulprüfung vor allem bei Lernenden unteren Leistungsbereichs seiner primären Aufgabe – der Überprüfung von Fähigkeiten und Kenntnissen im Bezug auf die Vorgaben des Rahmenplans (vgl. CKE 2010) – in abbildender Weise nachkommt.

Zweite Studie

Die in *Publikation IV* beschriebene *zweite Studie* zielt auf die Beleuchtung des zweiten Anliegens der vorliegenden Dissertation (vgl. Scheja & Castelli 2018). Sie stellt die Befunde zur Auswirkungen einer videobasierten Fortbildung auf die Entwicklung der Aufgabenkompetenz bzgl. des kognitiven Anspruchs von Mathematikaufgaben bei Lehrkräften dar (Durchführung Februar 2017 – Juli 2017). Ähnlich wie bereits in der ersten Studie waren dabei für das Konstrukt des kognitiven Anspruchs kognitive Kategorien aussagekräftig, die auf die Bildungsstandards der KMK und vor allem auf das COACTIV-Projekt zurückführbar sind; hierzu zählten: mathematisches Modellieren, Argumentieren, Umgang mit mathematischen Darstellungen, technisches Arbeiten und Umgang mit mathematischen Texten. Im Vergleich zu angrenzenden, großflächigen empirischen Untersuchungen (Jordan et al. 2008; Neubrand et al. 2013; Drüke-Noe 2014) besteht die Besonderheit der Studie darin, dass die Lehrerkognitionen und ihre Veränderungen im Sinne des Ebenenmodells nach Lipowsky (2010) “unmittelbar” (Ebene 2) untersucht wurden. Die Ergebnisse der Eingangsbefragung der explorativen Studie deuten zunächst darauf hin, dass die Lehrkräfte zum einen hinsichtlich der kategorialen Bandbreite nicht im erwarteten Maße über die Kenntnis anspruchsgenerierender Kategorien verfügen. Zum anderen sind sie allenfalls punktuell in der Lage, kategoriale Anspruchsniveaus zu identifizieren und differenziert auf Aufgaben bzw. deren Veränderungen anzuwenden. Dies ist tendenziell anschlussfähig an bereits vorliegende Befunde aus der fachdidaktischen

Forschung, in denen wiederholt aufgezeigt wurde, dass Lehrkräfte eher in geringem Maße in der Lage sind, Aufgabenpotentiale und somit auch Nutzungsmöglichkeiten von Aufgaben zu erkennen (vgl. Drüke-Noe et al. 2017; Hillje 2012; Karpinski et al. 2013), um so *gezielt* verständinsvolle Lerprozesse anzuregen; das Ergebnis stützt zugleich die Deutungen von Karpinski et al. (2013) sowie von Drüke-Noe, dass Lehrkräften „möglicherweise Aufgabenmerkmale jenseits inhaltlicher tatsächlich nur wenig bewusst sind” (2014, S. 252). Die Ergebnisse der Gegenüberstellung der Vor- und Nachuntersuchung (Eingangsbefragung: Februar 2017; Ausgangsbefragung: Juli 2017) deuten darauf hin, dass das kategoriale Anspruchswissen der Lehrkräfte auf der Grundlage der fünf zugrundegelegten kognitiven Kategorien mittels der videobasierten Fortbildung (April bis Juni 2017) erweitert werden kann. Hier weisen die Ergebnisse der abschließenden Fragebogenerhebung in die Richtung, dass nun erheblich mehr anspruchsgenerierende Kategorien beachtet werden, wobei dies zugleich innerhalb der einzelnen Kategorien deutlich differenzierter geschieht. Diese Tendenz zur Exaktifizierung des Konstruktes „kognitiver Anspruch” zeigt sich bei den teilnehmenden Lehrkräften zugleich im Rückgang von Aufgabenmerkmalen, wie beispielsweise Bearbeitungszeit, Werkzeugnutzung oder auch Maß der Einübung, deren Intensität bisher keine empirischen Hinweise auf die Aufgabenschwierigkeit liefert. Damit kann davon ausgegangen werden, dass die durchgeführte Fortbildung eine zielgerichtete Variation des kognitiven Anspruchsniveaus von Mathematikaufgaben begünstigt und somit auch positive Wirkungen auf der Ebene des *unterrichtspraktischen Handelns* (Ebene 3) erwartet werden können (vgl. Lipowsky 2010).

3.2 Diskussion

In diesem Abschnitt werden die empirischen Ergebnisse der beiden Studien nacheinander vor dem Hintergrund der zentralen Unterrichtsziele wie auch der jeweils gültigen curricularen Unterrichtsvorgaben diskutiert. Um dies *auch* aus einer einheitlichen normativen Perspektive leisten zu können, wurde im Rahmen von Publikation I zunächst aufgezeigt (vgl. Scheja 2017a), dass die prozessartigen Ziele des polnischen Rahmenplans und die allgemeinbildenden Kompetenzen der KMK-Bildungsstandards eine hohe Übereinstimmung aufweisen und die prozessartigen Ziele einen gleichverteilten Niederschlag in den drei Winter‘schen Grunderfahrungen finden; damit können sowohl die empirischen Ergebnisse der polnischen Mittelschulprüfung als auch die der ZP 10 vor dem Hintergrund des normativen Anspruchs der Winter‘schen Grunderfahrungen diskutiert werden.

Erste Studie

Die Verteilung der Tätigkeiten und ihrer Niveaus innerhalb des vollständigen Aufgabensatzes der polnischen Mittelschulprüfung (2002 – 2015) zeigt, dass die Konzeption nicht alle drei Winter‘schen Grunderfahrungen – Anwendungsorientierung, Strukturorientierung und Problemorientierung – umfassend verankert. Ursachen dessen sind im Wesentlichen auf zwei Problemfeldern verortet: So würde man *einerseits* neben

substantiellen Anteilen des technischen Arbeitens und des Umgehens mit mathematischen Darstellungen auch höhere Anteile des innermathematischen Modellierens erwarten. Damit scheint folglich vor allem die Anwendungsorientierung gegenüber der Strukturorientierung einen höheren Beitrag bei der Erfassung der mathematischen Allgemeinbildung zu leisten. Ein zu geringer Beitrag des mathematischen Argumentierens in den Lösungsprozessen zeigt eine dementsprechend unausgewogene Verteilung auch für die Grunderfahrung der Problemorientierung. Wie aus den Analysen deutlich wird, ist eine derart unausgewogene Berücksichtigung der Grunderfahrungen jedoch weniger auf eine eingeschränkte Konstruktvalidität der Testaufgaben zurückführbar, sondern eher ein starker Effekt einer in Zeitraum *vor der Reform* unzureichenden Korrespondenz zwischen dem übergeordneten Rahmenplan und den konzeptionell zugrundeliegenden Prüfungsstandards. So scheinen in der vorfindbaren Aufgabenkonstruktion die Prüfungsstandards, in denen annähernd ausschließlich anwendungs- und auch problemorientiertes Wissen und damit zentrale Merkmale des Literacy-Konzeptes als verbindlich aufgeführt werden, angemessen operationalisiert worden zu sein; des Weiteren wirkt gerade dieser enge Zusammenhang zwischen Prüfungsvorgaben und Testaufgaben einschränkend hinsichtlich der Abbildung der drei Grunderfahrungen in den polnischen Testaufgaben für den gesamten Untersuchungszeitraum.
Dieses Defizit scheint mit der Einbettung der Prüfungsvorgaben in den überarbeiteten Rahmenplan im Zuge der Reform erkannt und in hohem Maße behoben. So wird ab 2012 – entsprechend der prozessartigen Zielvorgaben wie auch der damit korrespondierenden Winter‘schen Grunderfahrungen – in den Testaufgaben insgesamt ein breiteres Spektrum mathematischer Tätigkeiten eingefordert, die insgesamt zum Anspruchsniveau der Testaufgaben beitragen. Die im Vergleich zum Einführungszeitraum wie auch zu den KMK-Bildungsstandards (vgl. Kühn & Drüke-Noe 2013) hohe *Steuerungswirksamkeit* des neuen Rahmenplans zeigt sich einerseits im fehlenden Schwerpunkt innerhalb der beiden Kontexte (inner- bzw. außermathematisch) und andererseits in einer deutlich stärkeren Berücksichtigung des mathematischen Argumentierens in den polnischen Mittelschulprüfungsaufgaben. Damit scheinen seit der Reform sowohl die Struktur- als auch die Problemorientierung einen zum Teil gestiegenen Beitrag zur Erfassung der Allgemeinbildung in den polnischen Testaufgaben zu leisten. Insgesamt weisen damit die Testaufgaben nach der Reform der Mittelschulprüfung offenbar ein höheres diagnostisches Potential auf als dies während des Einführungszeitraums der Fall war.
Die Analyseergebnisse der ZP 10-Aufgaben zeigen eher eine eingeschränkte Passung mit den Vorgaben der Bildungsstandards für das Fach Mathematik. Die in den Bildungsstandards explizit geforderte Umsetzung der drei Winter‘schen Grunderfahrungen (KMK 2003, S. 6), welche sich in Bezug auf Testaufgaben in erster Linie in einer hinreichenden Aktivierungsbreite niederschlagen sollte, kann empirisch nur eingeschränkt bestätigt werden. Dies zeigt sich zunächst daran, dass substantielle Anteile des innermathematischen Modellierens und in Teilen auch des mathematischen Argumentierens

fehlen. Damit scheinen die Grunderfahrungen der Strukturorientierung und teilweise auch der Problemorientierung (und die darunter subsumierten mathematischen Kompetenzen) gegenüber der Anwendungsorientierung offenbar in zu geringem Umfang verankert, womit das Ergebnis anschlussfähig an bereits vorliegende innerdeutsche Befunde aus der fachdidaktischen Implementationsforschung ist (vgl. Drüke-Noe & Kühn 2013; Neubrand & Neubrand 2010). Mit dieser Schwerpunktsetzung innerhalb der verankerten kognitiven Anforderungen weisen die Testaufgaben der ZP 10 – ähnlich wie die Testaufgaben des Einführungszeitraums der polnischen Mittelschulprüfung – tendenziell Merkmale von PISA-Aufgaben und somit offenbar auch des Literacy-Konzepts auf (vgl. Neubrand et al. 2002; Neubrand et al. 2011); eine wünschenswerte Orientierung an den Erfordernissen eines umfassenderen, systematischen Wissensaufbaus im Sinne *intelligenten Wissens*, das von Weinert „als wichtigstes Bildungsziel" (2000, S. 5) genannt wird und von Neubrand und Neubrand (2010, S. 181) als Vorbedingung und zugleich als Ziel von Literacy betrachtet wird, ist in den untersuchten Aufgaben der ZP 10 scheinbar eingeschränkt abgebildet.

Zweite Studie

Die Ergebnisse der zweiten Studie, die jedoch aufgrund der Anlage kaum verallgemeinerbar sind, weisen zunächst in die Richtung, dass offenbar nicht alle Lehrkräfte im erwarteten Maße über kognitive Kategorien als flexible Beschreibungs- und Analysewerkzeuge von Mathematikaufgaben verfügen. Dies zeigt sich in der Kenntnis einer eher geringen kategorialen Breite hinsichtlich des kognitiven Anspruchs und lässt insofern fraglich erscheinen, ob alle der in den KMK-Bildungsstandards geforderten drei Winter'schen Grunderfahrungen hinreichend – da offenbar selten gezielt – in den eingesetzten Aufgaben verankert werden können. Dies deckt sich nicht mit breit konsentierten Vorgaben an einen allgemeinbildenden Mathematikunterricht und stützt bereits vorliegende Annahmen aus der grenzübergreifenden fachdidaktischen Forschung, die den Lern- wie auch schulinternen Testaufgaben anregungsarme Charakteristika (kaum streuender kategorieller Beitag zum Anspruchsniveau) zuweisen (vgl. Drüke-Noe 2014; Karpinski et al. 2013; Neubrand et al. 2011), indem es auf mögliche Ursachen dessen hindeutet. Darüber hinaus scheinen die obigen Ergebnisse Drüke-Noes Vermutung (2014, S. 249), dass jenseits inhaltlicher Aufgabenmerkmale, eine kriterial geleitete Konzeption von Aufgaben nur punktuell erfolgt, zu plausibilisieren, da sie zugleich *auch* auf eine unzureichende Kenntnis anderer Kategoriensysteme hindeuten, zu denen in erster Linie das der verbindlich eingeführten Bildungsstandards im Fach Mathematik zählt.

Die spezifischere und zugleich differenziertere Kenntnis und Anwendung anspruchsgenerierender Kategorien im Anschluss an die videobasierte Fortbildung deutet darauf hin, dass es den teilnehmenden Lehrkräften gelingen kann, gezielter mit dem Konstrukt des kognitiven Anspruchs umzugehen. Dies lässt einen positiven Einfluss auf die „Qualität der professionellen Wahrnehmung des Aufgabenpotentials" (Hammer 2016, S. 190) annehmen, die im engen Zusammenhang mit einer lernwirksamen Vorbereitung der

Aufgabenimplementation (ebd.), wie auch der Unterrichtsplanung insgesamt, steht. Vor diesem Hintergrund kann erwartet werden, dass Lehrkräfte im Zuge ihrer Reihenplanung, -durchführung und -reflexion fundierter Fragefelder innerhalb des Professionswissens in den Fokus nehmen können, die vom Autor als eine wichtige Voraussetzungen für eine umfassende Verankerung der drei Grunderfahrungen betrachtet werden. Hierzu zählen beispielsweise:

- Gibt es kategoriale Schwerpunkte im vorbereiteten Aufgabenpool und falls ja, sind diese curricular intendiert (kategoriale Breite)?
- Werden innerhalb der Kategorien unterschiedliche kognitive Anspruchsniveaus verlangt (kategoriale Tiefe)?
- Welche Förderschwerpunkte können anhand des erhaltenen formativen und summativen Feedbacks auf der Meso- bzw. Mikroebene identifiziert werden (formativ-diagnostische Bedarfsorientierung)?

3.3 Ausblick

Auf der Grundlage der Erkenntnisse der beiden vorangegangenen Abschnitte werden nun vor dem Hintergrund einer normativen Perspektive Vorschläge für eine Weiterentwicklung des Umgangs mit dem kognitiven Anspruch der hier betrachteten Testaufgaben (*Erste Studie*) wie auch des Umgangs von Lehrkräften mit ebendiesem Konstrukt in Lern- wie auch internen Testaufgaben (*Zweite Studie*) formuliert. Um in der ersten Studie das Deutungspotential der Befunde aufgrund der curricular-organisatorischen Reform der polnischen Mittelschulprüfung im Schuljahr 2011/12 zu erhalten, werden die Jahre des Einführungszeitraums der polnischen Mittelschulprüfung (2002 – 2005) und des Zeitraums nach ihrer Reform (2012 – 2015) getrennt diskutiert.

Erste Studie

Gerade weil empirische Befunde aufzeigen, dass a) Testaufgaben zentraler (Abschluss-)Prüfungen zunehmend unterrichtssteuernd wirken (vgl. Maier et al. 2011) und b) dass kognitiv anspruchsvolle Aufgaben die Leistungsentwicklung der Lernenden im Fach Mathematik positiv beeinflussen (vgl. Baumert & Kunter 2011; Stein & Lane 1996), ist es für die mit der Aufgabenkonzeption betrauten Institute (CKE und das QUA-LiS) ratsam, eine hohe Abbildungsgüte der curricularen Vorgaben durch die Testaufgaben anzustreben. Aus der Perspektive der zugrundeliegenden Prüfungsstandards und des Rahmenplans, die für den Einführungszeitraum der polnischen Mittelschulprüfung als verbindlich betrachtet werden, lassen sich entsprechende Angleichungsvorschläge formulieren. Hier ist es in erster Linie die Problemlösefähigkeit in Form einer „Reflexion auf das eigene Denkhandeln“ (Winter 1995, S. 41), die in den polnischen Testaufgaben stark unterrepräsentiert vertreten scheint. Gerade also im Hinblick auf das mathematische Argumentieren, das nach Meyer (vgl. 2007) zum Ausgangspunkt und Katalysator tieferen Mathematikverständnisses werden kann, weisen die Testaufgaben des Einführungszeitraums der polnischen Mittelschulprüfung Entwicklungspotentiale auf. Die Berücksichtigung dieser Tätigkeit innerhalb der Anforderungen würde zugleich eine größere Facettierung im Beitrag zum kognitiven Anspruch der polnischen Testaufgaben nach sich ziehen. Andererseits wird deutlich: Da die Prüfungsstandards wie auch die Bildungspolitik „die Kenntnis und den Umgang mit Algorithmen in ihrer typischen Anwendung“ (CKE 2010, S. 61) für diesen Zeitraum als leitendes Ziel und zugleich Anforderungsschwerpunkt der Testaufgaben festlegen, ist folglich eine Erhöhung der festgestellten Anteile des innermathematischen Modellierens gegenüber dem außer-mathematischen Modellieren offenbar nicht erstrebenswert, da nicht intendiert. Bildungspolitisch wünschenswert scheint im Zeitraum zwischen 2002 und 2011 vielmehr die Verankerung einer funktionalen Anwendung mathematischer Kenntnisse in verschiedenen Kontexten, die sich auffallend stark am Literacy-Konzept orientieren (vgl. Neubrand et al. 2002); ein Vergleich der mittleren Anforderungsniveaus in den hier untersuchten Kategorien zwischen den Testaufgaben der polnischen Mittelschulprüfung und Testaufgaben von PISA 2003 bestätigt, dass dieses in vergleichbarer Weise umgesetzt wurde.

Die mit der Reform der polnischen Mittelschulprüfung im Schuljahr 2011/12 eingeführten und mit den Winter’schen Grunderfahrungen korrespondierenden prozessartigen Ziele werden in den Testaufgaben dieses Zeitraums nur mit geringen kategorialen Schwerpunktsetzungen abgebildet. Für eine umfassendere Realisierung beider Vorgaben sollten jedoch substantielle Anteile des mathematischen Argumentierens innerhalb des kognitiven Anforderungsspektrums der Testaufgaben verankert werden. Um das diagnostische Potential der polnischen Testaufgaben zu erhöhen, wäre es darüber hinaus wünschenswert, dass innerhalb der einzelnen Tätigkeiten schwerpunktlos verschiedene Anspruchsniveaus verlangt werden; die CKE spricht in diesem Kontext von einem Erfassen des – nicht operationalisierten – „Grades, in dem der Mittelschüler die curricularen Vorgaben erfüllt“ (CKE 2010, S. 61). Diesbezüglich zeigen die Ergebnisse der ersten Studie, dass sich die eingeforderten Tätigkeiten mit Ausnahme der Kategorie „Umgang mit mathematischen Darstellungen“ stets entweder auf dem einfachen oder auf dem mittleren Niveau konzentrieren. Eine derart geringe Aktivierungstiefe innerhalb der Kategorien liefert folglich eine weniger differenzierte Rückmeldung darüber, wie der Entwicklungstand bezüglich der einzelnen prozessartigen Ziele der Lernenden ist. Dass dieses Entwicklungspotential in konzeptioneller Hinsicht offenbar erkannt, kurz- wie auch mittelfristig dennoch ungenutzt zu bleiben scheint, belegt die *leitende Zielvorgabe*, die durch die CKE im Rahmen einer weiteren Reformen der zentralen Prüfungen im Schuljahr 2018/19 formuliert wird: „Die Testaufgaben werden das Fähigkeitsniveau überprüfen, das in den prozessartigen Zielen des Rahmenplans der Allgemeinbildung beschrieben wird: (...)“ (CKE 2017, S. 5). Damit wird im Gegensatz zu der vorangegangenen Reform der polnischen Mittelschulprüfung im Jahre 2012 ein Bezug zwischen einem Fähigkeitsniveau, das offenbar auf eine Differenzierung des Anspruchs abzielt, und den prozessartigen Zielen in den Blick genommen. Eine erste Einsicht in die curricular-organisatorischen Grundlagen, wie den Rahmenplan (vgl. MEN 2017) und das ergänzende Informationsmaterial (vgl. CKE 2017), der ab dem Schuljahr 2018/19 in Polen sukzessive eingeführten Reformen durch Scheja deutet jedoch darauf hin, dass eine Charakterisierung dafür notwendiger Anforderungsniveaus innerhalb der prozessartigen Ziele auch aktuell nicht erfolgt. Eine Umsetzung der leitenden CKE-Vorgabe, die durch eine Konkretisierung der Komplexitätsniveaus innerhalb der prozessartigen Ziele bedingt wird, bleibt somit bisher offen.

Die vorliegenden Befunde zu Nordrhein-Westfalen zeigen ebenfalls, dass aus der Perspektive der drei Winter'schen Grunderfahrungen für die Testaufgaben der ZP 10 Gestaltungsvorschläge formuliert werden können. Um hier einen umfassenderen Beitrag zur Realisierung der Struktur-, Anwendungs- und auch Problemorientierung zu erreichen, scheint in erster Linie eine höhere Gewichtung der Strukturorientierung wünschenswert. Diese sollte durch substantielle, innermathematisch verortete Problemkontexte, die sich zudem über alle Anspruchsniveaus – von Standardmodellierungen bis hin zu Modellreflexion- und Modellentwicklung – verteilen, umgesetzt werden. Das aktuelle kontextuelle Anforderungsprofil der ZP 10-Aufgaben (vgl. Scheja 2016) orientiert sich offenbar zu einseitig an Anwendungen, d. h. am Verständnis von Mathematik „zur Deutung und Modellierung, zum besseren Verständnis und zur Beherrschung primär nicht-mathematischer Phänomene" (Heymann 1996, S. 278). Da das Argumentieren ebenfalls eher in geringem Maße zum kognitiven Anspruchsniveau der ZP 10-Aufgaben und somit auch zur Entwicklung und Kultivierung einer argumentativen Grundhaltung im Mathematikunterricht beiträgt, ist auch hier eine Erhöhung des Beitrages dieser kognitiven Tätigkeit geboten. Ein vergrößertes Maß an Argumentationen ließe die Grunderfahrung der Problemorientierung intensiver realisieren.

Für Lernende, die kurz vor dem Übergang in die Schulformen der ISCED 3 Levels stehen, würde man mit einer derart vergrößerten Bandbreite der kognitiven Anforderungen potentiell anschlussfähiges Wissen und Kompetenzen fördern, die nach Neubrand und Neubrand hinreichend breite Vorstellungen zu zentralen Begriffen der Mathematik, „argumentatives Denken und die Einbettung der Einzelkenntnisse in einen systematisch geordneten Wissensbestand" (2010, S. 173) aufweisen.

Abschließend stellt sich die Frage nach möglichen Ursachenfeldern für das charakteristische Anforderungsprofil der ZP 10-Aufgaben. In der hierarchischen Anordnung der Konzeptionsabfolge KMK-Bildungsstandards → NRW-Kernlehrplan → ZP 10-Aufgaben erscheint es denkbar, dass letztere zwar, wie gezeigt, die Vorgaben der KMK-Bildungsstandards nur eingeschränkt abbilden, der nordrhein-westfälische Kernlehrplan jedoch durchaus in hohem Maße als steuerungswirksam identifizierbar ist. In diesem Falle wäre die Diskrepanz zwischen den KMK-Wirkungserwartungen und den Anforderungen der ZP 10-Aufgaben möglicherweise auf eine *Inkongruenz* der beiden verbindlichen Vorgaben – der KMK-Bildungsstandards sowie des NRW-Kernlehrplans – zurückzuführen.

Eine erste Analyse eines derart verortbaren Zusammenhangs weist in die Richtung, dass die ZP 10-Aufgaben die Kernlehrplanvorgaben hinreichend breit realisieren und liefert damit vorsichtige Hinweise auf eine Bestätigung dieser Inkongruenz-Vermutung (Neubrand & Neubrand 2010, S. 171). Diese zeigt sich bereits auf der Ebene der *mathematischen Kompetenzen* unter anderem darin (s. KMK 2003, S. 7 ff.), dass „mathematisch Argumentieren" (K1) und „Kommunizieren" (K6) im nordrhein-westfälischen Kernlehrplan zu einer prozessbezogenen Kompetenz, dem „Argumentieren/ Kommunizieren" (MSW 2004, S. 12), zusammengefasst wird, wodurch aufgrund der Obligatorik des Kernlehrplans der relativ geringe Beitrag des Argumentierens zum kognitiven Anspruchsniveau der ZP 10-Aufgaben erklärt werden könnte. Als ein weiterer Hinweis auf die Inkongruenz beider Vorgaben kann das Vorkommen der Kompetenzen „Mathematische Darstellungen verwenden" (K4) sowie „Mit symbolischen, formalen und technischen Elementen der Mathematik umgehen" (K5) (KMK 2003, S. 7 ff.) gesehen werden, die im NRW-Kernlehrplan jedoch *keine* Berücksichtigung finden, was wiederum im Zusammenhang mit den festgestellten substantiellen Defiziten in der Realisierung der Grunderfahrung der Strukturorientierung in den ZP 10-Aufgaben stehen kann. Weiterführende mathematikdidaktische Forschungsarbeiten, die a) die Steuerungswirksamkeit der Lehrpläne auf Länderebene und b) die Kongruenz der Lehrpläne der einzelnen Bundesländer einerseits und der KMK-Bildungsstandards andererseits beleuchten und somit den tatsächlichen Ansatzpunkt für konzeptionelle Steuerungs-Korrekturen in den Blick nehmen, fehlen jedoch bisher.

Zweite Studie

Es gilt als empirisch gesichert, dass kognitiv herausfordernde Lernaufgaben einen positiven Einfluss auf die Leistungsentwicklung der Lernenden haben, wodurch die Fähigkeit der Lehrkräfte zum bedarfsorientierten Erkennen, Vorbereiten und Verändern von Aufgabenpotentialen als eine der zentralen qualitätsbeeinflussenden Größen des Mathematikunterrichts betrachtet werden kann. Als ein tragender Bestandteil professioneller Kompetenz wird diese Fähigkeit in Lernprozessen erworben und bedarf „explizit geschaffener Lerngelegenheiten, wie sie in der Phase der Lehrerausbildung und später in der Weiterbildung zu finden sind" (Kunter et al. 2011, S. 60). Mit Blick auf die Ergebnisse der zweiten Studie (Scheja & Castelli 2018), die jedoch nur bedingt verallgemeinerbar sind, können vor diesem Hintergrund inhaltliche und organisatorische Entwicklungsvorschläge identifiziert werden. So bedarf es in der Aus- und Weiterbildung von Mathematiklehrkräften offenbar entsprechender Weiterbildungsmodule, in denen flexibel anwendbares *Wissen* bezüglich kognitiver Merkmale von Mathematikaufgaben generiert wird. Dieser Fortbildungsbedarf ist fachspezifisch und scheint vielmehr grundsätzliche Züge aufzuweisen (vgl. Drüke-Noe et al. 2017; Hammer 2016; Hillje 2012), da offenbar nicht nur anspruchsgenerierende Kategorien der vorliegenden Studie in hohem Maße unbekannt – zumindest jedoch unbewusst – sind, sondern auch bereits als implementiert geltende Kategoriensysteme obligatorischer Unterrichtsvorgaben (v. a. KMK-Bildungsstandards und NRW-Kernlehrplan). Vor diesem Hintergrund wäre in zukünftigen Studien eine detaillierte Erfassung der tatsächlichen Konzeptionskriterien und deren Differenzierung hinsichtlich Intensität bei Lern- wie auch Testaufgaben durch die Lehrkräfte wünschenswert, da darin potentielle Anknüpfungspunkte für zukünftige Fortbildungen gesehen werden.

Bei der Durchführung der videobasierten Fortbildung wurde deutlich, dass vermutlich aufgrund der engen inhaltlichen Ausrichtung der Videosequenzen nicht alle Lehrkräfte einen wünschenswerten Kompetenzzuwachs aufwiesen. So scheint hier eine erfolgreiche „Veränderung von Lehrerkognitionen“ im Sinne der Fortbildungsziele (Ebene 2, vgl. Lipowsky 2010) primär dadurch effektiver förderbar, dass die begleitenden Handreichungen sowohl eine hohe inhaltliche Passgenauigkeit zu den jeweiligen Sequenzen aufweisen als auch verwandte Aufgaben- bzw. Kodierungsbeispiele beinhalten (vgl. Ball & Cohen 1996), die vornehmlich auf Analogien zurückgreifen. Ein gänzlicher Verzicht auf eine persönliche Unterstützung durch Experten während der Umsetzung wird ebenfalls kritisch gesehen, da dies vermutlich die festgestellten Verzögerungen bzw. Schwierigkeiten im Wissensaufbau vereinzelter Lehrkräfte, die nur zum Teil dem erhöhten Lernaufwand zuzuschreiben ist, aufgrund fehlender flexibler Unterstützungsalternativen bestehen ließe. Dies gilt insbesondere für begriffliche Schwierigkeiten beim Erlernen und der flexiblen Anwendung der Anspruchsniveaus, die durch eine Berücksichtigung von strukturellen „Gelingensfaktoren“ (Dreher et al. 2018, S. 231), wie der intendierten kollektiven Partizipation (vgl. Garet et al. 2001), offenbar nicht gänzlich aufgefangen werden kann. Im Sinne der Effektivität erscheint es insofern sinnvoll, ähnliche inhaltlich eng ausgerichtete Fortbildungen, die videobasiert sind, beispielsweise durch Präsenzphasen mit professionellen Kommunikationsmöglichkeiten zu ergänzen (vgl. Maldonado 2002). Damit jedoch nachhaltige Wirkungen auf der Ebene des „unterrichtspraktischen Handelns“ (Ebene 3, vgl. Lipowsky 2010) zu erwarten sind und die mehrfach festgestellte Aufgabensteuerung durch zentrale Prüfungen damit zugleich weniger als ein „passives Kopieren“ von Aufgabenmerkmalen gesehen werden kann, ist es nach Auffassung des Autors notwendig, dass kollektive Partizipationen systemimmanent, vor allem im Sinne der Top-Down-Steuerung initiiert werden (Fachkonferenz → schulinterne fachbezogene Lehrplanarbeit → jahrgangsbezogene Reihenplanung).

4 Literatur

Anderson, L. W. & Krathwohl, D. R. (2001). *A Taxonomy for Learning, Teaching and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. New York: Longman.

Ball, D.L., & Cohen, D. (1996). Reform by the book: what is – or might be – the role of curriculum materials in teacher learning and instructional reform? *Educational Researcher*, *25* (9), 6 – 14.

Baumert, J., Lehmann, R., Lehrke, M., Schmitz, B.; Clausen, M., Hosenfeld, I., Köller, O., Neubrand, J. (1997). *TIMSS – Mathematisch-naturwissenschaftlicher Unterricht im internationalen Vergleich*. Deskriptive Befunde. Opladen.

Baumert, J. & Köller, O. (2000). Unterrichtsgestaltung, verständnisvolles Lernen und multiple Zielerreichung im Mathematik- und Physikunterricht der gymnasialen Oberstufe. In J. Baumert (Hrsg.), *TIMSS-III. Dritte Internationale Mathematik- und Naturwissenschaftsstudie. Mathematische und naturwissenschaftliche Bildungam Ende der Schullaufbahn. 2. Mathematische und physikalische Kompetenzen am Ende der gymnasialen Oberstufe.* (S. 271 – 315). Opladen: Leske u. Budrich.

Baumert, J. & Kunter, M. (2011). Das Kompetenzmodell von COACTIV. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 29 – 53). Münster: Waxmann.

Biermann, M., Wiegand, B. & Blum, W. (2003). Nicht „irgendwie", sondern zielgerichtet. Aufgaben – Lernen fördern – Selbstständigkeit entwickeln, *Friedrich Jahresheft* 2003, 32 – 35.

Black, P. J. & William, D. (1998). Assessment and Classroom Learning. *Assessment in Education,* 5, 7 – 74.

Blömeke, S., Risse, J., Müller, C., Eichler, D. & Schulz, W. (2006). Analyse der Qualität von Aufgaben aus didaktischer und fachlicher Sicht. Ein allgemeines Modell und seine exemplarische Umsetzung im Unterrichtsfach Mathematik. *Unterrichtswissenschaft*, 34 (4), 330 – 357.

Blum, W., Drüke-Noe, C., Hartung, R. & Köller, O. (2010). *Bildungsstandards Mathematik: konkret. Sekundarstufe I: Aufgabenbeispiele Unterrichtsanregungen, Fortbildungsideen*. Berlin: Cornelsen Verlag Scriptor.

Bruner, J. (1972). *Der Prozeß der Erziehung.* Berlin: Berlin-Verlag.

Büchter, A. & Leuders, T. (2005). *Mathematikaufgaben selbst entwickeln: Lernen fördern – Leistung überprüfen.* Berlin: Cornelsen Scriptor.

Büchter, A. & Pallack, A. (2012). Methodische Überlegungen und empirische Analysen zur impliziten Standardsetzung durch zentrale Prüfungen. *Journal für Mathematikdidaktik*, 33, 59 – 85.

Christiansen, B. & Walther, G. (1986). Task and Activity. In B. Christiansen, A. G. Howson & M. Otte, M. (Eds.), *Perspectives on mathematics education* (243 – 307).

Dordrecht: D. Reidel Publishing Company.

CKE (2006). *Egzamin w trzeciej klasie gimnazjum z zakresu przedmiotow matematyczno-przyrodniczych*. ttps://cke.gov.pl/images/stories/Arkusze2006gimn/gmp_a1.pdf. Zugegriffen: 10.09.2018.

CKE (2010). *Informator o egzaminie gimnazjalnym od roku szkolnego 2011/12*. http://www.cke.edu.pl/images/files/Gimnazjum_2011_2012/Informator_G1.pdf. Zugegriffen: 01.02.2018.

CKE (2011a). *Ramowy opis polskiego systemu egzaminów zewnetrznych.* www.bip.cke.edu.pl/bip_download.php?id=1293. Zugegriffen: 02.02.2018.

CKE (2011b). *Egzamin w klasie trzeciej gimnazjum z zakresu przedmiotow matematyczno-przyrodniczych.* https://cke.gov.pl/images/stories/0001_Gimnazja_2011/mat/gm-1-112.pdf. Zugegriffen am 12.01.2019.

CKE (2012). *Egzamin w klasie trzeciej gimnazjum. Czesc matematyczno-przyrodnicza - matematyka.* *https://cke.gov.pl/images/stories/0000000000000000000gim/mat/GM-M1-122.pdf*. Zugegriffen: 14.09.2018.

CKE (2014). *Egzamin w klasie trzeciej gimnazjum. Czesc matematyczno-przyrodnicza - matematyka.* https://www.cke.edu.pl/images/_EGZAMIN_GIMNAZJALNY/Arkusze-egzaminacyjne/2015/ matematyka/Zeszyt_zadan_GM-MX1-152.pdf. Zugegriffen: 02.10.2018.

CKE (2015). *Egzamin w klasie trzeciej gimnazjum. Czesc 2. matematyka. https://cke.gov.pl/images/_EGZAMIN_GIMNAZJALNY/Arkusze-egzaminacyjne/2015/matematyka/Zeszyt_zadan_GM-MX1-152.pdf. Zugegriffen: 10.08.2018.*

CKE (2017). *Informator o egzaminie ósmoklasisty z matematyki od roku szkolnego 2018/2019.* Warszawa: Centralna Komisja Egzaminacyjna.

Cohors-Fresenborg, E., Sjuts, J. & Sommer, N. (2004). Komplexität von Denkvorgängen und Formalisierung von Wissen. In M. Neubrand (Hrsg.), *Mathematische Kompetenzen von Schülerinnen und Schülern in Deutschland: Vertiefende Analysen im Rahmen von PISA 2000* (S. 109 – 144). Wiesbaden: VS Verlag für Sozialwissenschaften.

Deci, E. L. & Ryan, R. M. (1993). Die Selbstbestimmungstheorie der Motivation und ihre Bedeutung für die Pädagogik. *Zeitschrift für Pädagogik*. 39 (2), 223 – 238.

Deci, E. L., Koestner, R., Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin* 125, 627 – 668.

Doyle, W. (1983). Academic work. *Review of educational research*, 53 (2), 159 – 199.

Dreher, U., Holzäpfel, L., Leuders, T., Stahnke, R. (2018). Problemlösen lehren lernen – Effekte einer Lehrerfortbildung auf die prozessbezogenen mathematischen Kompetenzen von Schülerinnen und Schülern. *Journal für Mathematikdidaktik*, 39 (2), 227 – 256.

Drüke-Noe, C. (2014). *Aufgabenkultur in Klassenarbeiten im Fach Mathematik. Empirische Untersuchungen in neunten und zehnten Klassen.* Wiesbaden: Springer Spektrum.

Drüke-Noe, C., Maier, U., Kleinknecht, M., Metz, K., Hoppe, H. & Bohl, T. (2017). Lehrkräfte bei der Auswahl und Gestaltung von Aufgaben professionalisieren: Wie verändert eine Fortbildung zu Merkmalen kognitiv-aktivierender Aufgaben die aufgabenbezogene Unterrichtsplanung von Lehrkräften? In K. Zierer & S. Wernke (Hrsg.), *Die Unterrichtsplanung: Ein in Vergessenheit geratener Kompetenzbereich?! - Status Quo und Perspektiven aus Sicht der empirischen Forschung* (S. 208 – 223). Bad Heilbrunn: Klinkhard.

Dylak, S. (2013). *Architektura wiedzy w szkole*. Warszawa: Difin.

Evertson, C. M. & Worsham, M. E. (2006). Classroom management for elementary teachers – 7th ed. – Boston.

Gagne, R., Briggs, L. & Wager, W. (1992). *Principles of Instructional Design* (4. Aufl.). Fort Worth, TX: HBJ College Publishers.

Garet, M., Porter, A. C., Desimone, L., Birman, B.F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal*, *38* (4), 915 – 945.

Hammer, S. (2016). *Professionelle Kompetenz von Mathematiklehrkräften im Umgang mit Aufgaben in der Unterrichtsplanung: theoretische Grundlegung und empirische Untersuchung*. Dissertation: LMU München.

Hattie, J. & Timperley, H. (2007). The power of feedback. *Review of Educational Research,* 77, H. 1, 81 – 112.

Hattie, John A. C. (2012). *Visible Learning for Teachers – Maximizing impact on learning.* London & New York. Routledge.

Helm, C. (2016). Zentrale Qualitätsdimensionen von Unterricht und ihre Effekte auf Schüleroutcomes im Fach Rechnungswesen. *Zeitschrift für Bildungsforschung,* 6 (2), 101 – 119.

Helmke, A. (2004). *Unterrichtsqualität: Erfassen, Bewerten, Verbessern* (3. Aufl.). Seelze: Kallmeyersche Verlagsbuchhandlung.

Heymann, H. W. (1996). *Allgemeinbildung und Mathematik*. Weinheim und Basel: Beltz Verlag.

Hiebert, J. & Wearne, D. (1993). Instructional tasks, classroom discourse, and students learning in second-grade arithmetic. *American Educational Research Journal*, 30 (2), 393 – 425.

Hillje, M. (2012). *Professionelles Wissen von Mathematiklehrkräften und die didaktische Strukturierung – Analysen von Unterrichtsplanung und Unterrichtsdurchführung anhand von Aufgaben.* Dissertation: Universität Oldenburg.

Holzberger, D., & Kunter, M. (2016). Unterricht aus der Perspektive der Pädagogischen Psychologie und empirischen Unterrichtsforschung. In: J. Möller, M. Köller & T.

Riecke-Baulecke (Eds.), *Basiswissen Lehrerbildung: Schule und Unterricht – Lehren und Lernen* (S. 39 – 52). Seelze: Klett Kallmeyer.

Hošpesová, A. & Tichá, M. (2010). Reflexion der Aufgabenbildung als Weg zur Erhöhung der Lehrerprofessionalität, in Nuhrenbörger et al. (Hrsg.), *Mathematik im Denken der Kinder. Anregungen zur mathematik-didaktischen Reflexion* (S. 122 – 126), Seelze: Klett/Kalmeyer.

Jordan, A., Ross, N., Krauss, S., Baumert, J., Blum, W., Neubrand, M., Löwen, K., Brunner, M., & Kunter, M. (2006). *Klassifkationsschema für Mathematikaufgaben: Dokumentation der Aufgabenkategorisierung im COACTIV-Projekt.* Berlin: Max-Planck-Institut für Bildungsforschung.

Jordan, A., Krauss, S., Löwen, K., Blum, W., Neubrand, M., Brunner, M., Kunter, M., & Baumert, J. (2008). Aufgaben im COACTIV-Projekt: Zeugnisse des kognitiven Aktivierungspotentials im deutschen Mathematikunterricht. *Journal für Mathematikdidaktik*, 29, 83 – 107.

Karpinski, M., Grudniewska, M., & Zambrowska,M. (2013). *Nauczanie matematyki w gimnazjum - Raport z badania*. Warszawa: IBE.

Kasper, B. (2017). *Implementation von Schulqualität Governanceanalyse des Orientierungsrahmens Schulqualität in Niedersachsen*. Springer VS.

Klieme, E. & Bos, W. (2000): Mathematikleistung und mathematischer Unterricht in Deutschland und Japan: Triangulation quantitativer und qualitative Forschungs ansätze im Rahmen der TIMSS-Studie, *Zeitschrift für Erziehungswissenschaft*, 3 (3), 359 – 379.

Klieme, E., Pauli, C. & Reusser, K. (2009). The Pythagoras Study: Investigating effects of teaching and learning in Swiss and German classrooms. In T. Janik & T. Seidel (Hrsg.), *The power of video studies in investigating teaching and learning in the classroom* (S. 137 – 160). Munster: Waxmann.

Klieme, E.; Bürgermeister, A.; Harks, B.; Blum, W.; Leiß, D.; Rakoczy, K. (2010). Leistungsbeurteilung und Kompetenzmodellierung im Mathematikunterricht. Projekt Co2CA1. *Zeitschrift für Pädagogik*, Beiheft; 56, 64 – 74.

KMK (2003). Bildungsstandards im Fach Mathematik für den Mittleren Schulabschluss. München: Wolters Kluwer Deutschland GmbH.

Knoll, S. (2003). *Verwendung von Aufgaben in Einführungsphasen des Mathematikunterrichts.* Marburg: Tectum Verlag.

Konarzewski, K. (2004). *Reforma oswiaty. Podstawa programowa i warunki kształcenia.* Warszawa: ISP.

Konarzewski, K. (2008). *Przygotowanie uczniów do egzaminu: pokusa łatwego zysku. Raport badawczy.* Warszawa: ISP.

Kounin, J. S. (2006). *Techniken der Klassenführung* (Original der deutschen Ausgabe, 1976). Münster: Waxmann.

Krüger, M. (2015). *Aufgabenkultur in zentralen Abschlussprüfungen: Exploration und Deskription naturwissenschaftlicher Aufgabenstellungen im internationalen Vergleich.* Münster: Waxmann.

Kunter, M., Brunner, M., Baumert, J., Klusmann, U., Krauss, S., Blum, W., Jordan, A., & Neubrand, J. (2005). Der Mathematikunterricht der PISA-Schülerinnen und Schüler. *Zeitschrift für Erziehungswissenschaft*, 8 (4), 502 – 520.

Kunter, M. & Voss, T. (2011). Das Modell der Unterrichtsqualität in COACTIV: Eine multikriteriale Analyse. In: M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften – Ergebnisse des Forschungsprogramms COACTIV* (S. 85 – 113). Münster: Waxmann.

Kunter, M. & Baumert, J. (2011). Das COACTIV-Forschungsprogramm zur Unterstützung professioneller Kompetenz von Lehrkräften – Zusammenfassung und Diskussion. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften – Ergebnisse des Forschungs programms COACTIV* (S. 345 – 366). Münster: Waxmann.

Kunter, M., Kleichmann, T., Klussmann, U. & Richter, D. (2011). Die Entwicklung professioneller Kompetenz von Lehrkäften. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV. (S. 56 – 68). Münster: Waxmann.

Kühn, S., & Drüke-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., & Holzäpfel, L. (2011). Kognitive Aktivierung im Mathematikunterricht. *Unterrichtswissenschaft*, 39 (3), 213 – 230.

Leuders, T. (2015). Aufgaben in Forschung und Praxis. In R. Bruder, L. Hefendehl-Hebeker, B. Schmidt-Thieme & H.-G. Weigand (Eds.), *Handbuch Mathematik didaktik* (S. 433 – 458), Heidelberg: Springer.

Lipowsky, F. (2010). Lernen im Beruf. Empirische Befunde zur Wirksamkeit von Lehrerfortbildungen. In F. H. Müller, A. Eichenberger, M. Lüders & J. Mayr (Hrsg.), *Lehrerinnen und Lehrer lernen. Konzepte und Befunde zur Lehrerfortbildung* (S. 51 – 70), Münster: Waxmann.

Luthiger, H. (2014). *Differenz Lern- und Leistungssituationen, eine explorative Studie zu ihrer theoretischen Grundlegung und empirischen Überprüfung*. Münster Waxmann.

Maciejak, A., Mosiek, P., Okrasiński, W., Sęk, A. (2012). On Mathematical Modelling in High Shool. *Mathematica Applicanda*, 40, 131 – 144.

Maier, U., Kleinknecht, M., Metz, K. & Bohl, T. (2010). Ein allgemeindidaktisches Kategoriensystem zur Analyse des kognitiven Potenzials von Aufgaben. *Beitäge zur Lehrerbildung*, 28 (1), 84 – 96.

Maier, U., Bohl, T., Metz, K., & Kleinknecht, M. (2011). Einflüsse von Merkmalen des Testsystems und Schulkontextfaktoren auf die Akzeptanz und Rezeption von zentralen Testrückmeldungen durch Lehrkräfte. *Journal of Educational Research Online*, 3, 62 – 93.

Maldonado, L. (2002). *Effective professional development. Findings from research*. New York: The College Board.

MEN (2009). *Podstawa programowa z komentarzami. Tom 6: Edukacja matematyczna i techniczna w szkole podstawowej, gimnazjum i liceum.* Warszawa: Ministerstwo Edukacji Narodowej.

MEN (2013). *Programme for international student assessment. Wyniki badania 2012 w Polsce.* www.ifispan.waw.pl. Zugegriffen: 14.12.2016.

MEN (2017). *Podstawa programowa krztalcenia ogolnego – szkola podstawowa matematyka*. Warszawa: Ministerstwo Edukacji Narodowej.

Meyer, M. (2007). *Entdecken und Begründen im Mathematikunterricht. Von der Abduktion zum Argument*. Dissertation. Hildesheim: Franzbecker.

Mietzel, G. (2003). *Pädagogische Psychologie des Lernens und Lehrens*, 7. überarbeitete Auflage. Göttingen, Bern, Toronto, Seattle: Hogrefe.

Mrozek, E. (2015). Introducting additive compare problems – traditional vs. constructivist approach. *Didactica Mathematicae,* 37, 27 – 46.

MSW (2004). *Kernlehrplan für die Realschule in Nordrhein-Westfalen* (Heft 3302). Frechen: Ritterbach-Verlag.

Mullis, I. V. S. et al. (2003) : TIMSS Assessment Frameworks and Specifications 2003. Chestnut Hill, MA: Boston College 2. Aufl.

Neubrand, J. (2002). *Eine Klassifikation mathematischer Aufgaben zur Analyse von Unterrichtssituationen: Selbsttätiges Arbeiten in Schülerarbeitsphasen in den Stunden der TIMSS-Video-Studie*. Hildesheim: Verlag Franzbecker.

Neubrand, M., Klieme, E., Lüdtke, O. & Neubrand, J. (2002). Kompetenzstufen und Schwierigkeitsmodelle für den PISA-Test zur mathematischen Grundbildung *Unterrichtswissenschaft*, 30 (1), 100 – 119.

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 und Oldenburg.

Neubrand, M., Jordan, A., Krauss, S., Blum, W. & Löwen, K. (2011). Aufgaben im COACTIV Projekt: Einblicke in das Potenzial für kognitive Aktivierung im Mathematikunterricht. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 115 – 132). Münster: Waxmann.

OECD (2001). *Lernen für das Leben. Erste Ergebnisse der internationalen Schulleistungsstudie PISA 2000.* Paris: OECD.

Oelkers, J. & Reusser, K. (2008). *Expertise: Qualität entwickeln, Standards sichern, mit Differenz umgehen*. Bonn: BMBF.

Pauli, Ch., Drollinger-Vetter, B., Hugener, I, Lipowsky, F. (2008), Kognitive Aktivierung im Mathematikunterricht, *Zeitschrift für Pädagogische Psychologie* 22 (2), 127 – 133.

Piaget, J. (1928). *The language and thought of the child*. New York: Hartcourt Brace.

Pianta, R. C. & Hamre, B. K. (2009). Conceptualization, Measurement, and Improvement of Classroom Processes: Standardized Observation Can Leverage Capacity. *Educational Researcher*, 38 (2), 109 – 119.

Pieronkiewicz, B. (2017). Edukacja matematyczna z perspektywy aksjologiczno-teleologicznej, *Edukacja*, S. 55 – 66.

Prabhu, N.S. (1987). *Second Language Pedagogy.* Oxford: Oxford University Press.

Renkl, A. (1991). *Die Bedeutung der Aufgaben-und Rückmeldungsgestaltung für die Leistungsentwicklung im Fach Mathematik.* (Nicht veröffentlichte Dissertation). Universität Heidelberg.

Resnick, L. & Ford, W. (1981). *The Psychology of Mathematics for Instruction*. Hillsdale: Lawrence Erlbaum Associates.

Riecke-Baulecke, T. (2017). Unterrichtsqualität. In: Abshagen, Barzel, Kramer, Riecke-Baulecke, Rösken-Winter, Selter (Hrsg.): *Basiswissen Lehrerbildung: Mathematik unterrichten* (S. 149 - 167). Seelze: Kallmayer.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science* 18, 119 – 144.

Samborska, M. (2015). Using Student-Friendly Tasks to promote mathematical understanding in a middle school classroom. *Didactica Mathematicae,* 37, 47 – 74.

Scheja, B. (2007). *Der Wandel von Bildungssystemen am Beispiel von Deutschland und Polen unter besonderer Berücksichtigung der PISA-Studie.* Dissertation, Universität zu Köln.

Scheja, B. (2016). Cognitive activation through mathematics tasks in the context of centralised examinations on ISCED 2 Level, using the example of North-Rhine Westphalia. *Didactica Mathematicae*, 38, 175 – 202.

Scheja, B. (2017a). Kognitive Aktivierung durch Mathematikaufgaben zentraler Abschlussprüfungen - Eine Vergleichsanalyse der polnischen Mittelschulprüfung und der Zentralen Prüfung aus Nordrhein-Westfalen. *Journal für Mathematikdidaktik,* 38 (2), 291 – 322.

Scheja, B. (2017b). The changing cognitive requirement of test tasks in mathematics – A longitudinal study of the Polish middle school examinations. *Didactica Mathematicae*, 39, 101 – 130.

Scheja, B. & Castelli, S. (2018). Developing teacher competence regarding the cognitive requirement of mathematical tasks – a video-based intervention. *Didactica Mathematicae*, 40, 65 – 95.

Schiepe-Tiska, A., Reiss, K., Obersteiner, A., Heine, J.-H., Seidel, T. & Prenzel, M. (2013). Mathematikunterricht in Deutschland: Befunde aus PISA 2012. In M. Prenzel,C. Sälzer, E. Klieme & O. Köller (Hrsg.), *PISA 2012. Fortschritte und Herausforderungen in Deutschland* (S. 123 – 154). Münster: Waxmann.

Shayer, M. & Adhami, M. (2007). Fostering cognitive development through the context of Mathematics: Results of the CAME project. Educational studies in Mathematics, 64(3), 265 – 291.

Shuell, T. (1996). Teaching and learning in the classroom context. In C. D. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (S. 726 – 764). New York: Simon & Schuster Macmillan.

Schratz, M., Schwarz, J., Westfall-Greiter, T. (2011). Auf dem Weg zu einer Theorie lernseits von Unterricht. In: W. Meseth, M. Proske, F.-O. Radtke. (Hrsg.).: *Unterrichtstheorien in Forschung und Lehre* (S. 103 – 129). Bad Heilbrunn: Klinkhardt.

Stankov, L., Lee, J., Luo, W. & Hogan, D. (2012). Condence: A better predictor of academic achievement than self-efficacy, self-concept and anxiety? *Learning and Individual Differences*, 22 (6), 747 – 758.

Stein, M. & Lane, S. (1996). Instructional Tasks and the Development of Student Capacity to Think and Reason: An Analysis of the Relationship between Teaching and Learning in a Reform Mathematics Project. *Educational Research and Evaluation*, 2 (1), 50 – 80.

Stein, M. K., Grover, B. W. & Henningsen, M. (1996). Building Student Capacity for Mathematical Thinking and Reasoning. An Analysis of Mathematical Tasks Used in Reform Classrooms. *American Educational Research Journal* 33 (2), 455 – 488.

Terhart, E., Baumgart, F., Meder, N. & von Sychowski, G. (2009). Standardisierte Prüfungsverfahrenin der Erziehungswissenschaft: Kontext, Formen, Konsequenzen. *Erziehungswissenschaft*, 20, 9 – 36.

Turner, J., Meyer, D., Cox, K., Logan, C., DiCintio, M. & Thomas, C. (1998). Creating contexts for involvement in mathematics. *Journal of Educational Psychology*, 90 (4), 730 – 745.

UNESCO (2011). *Revision of the International Standard Classification of Education (ISCED)*. 36C/19 of the 36th Session of the General Conference based on 34C/Resolution 20, Paris.

Wang, M., Haertel, G., & Walberg, H. (1993). Toward a knowledge base for school learning. *Review of Educational Research*, 63 (3), 249 – 294.

Weinert, F. (2000). *Lehrern und Lernen für die Zukunft – Ansprüche an das Lernen in der Schule.* Vortrag gehalten am 29.3.2000 im Pädagogischen Zentrum Rheinland-

Pfalz in Bad Kreuznach. Sonderdruck.

Winter, H. (1995). Mathematikunterricht und Allgemeinbildung. *Mitteilungen der Gesellschaft für Didaktik der Mathematik*, 61, 37 – 46.

Wittmann, E. C. (1990). Wider die Flut der „bunten Hunde“ und der „grauen Päckchen“: Die Konzeption des aktiv-entdeckenden Lernens und des produktiven Übens In E. C. Wittmann & G. N. Müller (Hrsg.), *Handbuch produktiver Rechenübungen* (S. 152 – 166). Stuttgart: Klett Verlag.

Wunder, D. (2002). Im Spannungsfeld von Wissenschaft und Politik. Anmerkungen zur politischen und öffentlichen Rezeption von PISA. *Die Deutsche Schule*, 94, (2), 138 – 147.

Wygotski, L. S. (1978). *Mind in society. The development of higher psychological processes.* Cambridge: Harvard University Press.

Zahorska, M. (2002). *Szkola miedzy panstwem, spoleczenstwem a rynkiem*. Warszawa: Zak.

Zahorska, M. (2012). *Zawirowania systemu edukacji.* Warszawa: Wydawnictwa Uniwersytetu Warszawskiego.

5 Zusammenfassungen

5.1 Zusammenfassung

In Anbetracht einer, offenbar weitflächig verbreiteten Normierung von Gelegenheitsstrukturen durch Lern- wie auch Testaufgaben, gewinnt die Frage nach den *kognitiven Aufgabenmerkmalen* und deren Komposition seitens der fachdidaktischen Forschung zunehmend an Bedeutung. Hierzu gibt es bislang wenig Forschung: dieses Forschungsdesiderat wird im Rahmen der vorliegenden kumulativen Dissertation aufgeriffen. Dies geschieht, indem das Konstrukt der kognitiven Aktivierung, wie auch das Teilkonstrukt des kognitiven Anspruchs, zugrunde gelegt wird; diese werden jeweils als spezifische inhärente Aufgabenpotentiale definiert, die mittels Kategoriensystemen erfassbar werden. Ihre Anwendung ermöglicht eine schrittweise Erarbeitung der beiden zentralen Anliegen dieser Dissertation:

1. *Das kognitive Aktivierungspotential von Mathematikaufgaben der polnischen Mittelschulprüfung und der nordrhein-westfälischen ZP 10 soll erfasst, analysiert und bewertet werden (erste Studie).*

Die ebenfalls am Ende des ISCED 2 Levels durchgeführte ZP 10 eignet sich für eine derartige Gegenüberstellung in besonderem Maße. So ist der Anteil der ZP-10-Prüflinge an Prüflingen, die deutschlandweit an Abschlussprüfungen zum Erwerb des Mittleren Schulabschlusses teilnehmen, in Nordrhein-Westfalen im Ländervergleich am höchsten. Darüber hinaus weisen beide Prüfungsinstrumente eine hohe funktionale Äquivalenz (Selektion und Vergleichbarkeit) auf und der Umgang mit den Prüfungsergebnissen seitens der Bildungspolitik zeigt in Polen und in NRW zahlreiche Parallelen.

2. *Kennen, verstehen und nutzen Lehrkräfte kognitive Aufgabenmerkmale als Beschreibungsstrukturen von Mathematikaufgaben und inwiefern kann eine Lehrerfortbildung zu einer Steigerung dieser Kompetenz führen (zweite Studie)?*

Die empirischen Ergebnisse der ersten Studie zeigen unter anderem, dass sowohl die in der ZP 10 als auch die in der polnischen Mittelschulprüfung eingesetzten Aufgaben im Vergleich zu PISA- wie auch zu den deutschen Unterrichts-Aufgaben ein verhältnismäßig hohes kognitives Aktivierungspotential aufweisen. Hierzu tragen insbesondere sprachlogisch komplexe Aufgabenstellungen und mathematische Darstellungen bei, denen ebenfalls anspruchsvolle, da mehrschrittige Verarbeitungsprozesse nachgeordnet werden. Die Befunde des nordrhein-westfälisch-polnischen Vergleichs weisen dennoch auf zwei mit Einschränkungen einander entsprechende Aufgabenprofile hin. So zeigt der polnische Aufgabensatz ein im Vergleich zur ZP 10 ausgewogeneres Anforderungsspektrum, was primär durch eine im Hinblick auf Maß und Niveau unterschiedliche Berücksichtigung der Aufgabenkontexte (innermathematisch und außermathematisch) bedingt wird.

Die Ergebnisse der Eingangsbefragung der zweiten, explorativen Studie deuten zunächst darauf hin, dass die Lehrkräfte hinsichtlich der kategorialen Bandbreite verglichen mit den verbindlichen KMK-Bildungsstandards nicht im erwarteten Maße über die Kenntnis anspruchsgenerierender Kategorien verfügen. Sie sind zudem allenfalls punktuell in der Lage, kategoriale Anspruchsniveaus zu erkennen und differenziert auf Aufgaben bzw. deren Veränderungen anzuwenden. Dieses kategoriale Anspruchswissen der Lehrkräfte kann auf der Grundlage der fünf zugrundegelegten kognitiven Kategorien (mathematisches Modellieren, Argumentieren, Umgang mit mathematischen Darstellungen, technisches Arbeiten und Umgang mit mathematischen Texten) mittels der gestuften videobasierten Fortbildung erweitert werden. Hier deuten die Ergebnisse der Ausgangsbefragung darauf hin, dass anschließend erheblich mehr anspruchsgenerierende Kategorien beachtet werden, wobei dies zugleich innerhalb der einzelnen Kategorien deutlich differenzierter geschieht. Diese Tendenz zur Exaktifizierung des Konstruktes „kognitiver Anspruch" zeigt sich bei den teilnehmenden Lehrkräften zugleich im Rückgang von Aufgabenmerkmalen, wie beispielsweise Bearbeitungszeit, Werkzeugnutzung oder auch Maß der Einübung, deren Intensität bisher keine empirischen Hinweise auf die Aufgabenschwierigkeit liefert. Damit kann davon ausgegangen werden, dass die durchgeführte Fortbildung eine zielgerichtete Variation des kognitiven Anspruchsniveaus von Mathematikaufgaben begünstigt und somit auch positive Wirkungen auf der Ebene des unterrichtspraktischen Handelns erwartet werden können.

5.2 Abstract

In view of the apparently widespread standardization of opportunity structures through learning and test tasks, the question of cognitive task characteristics and their composition is becoming increasingly important for didactic research. There has been little research on this subject so far: this research desideratum will be dealt with in the context of the present cumulative dissertation. This is done by taking the construct of cognitive activation as a basis, as well as the partial construct of the cognitive claim; these are each defined as specific inherent task potentials that can be grasped by means of category systems. Their application enables a step-by-step elaboration of the two central concerns of this dissertation:

> *1. The potential for cognitive activation of mathematical tasks of the Polish Middle School examination and the North Rhine-Westphalian ZP 10 will be recorded, analysed and evaluated (first study).*

The ZP 10, which was also conducted at the end of the ISCED 2 level, is particularly suitable for such a comparison. For example, the proportion of ZP-10 candidates who take part in final examinations for the acquisition of the intermediate school leaving certificate throughout Germany is the highest in North Rhine-Westphalia compared to other federal states. In addition, both examination instruments have a high functional equivalence (selection and comparability) and the handling of examination results by

education policy shows numerous parallels.

2. Do teachers know, understand and use cognitive task characteristics as descriptive structures of mathematical tasks and to what extent can further teacher training lead to an increase in this competence (second study)?

The empirical results of the first study show, among other things, that the tasks used in both ZP 10 and the Polish secondary school examination have a relatively high potential for cognitive activation compared with PISA and German teaching tasks. This is in particular due to linguistically complex tasks and mathematical representations, which are also subordinated to demanding, as multi-step processing. The findings of the North Rhine-Westphalian-Polish comparison nevertheless point to two task profiles comparable with limitations. Thus the Polish task set shows a more balanced requirement spectrum compared to ZP 10, which is primarily due to a different consideration of the task contexts (intra-mathematical and extra-mathematical) with regard to measure and level.

The results of the initial survey of the second, explorative study initially indicate that the teachers do not have the expected knowledge of claim-generating categories with regard to the categorical range compared with the binding KMK educational standards. In addition, they are at most selectively able to recognise categorical levels of entitlement and apply them in a differentiated manner to tasks or their changes. This categorical claim knowledge of teachers can be expanded on the basis of the five underlying cognitive categories (mathematical modelling, argumentation, dealing with mathematical representations, technical work and dealing with mathematical texts) by means of tiered video-based advanced training. Here, the results of the initial survey indicate that considerably more claim-generating categories will then be considered, although this will also be much more differentiated within the individual categories. This tendency towards the exactification of the construct „cognitive requirement“ is also evident among the participating teachers in the decline of task characteristics, such as processing time, tool use or also the degree of practice, the intensity of which so far does not provide any empirical evidence of the task difficulty. It can thus be assumed that the further training provided favours a targeted variation of the cognitive demands level of mathematical tasks and that positive effects can therefore also be expected at the level of practical teaching action.

6 Anhang

Erklärung (gemäß § 4 Abs 1 Punkt 9)

„Ich versichere, dass ich die von mir vorgelegte Dissertation selbständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit − einschließlich Tabellen, Karten und Abbildungen −, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie − abgesehen von unten angegebenen Teilpublikationen − noch nicht veröffentlicht worden ist, sowie, dass ich eine solche Veröffentlichung vor Abschluss des Promotionsverfahrens nicht vornehmen werde. Die Bestimmungen der Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Prof. Dr. Benjamin Rott betreut worden.“

1. Artikel

Scheja, B. (2016). Cognitive activation through mathematics tasks in the context of centralised examinations on ISCED 2 Level, using the example of North-Rhine Westphalia. *Didactica Mathematicae*, 38, 175 – 202.

2. Artikel

Scheja, B. (2017a). Kognitive Aktivierung durch Mathematikaufgaben zentraler Abschlussprüfungen − Eine Vergleichsanalyse der polnischen Mittelschulprüfung und der Zentralen Prüfung aus Nordrhein-Westfalen. *Journal für Mathematikdidaktik,* 38(2), 291 − 322.

3. Artikel

Scheja, B. (2017b). The changing cognitive requirement of test tasks in mathematics – A longitudinal study of the Polish middle school examinations. *Didactica Mathematicae,* 39, 101 − 130.

4. Artikel

Scheja, B. & Castelli, S. (2018). Developing teacher competence regarding the cognitive requirement of mathematical tasks − a video-based intervention. *Didactica Mathematicae,* 40, 65 − 95.

Darlegung der eigenen Anteile an der videobasierten Studie und an der Publikation IV (Scheja & Castelli 2018)

Die Verantwortlichen der Studie sind Dr. Bruno Scheja (St.-Ursula-Gymnasium Brühl) und Sabine Castelli (Herder-Gymnasium Minden/Universität Bielefeld). Hierbei war der Verfasser der vorliegenden Dissertation an allen Phasen der Studie, d. h. bei der Konzeption, der Organisation, der Durchführung und abschließenden Berichtsabfassung maßgeblich beteiligt.

Während der konzeptionellen Phase umfasst dies die Auswahl, die Entwicklung und die Kodierung von Aufgaben, die innerhalb der Studie zum Einsatz kamen, die Fragebogenerstellung für die zwei Erhebungszeitpunkte der Studie (Vor- und Nachuntersuchung) sowie die Ausarbeitung der Fortbildungsunterlagen (Grundlagen der vier Videosequenzen und die Handreichungen für Lehrkräfte). Während der Organisation und Durchführung, die in Brühl, Köln und Minden stattfand, fielen einerseits die Auswahl der teilnehmenden Lehrkräfte und andererseits der Druck und die Administration der Handreichungen für Lehrkräfte in den Verantwortungsbereich des Verfassers. Die Verfilmung der vier Fortbildungssequenzen und ihre Platzierung in einem WIKI (sowie auf mobilen Datenträgern) erfolgten in Zusammenarbeit mit Sabine Castelli. Sämtliche Auswertungen der Fragebögen wurden durch den Verfasser gesammelt und in einem gemeinsamen Datensatz aufbereitet. Der Verfasser hat abschließend den Entwurf zum Bericht im Rahmen einer Publikation in dem englischsprachigen Journal *Didactica Mathematicae* eigenständig verfasst.