

Svenja Mareike Kühn/Christina Drüke-Noe

Qualität und Vergleichbarkeit durch Bildungsstandards und zentrale Prüfungen?

Ein bundesweiter Vergleich von Prüfungsanforderungen im Fach Mathematik zum Erwerb des Mittleren Schulabschlusses

Zusammenfassung: Zur Sicherung von Qualität und Vergleichbarkeit sollen alle Bundesländer zentrale Prüfungen zum Erwerb des Mittleren Schulabschlusses auf Basis der Bildungsstandards durchführen. Der Beitrag stellt die Befunde einer Längsschnittstudie vor, in der im Ländervergleich Prüfungsaufgaben ($N = 3530$) im Fach Mathematik mit Hilfe eines Kategoriensystems auf Basis der Bildungsstandards analysiert wurden. Die Untersuchung erfasst auf der Basis theoretischer Annahmen der Implementationsforschung, wie hoch die Übereinstimmung zwischen den bundesweit gültigen Bildungsstandards und der tatsächlichen Aufgabenpraxis in den Prüfungen der Länder zum Erwerb des Mittleren Schulabschlusses ist. Es zeigt sich, dass die Prüfungsaufgaben nur sehr unausgewogen die drei Dimensionen der Bildungsstandards (Leitideen, Kompetenzen, Anforderungsbereiche) abbilden. Darüber hinaus sind die Aufgaben durch eine erhebliche länderspezifische Heterogenität gekennzeichnet. Die Befunde weisen auf eine Diskrepanz zwischen bildungspolitischen Wirkungserwartungen und den Wirkungserfahrungen hin, da das Ziel vergleichbarer Prüfungsanforderungen über alle Länder hinweg offensichtlich bislang kaum erfüllt wird.

Schlagnote: Bildungsstandards, zentrale Abschlussprüfung, Implementation, Aufgabenanalyse, Mathematik

Mit dem Mittleren Schulabschluss (im Folgenden: MSA), welcher an allgemeinbildenden Schulen nach der zehnten Jahrgangsstufe vergeben wird, erwerben Schülerinnen und Schüler in Deutschland die Zugangsberechtigung zu weiterführenden berufs- und studienqualifizierenden Bildungswegen. Vor dem Hintergrund der zunehmenden Entkopplung von Schulform und Schulabschluss (Baumert, Cortina & Leschinsky, 2008; Klemm, 2004) sowie durch schulstrukturelle Ausdifferenzierungen nicht nur *zwischen* den 16 Ländern, sondern auch *innerhalb* dieser Länder zwischen verschiedenen Schulformen und Schulen, werden Qualität und Vergleichbarkeit des MSA seitens der Anschlussysteme (z. B. Ausbildungsbetriebe) immer wieder in Frage gestellt (DIHK, 2012). Empirische Studien bestätigen in diesem Zusammenhang, dass zwischen Absolvierenden und Absolventen unterschiedlicher Schulen, Schulformen und Länder erhebliche Leistungsunterschiede bestehen (z. B. Baumert & Watermann, 2000; Köller, Baumert, Cortina, Trautwein & Watermann, 2004; Neumann, Nagy, Trautwein & Lüdtkke, 2009), woraus eine unterschiedliche Wertigkeit formal gleicher Abschlüsse in Abhängigkeit von vergebender Einzelschule, Schulform bzw. Bundesland resultiert, was eine divergierende Akzeptanz des Abschlusszertifikates des Mittleren Schulabschlusses

und damit eine Hierarchisierung der Zugangschancen zum Ausbildungs- und Arbeitsmarkt – und so der Lebenschancen jedes Einzelnen – impliziert (Baumert, Cortina & Leschinsky, 2008; Schuchart, 2007).

Die Sicherung der Gleichwertigkeit schulischer Ausbildung und die Vergleichbarkeit der erworbenen Schulabschlüsse sind von großer Bedeutung für deren Akzeptanz seitens der abnehmenden Systeme und damit für die Akzeptanz des gesamten Berechtigungssystems (Böttcher & Rürup, 2007; Klemm, 2004). Im Kontext der bildungspolitischen Debatten über die Entwicklung schulischer Qualität und die Sicherung der Gleichwertigkeit schulischer Zertifikate wurden daher zahlreiche ergebnisorientierte Steuerungsmaßnahmen implementiert, die primär auf schulextern definierte Standards und deren Überprüfung durch zentrale Instanzen fokussieren (Maag Merki, 2010; Wacker, Maier & Wissinger, 2012). Für den MSA beschloss die KMK im Zuge dieses vielfach beschriebenen ‚steuerungspolitischen Paradigmenwechsels‘ (Steffens, 2007) zu Beginn der letzten Dekade die Einführung von Bildungsstandards für bestimmte Schulabschlüsse und Fächer (KMK, 2003a), verbunden mit länderspezifischen Maßnahmen, um zu überprüfen, über welche Kompetenzen Schülerinnen und Schüler zu bestimmten Zeitpunkten der Schullaufbahn verfügen. Bildungsstandards stellen damit eine Zielebene sowohl für das unterrichtliche Handeln der Lehrerinnen und Lehrer als auch für die entsprechenden standardbasierten Überprüfungen (z. B. Lernstandserhebungen) dar (Köller, 2011) und *konkretisieren sich durch Aufgaben* in unterrichtlichen Lerngelegenheiten, in traditionellen Bewertungskontexten zur Überprüfung des Unterrichtserfolgs (z. B. Klassenarbeiten), in normierten Testaufgaben sowie in Prüfungen zur Zertifizierung von Schulabschlüssen (Blum, Drüke-Noe, Hartung & Köller, 2006; Gehrman, Hericks & Lüders, 2010; 0, 2007). Ob und inwieweit die Bildungsstandards tatsächlich in den schulischen und unterrichtlichen, aber auch administrativen Prozessen der Länder implementiert sind (vgl. auch KMK, 2006), ist bislang nur unzureichend bzw. noch nicht erfasst (Berner, Oelkers & Reusser, 2008). So liegen auch keine empirischen Befunde zum Grad der Standardorientierung in der Prüfungsaufgabenpraxis der Länder zum Erwerb des MSA vor, weshalb unklar ist, inwiefern die länderspezifischen Aufgaben – wie von der KMK beabsichtigt – bundesweit *vergleichbar* im Sinne von *gleichwertig* sind. Ob die mit der Einführung der Bildungsstandards intendierten Steuerungswirkungen – die Sicherung von Qualität und Vergleichbarkeit der Prüfungsanforderungen zum Erwerb des MSA – erreicht werden, bedarf daher empirischer Klärung.

In diesem Beitrag wird eine Studie vorgestellt, die im Rahmen der DFG-Nachwuchsakademie *Empirische Schul- und Unterrichtsforschung* entstanden ist.¹ Die retrospektiv angelegte Längsschnittstudie greift das skizzierte Forschungsdesiderat auf und betrachtet kategoriengeleitete Prüfungsaufgaben im Fach Mathematik im bundesweiten Vergleich. Sie möchte einen Beitrag zur Implementationsforschung (Gräsel, 2010; Prenzel, 2010) leisten, indem erfasst wird, wie hoch die Übereinstimmung zwischen bundesweit gültigen Bildungsstandards und der tatsächlichen Aufgabenpraxis in den Prüfungen der

1 Das diesem Beitrag zu Grunde liegende Vorhaben wurde mit Mitteln der Deutschen Forschungsgemeinschaft unter dem Förderkennzeichen KU 2798/1-1 gefördert.

Länder zum Erwerb des MSA ist. Einleitend sollen zunächst die bildungspolitische Ausgangslage, der nationale und internationale Stand der Forschung zur Implementation von Bildungsstandards sowie grundlegende theoretische Vorannahmen dargelegt werden. Es folgt eine Erläuterung des methodischen Vorgehens der Studie, an die sich die Darstellung und Diskussion zentraler Ergebnisse anschließt.

1. Theoretischer Hintergrund

Im Zuge der zahlreichen bildungspolitischen Reformen zur Sicherung und Weiterentwicklung der Qualität und Vergleichbarkeit schulischer Bildung beschloss die KMK zu Beginn der letzten Dekade die Einführung von Bildungsstandards für den MSA für die Fächer Deutsch, Mathematik und erste Fremdsprache sowie für die naturwissenschaftlichen Unterrichtsfächer.² Bildungsstandards beruhen auf breit verstandenen, fachlich verankerten und die jeweilige Fachstruktur widerspiegelnden Bildungszielen, die in den Schulen erreicht werden sollen und die wiederum eingebettet sind in fachübergreifende Bildungsziele. Sie sind Leistungsstandards, die fachbezogene Kompetenzen benennen, die für die weitere schulische und berufliche Ausbildung von Bedeutung sind und die anschlussfähiges Lernen ermöglichen. Vor dem Hintergrund der einleitend skizzierten Pluralität von Bildungswegen und einer Orientierung an erfolgreiche(re)n Staaten wurden bewusst keine schulformspezifischen, sondern abschlussbezogene Bildungsstandards verabschiedet. Diese erfüllen eine Orientierungsfunktion und setzen verbindliche Kompetenzerwartungen, sie erfüllen eine Evaluationsfunktion für externe und interne Evaluationen und dienen der Qualitätsentwicklung im Unterricht (Klieme et al., 2003). Als Beitrag zur Bildung soll der Mathematikunterricht allen Schülerinnen und Schülern insbesondere die drei Grunderfahrungen von Winter (1995) ermöglichen – Anwendungs-, Struktur- und Problemorientierung –, deren Realisierung in den Bildungsstandards gefordert wird (vgl. KMK, 2003b), was eine ausgewogene Verteilung der Kompetenzen voraussetzt. Da Leistungsaufgaben ein Abbild von Unterrichtsaufgaben sind (u. a. Sacher, 2000), lässt sich dieser in den Bildungsstandards formulierte Anspruch (wenigstens weitgehend) auf Aufgaben in Abschlussprüfungen übertragen. Bildungsstandards spannen eine Inhalts-, eine Prozess- und eine Anspruchsdimension auf: Speziell im Fach Mathematik sind dies in fünf Leitideen festgelegte verbindliche Inhalte, wobei eine Leitidee Inhalte mehrerer Stoffgebiete (Arithmetik, Algebra, Geometrie, Stochastik) umfassen kann. Die Leitideen sind L1 *Zahl*, L2 *Messen*, L3 *Raum und Form*, L4 *Funktionaler Zusammenhang* sowie L5 *Daten und Zufall*. In einer weiteren Dimension kommen die prozessbezogenen Kompetenzen hinzu, mit denen zentrale Aspekte des mathematischen Arbeitens in hinreichender Breite erfasst werden können (Blum et al., 2006) und welche Schülerinnen und Schüler mit Erwerb des MSA im Fach

2 Dieser Beitrag fokussiert ausschließlich die Entwicklungen für die Sekundarstufe I (mit Blick auf den MSA); die Entwicklungen für den Hauptschulabschluss, den Primarbereich sowie das Abitur werden hier nicht berücksichtigt.

Mathematik in der Auseinandersetzung mit mathematischen Inhalten erworben haben sollten. Hierzu gehören: K1 *mathematisch argumentieren*; K2 *Probleme mathematisch lösen*; K3 *mathematisch modellieren*; K4 *mathematische Darstellungen verwenden*; K5 *mit symbolischen, formalen und technischen Elementen der Mathematik umgehen*; und K6 *kommunizieren*. Mathematische Kompetenzen zeigen sich insbesondere in Form von Performanz beim Lösen von Aufgaben (Blum et al., 2006). In der dritten Dimension beschreiben Anforderungsbereiche kognitiv unterschiedliche Anspruchsniveaus – sind also *nicht* mit dem in den EPA gleichlautenden Begriff gleichbedeutend –, die für die erfolgreiche Bearbeitung und Lösung einer Aufgabe erforderlich sind. Die Bildungsstandards unterscheiden drei Anforderungsbereiche (AFB), und zwar: AFB 1 *Reproduzieren*, AFB 2 *Zusammenhänge herstellen* und AFB 3 *Verallgemeinern und Reflektieren*. Dabei nehmen Anspruch und damit kognitive Komplexität, wie auch meist die damit verbundene empirische Schwierigkeit, im Allgemeinen von Anforderungsbereich zu Anforderungsbereich zu. Insbesondere die Dimension Anforderungsbereich unterscheidet Bildungsstandards in ihrer Konzeption wesentlich von (traditionell stark inhaltlich ausgerichteten) Lehrplänen.

So liegt für alle Länder und für jede einzelne Schule ein bundesweit einheitlicher und gemeinsam vereinbarter, seit Beginn des Schuljahres 2004/2005 verbindlich umzusetzender Bezugsrahmen vor, um die Qualität und Vergleichbarkeit schulischer Bildungsgänge und ihrer Abschlüsse zu sichern und weiterzuentwickeln (Köller, 2011; Maag Merki, 2010). Dabei sieht die KMK mit Blick auf die föderale Vielfalt bewusst von Detailregelungen ab und betrachtet die Bildungsstandards als Rahmenvorgaben, auf deren Basis die Länder unterschiedliche, aber gleichwertige Maßnahmen implementieren sollen (KMK, 2009), was zu differenziellen Entwicklungen in den Bundesländern führt, beispielsweise im Hinblick auf die Lehrplanarbeit, Schulbuchentwicklung oder Fortbildungsmaßnahmen (Gördel, 2008; Köller, 2009; Zeitler, Köller & Tesch, 2010).

Ob und inwieweit die Bildungsstandards in den Ländern tatsächlich implementiert sind und damit die intendierten Steuerungswirkungen erfüllen, ist bislang nur unzureichend untersucht worden (Berner et al., 2008). Die wenigen bislang vorliegenden Forschungsarbeiten (Böttcher & Dicke, 2008; Pant, Vock, Pöhlmann & Köller, 2008a, 2008b; Tesch, 2010; Zeitler, 2010) weisen zunächst darauf hin, dass die mit der Implementation der Bildungsstandards verbundenen Reformervorgaben und Optimierungsversprechen auf unterrichtlicher Ebene offensichtlich kaum erfüllt werden (Wacker et al., 2012). Es zeigt sich insgesamt, dass offenbar erhebliche Vorbehalte gegenüber der bisherigen Form der Standardimplementierung und den damit verbundenen Unterstützungsmaßnahmen existieren. Im internationalen Vergleich lassen sich diese Skepsis und Distanz – in differierenden Ausmaßen – auch in anderen Ländern zu Beginn der Entwicklung von Standards nachzeichnen (z. B. Altrichter & Posch, 2007; Hochuli, 2011; Klieme et al., 2003).

Neben ihrer Bedeutung für die Unterrichtsentwicklung geben die Bildungsstandards eine Zielperspektive vor, deren Erreichung gemäß KMK durch landesweite bzw. länderübergreifende Tests oder in zentralen oder dezentralen Abschlussprüfungen am Ende von Schullaufbahnen überprüft werden soll (KMK, 2003a, S. 4). In welchem Maße

die in den Bildungsstandards ausgewiesenen Kompetenzen von den Schülerinnen und Schülern erreicht werden, wird durch unterschiedliche Verfahren erfasst, die auf Grund differierender Zielsetzungen und des Grades der Standardorientierung konzeptionelle Unterschiede aufweisen (Maag Merki, 2010): Schullaufbahnbegleitende, länderübergreifende (Köller, Knigge & Tesch, 2010) bzw. landesweite Vergleichsuntersuchungen (z. B. Steffens, 2010) sowie Lernstandserhebungen unter Federführung des IQB prüfen, inwieweit die in den Bildungsstandards formulierten Kompetenzen von den Schülerinnen und Schülern (schon) erreicht werden. Diese *standardbasierten Leistungstests* beziehen sich weniger auf spezifische Lehrplaninhalte, sondern orientieren sich unmittelbar an den domänenspezifischen Anforderungen der Bildungsstandards und erfassen auf unterschiedlichen in den Bildungsstandards ausgewiesenen Anspruchsniveaus (sog. Anforderungsbereiche) und in einer ausgewogenen Verteilung inhalts- und prozessbezogene Kompetenzen, die Schülerinnen und Schüler in ihrer gesamten bisherigen Schulzeit erworben haben. Darüber hinaus werden (mit Ausnahme von Rheinland-Pfalz) in allen übrigen Ländern zentrale Abschlussprüfungen durchgeführt, in denen alle Schülerinnen und Schüler eines Landes in bestimmten Fächern die gleichen schriftlichen, landesweit einheitlichen Prüfungsaufgaben bearbeiten (zur Heterogenität der Prüfungsverfahren vgl. Kühn, 2013). Die Aufgaben in diesen *curricular validen, standardorientierten Prüfungen* beziehen sich auf die landesspezifischen Lehrpläne (bzw. Kerncurricula, Rahmenrichtlinien), die sich an den Bildungsstandards orientieren (sollten).³ Mit Blick auf die inhaltlichen Anforderungen der Prüfungen (Kühn, 2013) legt die überwiegende Mehrheit der Länder die Inhalte der gesamten Sekundarstufe I zu Grunde, während einige Länder nur die Inhalte der (beiden) letzten Klassenstufe(n) vor dem Abschluss prüfen (Kühn, 2013). Damit müssen Absolventinnen und Absolventen in der Abschlussprüfung – je nach Land – (ausgewählte) Inhalte und Kompetenzen aus einem bis sechs Schuljahren der Sekundarstufe I nachweisen, was eine differenzielle Verwendungspraxis der Bildungsstandards bei der Prüfungsaufgabengestaltung – gerade im Hinblick auf die Verteilung der Leitideen – erwarten lässt. Ob und inwieweit die Bildungsstandards in den länderspezifischen Abschlussprüfungsverfahren zum Erwerb des MSA tatsächlich umgesetzt werden und wie schriftliche Prüfungsaufgaben, an denen der Grad der Umsetzung gemessen werden kann, ausgestaltet sind, ist jedoch bislang nicht umfassend untersucht worden: Die wenigen bislang vorliegenden, ausschließlich auf das Fach Mathematik bezogenen Arbeiten von Roppelt, Keller, Leiß, Drücke-Noe und Blum (2008, im Fokus: Berlin und Brandenburg, Prüfungsaufgaben 2007) sowie von Neubrand & Neubrand (2010, im Fokus: NRW, Prüfungsaufgaben 2008) zur Analyse von Prüfungsaufgaben hinsichtlich ihrer Orientierung an den Bildungsstandards weisen jedoch darauf hin, dass bezüglich des Ausschöpfens der Breite des Kompetenzspektrums

3 Köller (2009) weist darauf hin, dass die Entwicklung in den Ländern im Bereich der Lehrplangestaltung ganz unterschiedlich weit vorangeschritten ist: Während einige Länder die Bildungsstandards konsequent in kompetenzorientierten Curricula umsetzen, ist die Einführung der Bildungsstandards in anderen Ländern weitgehend an der Lehrplanarbeit vorbeigegangen, sodass hier weiterhin die traditionelle Inhaltsorientierung der Lehrpläne vorherrscht.

der Bildungsstandards noch ‚Entwicklungsbedarf‘ besteht, da bestimmte (insbesondere anspruchsvollere) Kompetenzen und der höchste Anforderungsbereich kaum berücksichtigt werden. Diese Einschätzungen sind anschlussfähig an bereits vorliegende Befunde zur Aufgabenkultur im Mathematikunterricht und in Klassenarbeiten, die in der Regel aus vielen ähnlichen Einzelaufgaben ohne einen Anwendungs- oder Lebensweltbezug bestehen und vorwiegend die Beherrschung und Automatisierung grundlegender Routineverfahren (formal-technisches Arbeiten) mit dem Ziel erfordern, eine bestimmte Lösung zu erarbeiten (z. B. BLK, 1998; Bromme, Seeger & Steinbring, 1990; Henn & Kaiser, 2001; Leuders, 2004; Drüke-Noe, 2012). Die Arbeiten von Roppelt et al. (2008) sowie Neubrand und Neubrand (2010) sind jedoch regional begrenzt – Analysen, die Aufgaben aus einer steuerungstheoretischen Perspektive hinsichtlich der angestrebten nationalen Vergleichbarkeit in Form eines umfassenden Vergleichs der MSA-Prüfungen aller Bundesländer in den Blick nehmen, fehlen bislang. Zudem beziehen sich die o. g. Studien auf den Zeitraum kurz nach der Einführung der Bildungsstandards, sodass keine merklichen Effekte auf die Aufgabengestaltung zu erwarten sind (Gräsel, 2010; Prenzel, 2010). Aktuelle Analysen mit einer zeitlichen Distanz zum Implementationsbeginn liegen ebenfalls nicht vor. Zusammenfassend kann festgehalten werden, dass die Wirkungen der bundesweit einheitlichen Bildungsstandards auf die Gestaltung der länderspezifischen Prüfungsaufgaben bislang nicht umfassend empirisch untersucht worden sind. Somit liegen keine Erkenntnisse darüber vor, inwiefern Bildungsstandards im Hinblick auf die Qualität und Vergleichbarkeit von Prüfungsanforderungen zum Erwerb des MSA tatsächlich steuerungswirksam sind.

Aus erziehungswissenschaftlicher Sicht ist vor dem Hintergrund neuer Steuerungstheorien (Altrichter & Maag Merki, 2010) und im Rekurs auf die Innovations-/Implementationsforschung (Gräsel, 2010; Prenzel, 2010) davon auszugehen, dass die Gestaltung landeszentral gestellter Prüfungsaufgaben nicht wie beabsichtigt ausschließlich auf Basis der Bildungsstandards erfolgt, sondern durch persönliche Dispositionen und Handlungsroutinen einzelner Akteure, organisationale Strukturen und (Fach-)Kulturen sowie spezifische Kontextfaktoren multifaktoriell beeinflusst wird (Altrichter & Maag Merki, 2010), was eine differenzielle Gestaltung der Prüfungsaufgaben erwarten lässt. Ergebnisse der Implementationsforschung weisen in diesem Zusammenhang auf eine systematische Diskrepanz zwischen bildungspolitischen Absichten in Form formaler Weisungen (wie vorgegebenen Bildungsstandards) und ihrer Umsetzung in der Praxis hin (Gräsel & Parchmann, 2004) bzw. machen deutlich, dass Innovationen im Bildungsbereich nur schwer Verbreitung finden und es häufig zu einem „Versanden“ kommt (Gräsel, 2010).

2. Methodisches Vorgehen

Die vorliegende Studie untersucht die Umsetzung der Bildungsstandards bei der Zusammenstellung schriftlicher Prüfungsaufgaben zum Erwerb des MSA im Hinblick auf die Qualität und Vergleichbarkeit der länderspezifischen Prüfungsanforderungen, wobei

hier Aufgabenstellungen aus dem Prüfungsfach Mathematik aller 15 zentral prüfenden Bundesländer analysiert werden. Sie leistet damit einen Beitrag zur Implementationsforschung, indem erfasst wird, ob und inwieweit die Bildungsstandards in den Prüfungsaufgaben der Länder zum Erwerb des MSA tatsächlich implementiert sind.

Die Aufgabenanalyse basiert auf einem fachspezifischen Kategoriensystem und erfasst klar definierte Merkmale von Prüfungsaufgaben, die sich aus den Vorgaben der Bildungsstandards für das Fach Mathematik ergeben (s. o.). Daraus ergibt sich folgende Forschungsfrage:

- 1) Welche Qualitätsmerkmale charakterisieren die schriftlichen Aufgaben im Rahmen der Prüfungsverfahren zum Erwerb des MSA im Fach Mathematik in den 15 zentral prüfenden Bundesländern?

Dabei geht es nicht nur um die Deskription bestimmter Merkmale von Aufgaben, sondern auch um die Erfassung nicht-intendierter Effekte (z. B. die Fokussierung auf bestimmte Aufgabenformate und -inhalte).

Darüber hinaus ermöglicht der Vergleich zwischen den Bundesländern auch eine Identifizierung von Kongruenzen und Inkongruenzen zwischen diesen und damit die Beantwortung der Frage, ob und inwieweit Prüfungsanforderungen zum Erwerb des MSA innerhalb Deutschlands vergleichbar im Sinne von gleichwertig sind. Mit Blick auf diese *vergleichsorientierte Perspektive* lässt sich folgende Forschungsfrage ableiten:

- 2) Inwiefern werden die Anforderungen der Bildungsstandards im Rahmen der länderspezifischen Prüfungsverfahren umgesetzt und damit steuerungswirksam?

Insgesamt gelten Bildungsstandards als Instrument zur Qualitätssicherung und *-entwicklung* im Schulsystem. Daher geht die hier dokumentierte Studie über die Beschreibung eines Ist-Zustandes hinaus und integriert zusätzlich die Entwicklungsperspektive, indem sie als Längsschnittstudie angelegt wird und die Analyse von Prüfungsaufgaben zum Erwerb des MSA über einen Zeitraum von 5 Jahren umfasst (2007-2011), um mögliche Veränderungen im Zeitverlauf aufzeigen zu können. So lässt sich darstellen, ob und inwieweit innerhalb eines Landes eine Weiterentwicklung stattfindet oder ob in aufeinanderfolgenden Jahren immer ähnliche Prüfungsaufgaben gestellt werden. Zudem gilt es zu prüfen, ob sich die Prüfungsaufgaben der einzelnen Bundesländer vor dem Hintergrund der einheitlichen Bildungsstandards im Zeitverlauf einander angleichen. Mit Blick auf diese entwicklungsorientierte Perspektive ergibt sich folgende dritte Forschungsfrage:

- 3) Verändern sich die länderspezifischen Prüfungsaufgaben zum Erwerb des MSA im Fach Mathematik im Zeitverlauf?

Zur Beantwortung dieser Fragestellungen wird eine kategoriengeleitete Analyse zentral gestellter Prüfungsaufgaben in Mathematik aus allen 15 Bundesländern im Rah-

men einer retrospektiven Längsschnittstudie (2007-2011) durchgeführt. Dabei wurden ausschließlich die Prüfungsaufgaben des Haupttermins⁴ berücksichtigt und dabei – sofern die Länder schulformspezifische Prüfungsverfahren vorsehen – nur die Aufgaben für die Schulform, an welcher der MSA am häufigsten vergeben wird (zu den Schulformspezifika vgl. Kühn, 2013).⁵ Insgesamt wurden 77 Abschlussprüfungen mit 3530 *Handlungsaufforderungen* analysiert. Der Begriff der Handlungsaufforderung knüpft an den in der Mathematikdidaktik etablierten „engen Aufgabenbegriff“ (z. B. Jordan et al., 2006) an und ist definiert als kleinste zusammenhängende inhaltsbezogene Aufforderung an den Bearbeiter bzw. die Bearbeiterin einer Aufgabe, eine prozess- oder produktorientierte mathematische Tätigkeit auszuführen (z. B. *Berechne* den Flächeninhalt des Dreiecks). Wenn im Folgenden von ‚Aufgabe‘ oder ‚Aufgabenstellung‘ gesprochen wird, ist damit immer die Analyseeinheit ‚Handlungsaufforderung‘ gemeint.

Die Studie zielt auf die kategorienbasierte Analyse eindeutig definierter Merkmale der Aufgabenstellung, die sich aus den Vorgaben der Bildungsstandards zur Aufgabenkonstruktion ergeben, und umfasst die dort formulierte Prozess-, Inhalts- und Anspruchsdimension der Aufgabengestaltung, d. h. die sechs Kompetenzen, die fünf Leitideen und die drei Anforderungsbereiche (s. o.).

Die Entwicklung des Kategoriensystems, das neben den drei Dimensionen der Bildungsstandards zahlreiche weitere Kategorien umfasst, orientierte sich an inhaltsanalytischen Verfahren (Mayring, 2000) und etablierten Forschungsarbeiten zum Umgang mit Kategoriensystemen (z. B. Seidel, 2003). Zur Sicherung der Qualität des Kategoriensystems wurden verschiedene Maßnahmen durchgeführt; dazu gehörte insbesondere ein intensives, mehrstufiges Training der an der Aufgabenanalyse beteiligten Beurteiler (Langer & Schulz von Thun, 2007). Um festzustellen, inwieweit diese hinreichend trainiert sind, wurde die Beurteilerübereinstimmung berechnet (Wirtz & Caspar, 2002), die nach Abschluss des Trainings für alle Kategorien Werte von $\kappa \geq .75$ und eine prozentuale Übereinstimmung zwischen beiden Beurteilern von wenigstens 88 Prozent aufwies. Damit ist das entwickelte Kategoriensystem ein verlässliches Erhebungsinstrument zur Analyse von Prüfungsaufgaben zum Erwerb des MSA im Fach Mathematik. Die abschließende Analyse wurde von zwei Beurteilern durchgeführt.

Der Fokus der Auswertung liegt auf der Deskription von Aufgabenmerkmalen und erfolgt mittels Kontingenztabellenanalysen (Bortz, Boehnke & Lienert, 2008) für multiple Vergleichsgruppen (hier: Bundesländer). χ^2 -Tests dienen dabei zur Überprüfung

4 Zusätzlich zum Haupttermin, an dem die Prüfung im Regelfall landesweit geschrieben wird, gibt es bis zu drei Nachtermine mit vergleichbaren Aufgaben, an denen die Prüfung aus triftigem Grund nachgeholt werden kann.

5 In Bayern wird der MSA am häufigsten an Realschulen vergeben. Schülerinnen und Schüler bearbeiten in den Abschlussprüfungen je nach gewähltem Schwerpunkt entweder die Aufgaben ‚Mathematik I‘ (bei mathematisch-naturwissenschaftlich-technischem Schwerpunkt) oder ‚Mathematik II‘ (bei wirtschaftlichem bzw. fremdsprachlichem Schwerpunkt). Die empirischen Befunde zeigen – mit zwei Ausnahmen – keine statistisch bedeutsamen Unterschiede zwischen den beiden Aufgabengruppen, sodass die Ergebnisse für Bayern aggregiert berichtet werden.

der statistischen Signifikanz, welche in Form der Zusammenhangsmaße ϕ (Phi, nur für 2×2 -Tabellen) und *Cramers V* quantifiziert wird. Bei der Interpretation der Koeffizienten ist in Anlehnung an Cohen (1988) von einem geringen Zusammenhang auszugehen, wenn ϕ bzw. *Cramers V* $> .10$, von einem mittleren Zusammenhang, wenn ϕ bzw. *Cramers V* $> .30$, und von einem starken Zusammenhang, wenn ϕ bzw. *Cramers V* $> .50$. Auf Grund der Stichprobengröße ($N = 3530$) werden alle Werte ϕ bzw. *Cramers V* $< .20$ als Unterschiede trivialer Größenordnung angesehen, die zwar statistisch signifikant sein mögen, jedoch praktisch bedeutungslos sind und daher nicht weiter interpretiert werden (Sink & Stroh, 2006; Thompson, 2002).

3. Empirische Befunde

Die Darstellung der Befunde erfolgt aus qualitätsorientierter, vergleichsorientierter und entwicklungsorientierter Perspektive entsprechend der zuvor formulierten Forschungsfragen.

3.1 Die qualitätsorientierte Perspektive

Wie bereits dargestellt, umfasst die standardorientierte Aufgabengestaltung in Mathematik eine Prozess-, eine Inhalts- und eine Anspruchsdimension in Form von Kompetenzen, Leitideen und Anspruchsniveaus. Eine Zusammenstellung von Aufgaben im Sinne der Bildungsstandards zeichnet sich zunächst durch eine inhaltliche Reichhaltigkeit aus, indem sie alle Leitideen berücksichtigt. Die folgende Abbildung 1 gibt einen Überblick, inwieweit dies bei der Gestaltung der Prüfungsaufgaben berücksichtigt wird, und differenziert dabei zwischen Pflicht- und Wahlaufgaben⁶.

Zunächst fällt auf, dass sich Pflichtaufgaben hinsichtlich der Verteilung der Leitideen *nicht* substanziell von den Wahlaufgaben unterscheiden: Dort, wo sich statistisch bedeutsame Unterschiede zeigen, sind diese nur von trivialer Größenordnung (alle *Cramers V* $< .10$) und gelten als praktisch nicht bedeutsam. Insgesamt werden die Leitideen *Zahl* und *Messen* in den Prüfungsaufgaben zu gleichen Teilen berücksichtigt, ebenso (jedoch mit geringerem Anteil) *Raum und Form* sowie *Daten und Zufall*. Innerhalb von L5 entfallen allerdings etwa vier Fünftel der Aufgaben auf den Teilbereich *Daten*, weshalb nur sehr wenige Aufgaben insgesamt zum zweiten Teilbereich *Zufall* enthalten sind. Zudem fällt die Dominanz der Leitidee *Funktionaler Zusammenhang* auf.

6 Pflichtaufgaben sind solche Aufgaben, die von *allen* Prüflingen bearbeitet werden müssen (entweder, weil keine Aufgabenauswahlmöglichkeit in der Prüfung besteht, oder weil sich die Prüfung in einen Pflicht- und Wahlpflichtteil gliedert). Wahlaufgaben umfassen einen einheitlichen Aufgabenpool (z. B. mehrere Themengebiete), wobei hier eine Auswahlmöglichkeit besteht, sodass nur eine bestimmte Anzahl dieser Aufgaben bearbeitet werden muss.

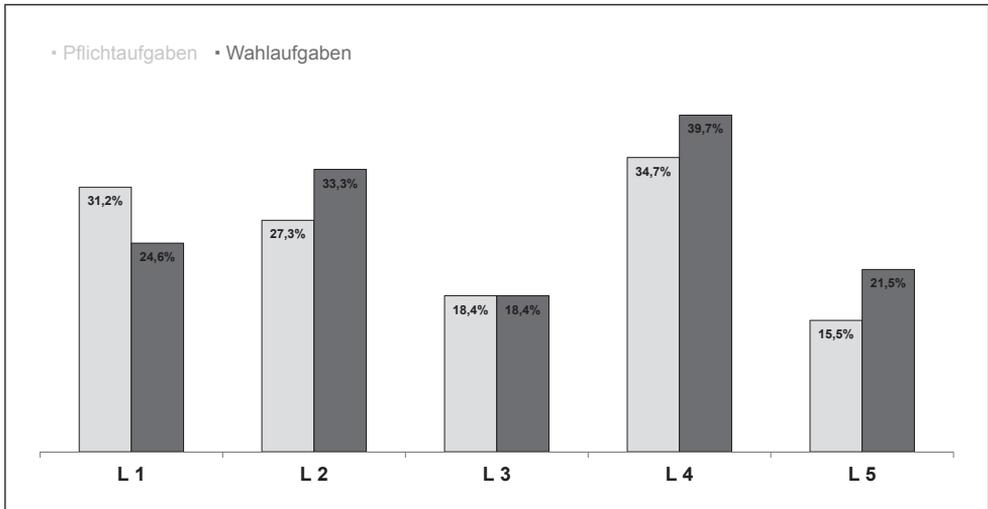


Abb. 1: Verteilung der Leitideen in Prüfungsaufgaben zum Erwerb des MSA in Prozent (N = 3530)

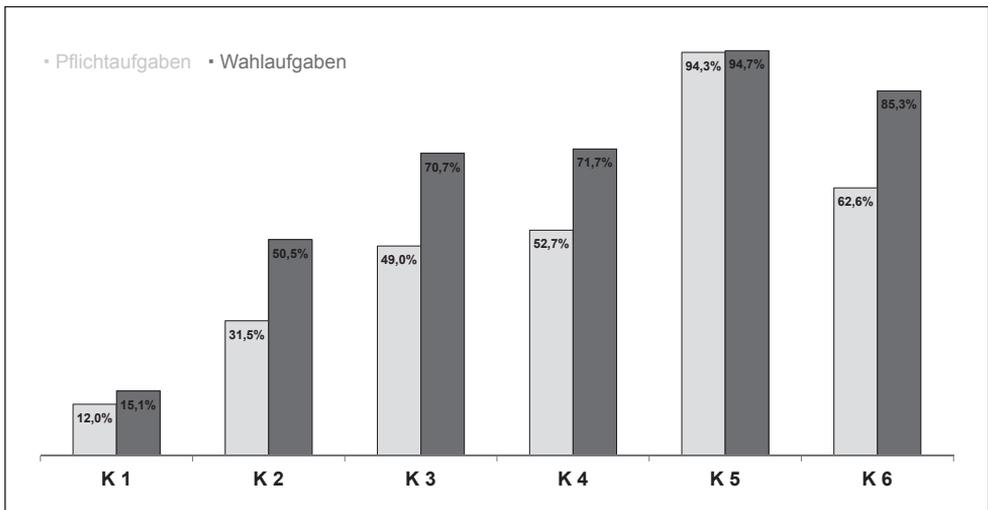


Abb. 2: Verteilung mathematischer Kompetenzen in Prüfungsaufgaben zum Erwerb des MSA in Prozent (N = 3530)

Bildungsstandardkonforme Aufgabenzusammenstellungen zeichnen sich darüber hinaus durch eine hinreichende Breite der geforderten Kompetenzen ohne eine einseitige Bevorzugung bestimmter Kompetenzen – hier ist gerade auch das symbolisch/technisch/formale Arbeiten (K5) zu nennen – aus. Abbildung 2 stellt dar, welche mathematischen Kompetenzen bei der Bearbeitung aller Prüfungsaufgaben erforderlich sind, und differenziert dabei wieder Pflicht- und Wahlaufgaben. Bei der Interpretation der Ergebnisse ist zu berücksichtigen, dass eine Aufgabe durchaus die Anwendung mehrerer Kompetenzen erfordern kann; dies ist gerade bei anspruchsvolleren Aufgaben der Fall.

Mit Blick auf die mathematischen Kompetenzen fällt zunächst wieder auf, dass sich diese innerhalb der einzelnen Aufgabengruppen unausgewogen verteilen. Dabei dominiert in beiden erwartungskonform besonders das formal-technische Arbeiten, also das Anwenden von Faktenwissen (z. B. die Nennung von Regeln, Definitionen etc.) oder Fertigkeiten (d. h. die Anwendung von Rechenverfahren oder das Umgehen u. a. mit Termen oder Diagrammen). Anspruchsvollere mathematische Kompetenzen werden seltener (Problemlösen) bzw. nur sehr selten gefordert (mathematisches Argumentieren). Diese unausgewogene Verteilung der einzelnen Kompetenzen ist innerhalb der Pflichtaufgaben besonders deutlich. Demgegenüber zeigen die Wahlaufgaben andere Akzentuierungen, und sie erfordern neben dem formal-technischen Arbeiten vermehrt auch andere Kompetenzen. Die Unterschiede zwischen Pflicht- und Wahlaufgaben sind jedoch höchstens von geringer Größenordnung ($p < .001$, alle *Cramers V* $< .25$) und damit nicht bedeutsam. Gerade in dieser Verteilung unterscheiden sich diese zentralen Prüfungsaufgaben deutlich von jenen der Lernstandserhebungen, wie sie u. a. in Klasse 8 geschrieben werden (vgl. hierzu auch Roppelt, Katzenbach & Pant, 2010).

Mathematische Kompetenzen können auf unterschiedlichen kognitiven Anspruchsniveaus angewendet werden. Aufgaben im Sinne der Bildungsstandards zeichnen sich daher auch durch eine angemessene Verteilung der realisierten Anforderungsbereiche ohne einseitige Fokussierung auf einen bestimmten Komplexitätsgrad aus. Insbesondere Aufgaben zum AFB 3 sollten Berücksichtigung finden, denn erst diese erfordern bewertende oder reflektierende Aktivitäten. Tabelle 1 zeigt, welches kognitive Anspruchsniveau bei der Anwendung der mathematischen Kompetenzen zur Bearbeitung der Prüfungsaufgaben erforderlich ist.

| | Pflichtaufgaben | Wahlaufgaben |
|---|-----------------|--------------|
| unter Anforderungsbereich 1 | 1,5% | 0,1% |
| Anforderungsbereich 1 (Reproduzieren) | 73,2% | 50,1% |
| Anforderungsbereich 2 (Zusammenhänge herstellen) | 24,7% | 49,4% |
| Anforderungsbereich 3 (Verallgemeinern und Reflektieren) | 0,6% | 0,4% |

Tab. 1: Kognitives Anspruchsniveau der Prüfungsaufgaben zum Erwerb des MSA in Prozent
($N = 3530$)

Mit Blick auf die drei in den Bildungsstandards benannten Anforderungsbereiche fällt zunächst auf, dass es im Rahmen der Prüfungsverfahren zum Erwerb des MSA auch einzelne Aufgabenstellungen gibt, die *unterhalb von Anforderungsbereich 1* anzusiedeln sind – darunter fallen z. B. elementare Fertigkeiten (Routineverfahren) im Umgang mit Zahlen und Größen sowie einfachste geometrische Grundkenntnisse (Blum et al., 2006, S. 34). Insgesamt sind die meisten Aufgabenstellungen dem Anforderungsbereich 1 zuzuordnen. Aufgaben dieses Anforderungsbereichs erfordern *einfache Routinetätigkeiten*, z. B. die Wiedergabe und direkte Anwendung von grundlegenden Begriffen, Sätzen und Verfahren in einem abgegrenzten Gebiet und in einem bekannten Zusammenhang. Der Anteil entsprechender Aufgabenstellungen ist bei den Wahlaufgaben – zu Gunsten des Anforderungsbereichs 2 – signifikant niedriger als im Pflichtbereich ($\chi^2(1, 3479) = 220.44, p < .001, \phi = .25$); gleichwohl wird die Gesamtheit der Aufgaben vom Anforderungsbereich 1 mit einem Anteil von knapp zwei Dritteln dominiert. Kognitiv anspruchsvolle Aufgabenstellungen des Anforderungsbereichs 3, die das Bearbeiten komplexer Fragestellungen u. a. mit dem Ziel erfordern, zu eigenen Problemformulierungen, Reflexionen oder Wertungen zu gelangen, finden sich hingegen praktisch gar nicht.

Da sich bildungsstandardkonforme Zusammenstellungen von Aufgaben im Fach Mathematik durch eine inhaltliche Reichhaltigkeit sowie eine hinreichende Breite der geforderten Kompetenzen ohne eine einseitige Bevorzugung bestimmter Kompetenzen und Komplexitätsgrade auszeichnen sollen, um eine hinreichende Realisierung der drei Winterschen Grunderfahrungen erkennen zu lassen, kann als Zwischenfazit und mit Blick auf die zuvor betrachteten Qualitätsmerkmale festgehalten werden, dass die tatsächliche Aufgabenpraxis nur eine geringe Passung mit den Vorgaben der Bildungsstandards aufweist. Zahlreiche Aspekte einer postulierten kompetenzorientierten Aufgabenzusammenstellung werden nur unzureichend erfüllt; dies gilt insbesondere hinsichtlich der Dominanz des symbolisch/technisch/formalen Arbeitens, die hier nicht im erwarteten Maße durch die Realisierung der anderen Kompetenzen relativiert wird. Vor diesem Hintergrund muss die Implementation der Bildungsstandards in den Prüfungsverfahren zum Erwerb des MSA zum gegenwärtigen Stand mit Blick auf das Spektrum der Aufgaben insgesamt als *wenig erfolgreich* bezeichnet werden.

3.2 Die vergleichsorientierte Perspektive

Die zuvor skizzierten Befunde gelten für die überwiegende Mehrheit aller im Rahmen dieser Studie untersuchten Prüfungsaufgaben. Mit Blick auf die Verteilung der Kompetenzen, Leitideen und Anforderungsbereiche zeigen sich allerdings länderspezifisch teilweise beträchtliche Unterschiede, wie Abbildung 3 anhand ausgewählter Länderbeispiele exemplarisch für die Verteilung der Leitideen verdeutlicht.

Die Prüfungsaufgaben der hier exemplarisch dargestellten Bundesländer unterscheiden sich hinsichtlich ihrer inhaltlichen Breite sehr deutlich: In den Bundesländern Bayern und Thüringen fallen die dominanten Anteile von Aufgaben zu L4 auf. Diese sind je-

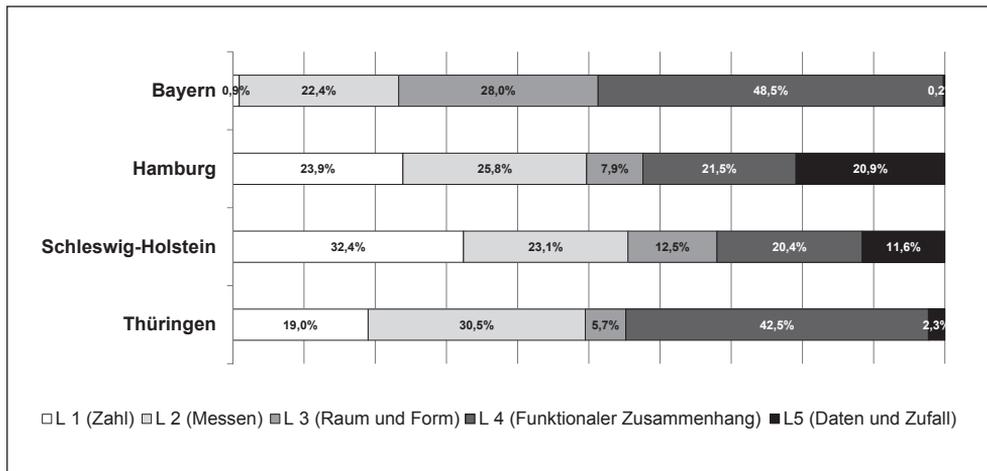


Abb. 3: Verteilung der Leitideen in Prüfungsaufgaben zum Erwerb des MSA in ausgewählten Bundesländern in Prozent (N = 1165); Grundlage ist jeweils die Gesamtheit aller Pflicht- und Wahlaufgaben

doch gut damit erklärbar, dass in beiden Bundesländern vorwiegend curriculare Inhalte der Klasse 10 Grundlage der Prüfungsaufgaben sind. Während in Bayern in den MSA-Abschlussprüfungen keine Aufgaben zu L5 gestellt werden, sind dies in Thüringen immerhin wenige. In Bayern fällt zudem auf, dass Aufgaben zu L1 praktisch nicht vorkommen, und es somit nur in Ausnahmefällen Aufgaben z. B. zur Prozentrechnung bzw. keine zur Kombinatorik gibt. Demgegenüber fällt in den Ländern Hamburg und Schleswig-Holstein auf – beide legen Inhalte der Klassenstufen 5 bis 10 zu Grunde –, dass nicht nur L4 eine geringere und L1 eine bedeutendere Rolle spielen, sondern dass insgesamt die Aufgabenzusammenstellungen hinsichtlich der realisierten Leitideen deutlich ausgewogener und in diesem Sinne folglich auch deutlicher bildungsstandardkonform sind. Der in Schleswig-Holstein hohe Anteil von Aufgaben zu L1 ist mit vielen ‚kurzen‘ Pflichtaufgaben zu dieser Leitidee erklärbar.

Gleichwohl zeigen vertiefende Analysen mit Blick auf die länderspezifische curriculare Grundlage der Prüfungen (s. o.), dass es kaum grundsätzliche, signifikante Unterschiede zwischen jenen Ländern gibt, welche die Inhalte der *gesamten* Sekundarstufe I prüfen, und jenen Ländern, die *nur* auf die Inhalte der Abschlussklasse(n) fokussieren. Vielmehr zeigen die Befunde erhebliche Divergenzen zwischen *allen* Ländern, und dies auch mit Blick auf die Inhalts- und Anspruchsdimension der Aufgaben. Insgesamt können *keine* spezifischen Ländertypen oder Vergleichsgruppen identifiziert werden, die in der Mehrheit der Dimensionen der Aufgabengestaltung (Leitideen, Kompetenzen, Anforderungsbereiche) Kongruenzen aufweisen. Gleichwohl ist der Deckungsgrad zwischen den Vorgaben der Bildungsstandards und der tatsächlichen Aufgabengestaltung mit Blick auf einzelne Dimensionen offenbar länderspezifisch unterschiedlich stark ausgeprägt, was anschlussfähig an die Befunde von Köller (2009) zum Stand der Umset-

zung der Bildungsstandards in den länderspezifischen Lehrplänen ist. Diesbezüglich scheinen weitere Analysen notwendig.

Es kann festgehalten werden: Die Prüfungsaufgaben im Fach Mathematik sind durch eine erhebliche länderspezifische Heterogenität gekennzeichnet – ob und inwieweit durch die Länderspezifika noch von *vergleichbaren* (gleichwertigen) Prüfungsanforderungen zum Erwerb des MSA gesprochen werden kann, muss zumindest kritisch hinterfragt werden.

3.3 Die entwicklungsorientierte Perspektive

Im Rekurs auf die Implementationsforschung (s. o.) ist davon auszugehen, dass kurz nach der Einführung der Bildungsstandards noch keine merklichen Effekte auf die Aufgabengestaltung zu erwarten sind. Jedoch kann angenommen werden, dass durch die zeitliche Distanz zum Implementationsbeginn eine Weiterentwicklung der Prüfungsaufgabenkultur hin zu einer stärker standardbezogenen Aufgabengestaltung stattfindet. Die retrospektive Längsschnittstudie (2007-2011) ermöglicht es zu prüfen, inwiefern die intendierte Weiterentwicklung im Sinne einer Qualitätssteigerung stattfindet. Gleichwohl ist darauf hinzuweisen, dass der hier betrachtete Zeitraum von fünf Jahren nur wenig umfangreich ist und die Befunde daher nur erste Hinweise auf mögliche Entwicklungstrends geben können.

Auf Grundlage der retrospektiven Längsschnittstudie lässt sich jedoch länderübergreifend feststellen, dass eine Implementation der Bildungsstandards kaum geschehen ist – vielmehr werden über den gesamten Zeitraum hinweg vergleichbare Prüfungsaufgaben innerhalb der 15 Länder gestellt. Tabelle 2 gibt einen Überblick über die Verteilung der Kompetenzen, Leitideen und Anspruchsniveaus in den Jahren 2007 und 2011; auf Grund der geringen Unterschiede, die sich zwischen den Pflicht- und Wahlaufgaben gezeigt haben (s. o.), werden die Befunde für die Gesamtheit der Aufgaben dargestellt:

Insgesamt, so lässt sich die Tabelle zusammenfassen, zeigen sich – mit zwei Ausnahmen – keine signifikanten Unterschiede zwischen beiden Analysezeitpunkten; dort, wo statistisch bedeutsame Divergenzen bestehen, sind diese nur von trivialer Größenordnung. Gleichwohl kann die beinahe Verdopplung an Aufgabenstellungen zur Leitidee *Daten und Zufall* als positive Tendenz bewertet werden, die jedoch – so zeigen vertiefende Analysen – nur auf die Entwicklung in einzelnen Ländern (z. B. Baden-Württemberg, Berlin) zurückzuführen ist und in ihrer Einschätzung der Unterscheidung bedarf, inwieweit *beide* thematischen Bereiche dieser Leitidee – also *Daten und Zufall* – adäquat repräsentiert sind.

Es kann zusammengefasst werden: Alle Länder weisen – trotz länderspezifischer Weiterentwicklungstendenzen mit Blick auf bestimmte Einzelaspekte – von Jahr zu Jahr eine erhebliche Konstanz in der Gestaltung der Prüfungsaufgaben auf. Von einer Qualitätssteigerung in den Abschlussprüfungsverfahren im Sinne einer *Weiterentwicklung*, die mit der Einführung der Bildungsstandards intendiert wurde, kann damit *nicht* gesprochen werden.

| | 2007 | 2011 | χ^2 | Sig. | Cramer's V |
|--|-------|-------|----------|-------|------------|
| Inhaltsbezogene Veränderungen | | | | | |
| L1 (Zahl) | 16,6% | 16,5% | <1 | n. s. | |
| L2 (Messen) | 24,5% | 21,5% | <1 | n. s. | |
| L3 (Raum und Form) | 13,2% | 12,5% | <1 | n. s. | |
| L4 (Funktionaler Zusammenhang) | 36,7% | 32,2% | <1 | n. s. | |
| L5 (Daten und Zufall) | 8,7% | 16,9% | 23.52 | <.001 | .13* |
| Prozessbezogene Veränderungen | | | | | |
| K1 (Argumentieren) | 11,2% | 14,8% | <1 | n. s. | |
| K2 (Problemlösen) | 44,2% | 36,8% | <1 | n. s. | |
| K3 (Modellieren) | 56,2% | 58,8% | <1 | n. s. | |
| K4 (Darstellungen verwenden) | 64,9% | 61,2% | <1 | n. s. | |
| K5 (technisches Arbeiten) | 96,3% | 93,1% | <1 | n. s. | |
| K6 (Kommunizieren) | 70,8% | 74,1% | 27.16 | <.001 | .14* |
| Anspruchsbezogene Veränderungen | | | | | |
| < AFB I | 0,1% | 1,4% | <1 | n. s. | |
| AFB I (Reproduzieren) | 63,3% | 60,5% | <1 | n. s. | |
| AFB II (Zusammenhänge herstellen) | 36,1% | 37,1% | <1 | n. s. | |
| AFB III (Verallgemeinern & Reflektieren) | 0,4% | 1,0% | <1 | n. s. | |

* statistisch bedeutsamer Zusammenhang trivialer Größenordnung

Tab. 2: Inhalts-, Prozess- und Anspruchsdimension der Prüfungsaufgabenzusammenstellungen in den Jahren 2007 und 2011 in Prozent (N = 3530)

4. Zusammenfassung und Diskussion

Der Beitrag stellt die Befunde einer Analyse von Prüfungsaufgaben im Fach Mathematik (N = 3530) aus allen 15 Bundesländern mit zentralen Abschlussprüfungen zum Erwerb des MSA über einen Zeitraum von fünf Jahren (2007-2011) dar, wobei mit Hilfe eines Kategoriensystems auf Basis der Bildungsstandards empirisch überprüft wurde, ob und inwieweit Bildungsstandards hinsichtlich der Qualität und Vergleichbarkeit von Prüfungsanforderungen zum Erwerb des MSA steuerungswirksam sind.

Insgesamt zeigt sich für den gesamten Analysezeitraum, dass die Prüfungsaufgaben – bei einer hohen jährlichen Konstanz hinsichtlich der länderspezifischen Aufgabenzusammenstellung – mehrheitlich nur eine geringe Passung mit den Vorgaben der

Bildungsstandards aufweisen, was anschlussfähig an bereits vorliegende Befunde aus der fachdidaktischen Forschung (Neubrand & Neubrand, 2010; Roppelt et al., 2008) ist und die Annahmen der Implementationsforschung (Gräsel, 2010; Gräsel & Parchmann, 2004; Prenzel, 2010) sowie neue Steuerungsansätze (Altrichter & Maag Merki, 2010; Altrichter, Brüsemeister & Wissinger, 2007) bestätigt. Die in den Bildungsstandards geforderte Umsetzung der drei Winterschen Grunderfahrungen, die ebenfalls geforderte inhaltliche Reichhaltigkeit der Prüfungsaufgaben sowie eine hinreichende Breite der geforderten Kompetenzen ohne eine einseitige Bevorzugung bestimmter Kompetenzen und Komplexitätsgrade bei der Aufgabengestaltung konnten im Rahmen der durchgeführten Aufgabenanalysen nicht bestätigt werden. Während das symbolisch/technisch/formale Arbeiten zumindest einen Beitrag zur Realisierung der Grunderfahrung der Strukturorientierung leistet, fehlt erkennbar ein ebenfalls notwendiger substantieller Beitrag des Argumentierens, ist es doch gerade diese Kompetenz, die für das Fach Mathematik charakteristische Tätigkeiten wie z. B. Begründen oder folgerichtiges Schließen umfasst. Auch die Grunderfahrung der Problemorientierung – unter dieser lassen sich insbesondere die Kompetenzen Problemlösen und (auch) Kommunizieren subsumieren – scheint nicht im erforderlichen Maße realisiert, denn man würde wohl insgesamt mehr Aufgaben mit Problemcharakter erwarten, speziell im Bereich der Pflichtaufgaben. Sicherlich kann die Grunderfahrung der Anwendungsorientierung noch am ehesten als realisiert angesehen werden, legt man hier vor allem die drei Kompetenzen Modellieren, Darstellungen und (auch) Kommunizieren zu Grunde. Um einen umfassenderen Beitrag zur Realisierung aller drei Grunderfahrungen zu leisten, scheint daher insbesondere eine Erhöhung des Anteils des Argumentierens geboten, aber auch andere Kompetenzen sollten vermehrt vorkommen, um die einseitige Orientierung an Verfahren (K5) zu reduzieren und gleichzeitig ein ausgewogeneres Mathematikbild in den Prüfungsaufgaben abzubilden – gleichermaßen in Pflicht- wie in Wahlaufgaben.

Es stellt sich die Frage nach möglichen Ursachen für die so charakterisierte Aufgabekultur der MSA-Prüfungen. Dabei wird nicht zuletzt an den je nach Bundesland sehr unterschiedlichen Anteilen der einzelnen Leitideen deutlich, dass es offenkundig regional unterschiedliche Deutungen der Realisierung der Bildungsstandards gibt. Dies zeigt sich auch daran, dass die MSA-Prüfungen praktisch keine Aufgaben enthalten, die dem höchsten Anforderungsbereich zuzuordnen sind, weshalb insgesamt Aufgaben fehlen, die reflektierende, verallgemeinernde oder bewertende Anforderungen zum Gegenstand haben. Möglicherweise ist das Fehlen derartiger Aufgaben auch auf ihre empirische Schwierigkeit zurückzuführen, weshalb das Stellen solcher Aufgaben in Prüfungen mit Zertifizierungsfunktion als riskant eingestuft wird, gerade deshalb, weil keine umfassenden Erprobungen der Aufgaben im Vorfeld stattfinden; insbesondere eine solche qualitätssichernde Maßnahme erscheint somit wünschenswert. Andere Erklärungen könnten darin bestehen, dass die Anspruchsdimension der Bildungsstandards nicht hinreichend bewusst oder möglicherweise nicht hinreichend in den Vorgaben der Kommissionen, die die Prüfungsaufgaben entwickeln, verankert ist. Gerade das Fehlen des höchsten Anforderungsbereichs lässt Abstimmungsbedarf der Aufgaben der MSA-Prüfungen mit den standardbasierten Aufgaben der Lernstandserhebungen erkennen. Dieser Abstim-

mungsbedarf wird auch im Hinblick auf die je nach Bundesland sehr unterschiedliche Verteilung der Leitideen deutlich. Dabei ist die Frage, inwiefern die Aufgaben der Lernstandserhebungen sowie jene der MSA-Prüfungen sich in ähnlicher Weise über die drei Dimensionen der Bildungsstandards verteilen sollen, zwar nur aus normativer Sicht zu diskutieren, doch wirkt ihre Beantwortung unmittelbar auf die Aufgabenkultur im Unterricht sowie auf jene in Klassenarbeiten zurück. In diesem Zusammenhang scheint mit Blick auf die Ergebnisse der Implementationsforschung denkbar, dass sich die unausgewogene Verteilung der Leitideen, der Kompetenzen und der Anforderungsbereiche, die über die Jahre 2007 bis 2011 nahezu unverändert geblieben ist, auch mit der Aufgabenkultur im Unterricht erklären lässt, in die die Bildungsstandards nicht im erwarteten Maße Eingang gefunden haben. Auch eine Bestätigung dieser Vermutung bedarf jedoch weiterer Untersuchungen und sollte u. a. Ergebnisse aus Schulinspektionen einbeziehen, die den Implementationsgrad der Bildungsstandards auf Schulebene betreffen. Hier scheinen jedoch der Anspruch an eine Implementation der Bildungsstandards sowie die Erwartung verlässlicher Anforderungen der MSA-Prüfungsaufgaben in gewissem Widerspruch zu stehen. Gerade im Hinblick auf eine wünschbare stärkere Konvergenz der Prüfungsanforderungen der verschiedenen Bundesländer scheinen hier jährliche moderate Weiterentwicklungen wünschenswert, die jedoch zunächst verstärkte Implementationsbemühungen auf der Ebene einzelner Bundesländer voraussetzen, bevor in der Folge Konvergenzen der Anforderungen zwischen den Bundesländern erwartbar sind.

Auf Grund der dokumentierten Befunde zu den Qualitätsmerkmalen der Prüfungsaufgaben muss die Implementation der Bildungsstandards insgesamt als *wenig erfolgreich* bezeichnet werden – inwieweit so das anschlussfähige Wissen und Können der Absolventinnen und Absolventen im Sinne der ‚Ausbildungsreife‘ überprüft werden kann, muss zumindest kritisch hinterfragt werden und bedarf empirischer Klärung, insbesondere auch mit Blick auf den der Prüfung vorgelagerten Unterricht. Bedingt durch die Länderspezifika muss zudem die bundesweite Vergleichbarkeit der Prüfungsanforderungen zum Erwerb des MSA deutlich in Frage gestellt werden. Inwiefern auf Grundlage der hier dargelegten Ergebnisse noch von einer Gleichwertigkeit schulischer Ausbildung gesprochen werden kann und welche Konsequenzen sich daraus für die Akzeptanz des Zertifizierungs- und Berechtigungssystems seitens der Anschlussysteme ergeben, sollte in weiterführenden Forschungsarbeiten betrachtet werden.

Literatur

- Altrichter, H., Brüsemeister, T., & Wissinger, J. (Hrsg.) (2007). *Educational Governance. Handlungskoordination und Steuerung im Bildungssystem*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Altrichter, H., & Maag Merki, K. (2010). Steuerung der Entwicklung des Schulwesens. In Dies. (Hrsg.), *Handbuch Neue Steuerung im Schulsystem* (S. 15-39). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Altrichter, H., & Posch, P. (2007). Analyse erster Erfahrungen mit der Implementation von Bildungsstandards. *Erziehung und Unterricht*, 157(7-8), 654-671.

- Baumert, J., Cortina, K. S., & Leschinsky, A. (2008). Grundlegende Entwicklungen und Strukturprobleme im allgemeinbildenden Schulwesen. In K. S. Cortina, J. Baumert, A. Leschinsky, K. U. Mayer & L. Trommer (Hrsg.), *Das Bildungswesen in der Bundesrepublik Deutschland* (S. 53-130). Reinbek: Rowohlt.
- Baumert, J., & Watermann, R. (2000). Institutionelle und regionale Variabilität und die Sicherung gemeinsamer Standards in der gymnasialen Oberstufe. In J. Baumert, W. Bos & R. Lehmann (Hrsg.), *TIMSS/III. Dritte Internationale Mathematik- und Naturwissenschaftsstudie. Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn* (S. 317-372). Opladen: Leske + Budrich.
- Berner, E., Oelkers, J., & Reusser, K. (2008). Implementation von Bildungsstandards. Bedingungen des Gelingens (und Scheiterns) aus internationaler Sicht. *Zeitschrift für Pädagogik*, 53. Beiheft, 210-226.
- BLK = Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung (1998). *Gutachten zur Vorbereitung des Programms Steigerung der Effizienz des mathematisch-naturwissenschaftlichen Unterrichts, Heft 60*. Bonn: BLK.
- Blum, W., Drücke-Noe, C., Hartung, R., & Köller, O. (Hrsg.) (2006). *Bildungsstandards Mathematik: konkret. Sekundarstufe I: Aufgabenbeispiele, Unterrichts Anregungen, Fortbildungsideen*. Berlin: Cornelsen Skriptor.
- Bortz, J., Boehnke, K., & Lienert, G. A. (2008). *Verteilungsfreie Methoden in der Biostatistik*. Berlin: Springer.
- Böttcher, W., & Dicke, J. N. (2008). Implementation von Standards. Empirische Ergebnisse einer Umfrage bei Deutschlehrern. In W. Böttcher, W. Bos, H. Döbert & H. G. Holtappels (Hrsg.), *Bildungsmonitoring und Bildungscontrolling in nationaler und internationaler Perspektive* (S. 143-156). Münster: Waxmann.
- Böttcher, W., & Rürup, M. (2007). Föderale Struktur des Bildungswesens und Schulentwicklung. In J. van Buer & C. Wagner (Hrsg.), *Qualität von Schule. Ein kritisches Handbuch* (S. 153-166). Frankfurt a. M.: Peter Lang.
- Bromme, R., Seeger, F., & Steinbring, H. (1990). Aufgaben, Fehler und Aufgabensysteme. In Dies. (Hrsg.), *Aufgaben als Anforderungen an Lehrer und Schüler* (S. 1-30). Köln: Aulis Verlag Deubner.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2. Aufl.). Hillsdale: Lawrence Erlbaum Associates.
- DIHK (2012). *Ausbildung 2012. Ergebnisse einer DIHK-Online-Unternehmensbefragung*. www.dihk.de/ressourcen/downloads/dihk-ausbildungsumfrage-2012.pdf [26.06.2013].
- Drücke-Noe, C. (2012). Wer Kalküle kann, schafft eine Klassenarbeit. Stimmt das? *Beiträge zum Mathematikunterricht*, 1, 213-216.
- Gehrmann, A., Hericks, U., & Lüders, M. (Hrsg.) (2010). *Bildungsstandards und Kompetenzmodelle. Beiträge zu einer aktuellen Diskussion über Schule, Lehrerbildung und Unterricht*. Bad Heilbrunn: Klinkhardt.
- Gördel, B. (2008). Die Einführung der nationalen Bildungsstandards in drei Bundesländern – eine explorative Studie zu Implementierungsstrategien. In R. Langer (Hrsg.), *Warum tun die das? Governanceanalysen zum Steuerungshandeln in der Schulentwicklung* (S. 193-220). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Gräsel, C. (2010). Stichwort: Transfer und Transferforschung im Bildungsbereich. *Zeitschrift für Erziehungswissenschaft*, 13(1), 7-20.
- Gräsel, C., & Parchmann, I. (2004). Implementationsforschung – oder: der steinige Weg, Unterricht zu verändern. *Unterrichtswissenschaft*, 32(3), 196-214.
- Henn, H.-W., & Kaiser, G. (2001). Mathematik – Ein polarisierendes Schulfach. *Zeitschrift für Erziehungswissenschaft*, 4(3), 359-380.

- Hochuli, L. A. (2011). *Bildungsstandards: Der Blick von Experten auf die Entwicklung von Bildungsstandards in der Schweiz und die Folgen für die Akteurkonstellation im Schulsystem*. http://geb.uni-giessen.de/geb/volltexte/2011/8336/pdf/AnnenHochuliLuzia_2011_06_22.pdf [19.06.2012].
- Jordan, A., Ross, N., Krauss, S., Baumert, J., Blum, W., Neubrand, M., Löwen, K., Brunner, M., & Kunter, M. (2006). *Klassifikationsschema für Mathematikaufgaben: Dokumentation der Aufgabenkategorisierung im COACTIV-Projekt*. Berlin: MPIB.
- Klemm, K. (2004). Strukturfragen und kein Ende. In H. G. Holtappels, K. Klemm, H. Pfeifer, H.-G. Rolff & R. Schulz-Zander (Hrsg.), *Jahrbuch der Schulentwicklung, Band 13: Daten, Beispiele und Perspektiven* (S. 83-96). Weinheim/München: Juventa.
- 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*. Bonn/Berlin: BMBF.
- KMK = Kultusministerkonferenz (Hrsg.) (2003a). *Vereinbarung über Bildungsstandards für den Mittleren Schulabschluss (Jahrgangsstufe 10). Beschluss der Kultusministerkonferenz vom 04.12.2003*.
- KMK = Kultusministerkonferenz (2003b). *Bildungsstandards im Fach Mathematik für den Mittleren Schulabschluss. Beschluss vom 04.12.2003*. München: Luchterhand.
- KMK = Kultusministerkonferenz (2006). *Gesamtstrategie der Kultusministerkonferenz zum Bildungsmonitoring*. Bonn: Wolters Kluwer Deutschland GmbH, München.
- KMK = Kultusministerkonferenz (Hrsg.) (2009). *Vereinbarung über Schularten und Bildungsgänge in der Sekundarstufe I. Beschluss der Kultusministerkonferenz vom 03.12.1993 i. d. F. vom 09.10.2009*.
- Köller, O. (2009). Die institutionelle Umsetzung durch das IQB – Konzeption und Kritik. In R. Wernstedt & M. John-Ohnesorg (Hrsg.), *Bildungsstandards als Instrument schulischer Qualitätsentwicklung* (S. 28-32). Berlin: Friedrich-Ebert-Stiftung.
- Köller, O. (2011). Standardsetzung im Bildungssystem. In H. Reinders, H. Ditton, C. Gräsel & B. Gniewosz (Hrsg.), *Empirische Bildungsforschung. Strukturen und Methoden* (S. 179-192). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Köller, O., Baumert, J., Cortina, K. S., Trautwein, U., & Watermann, R. (2004). Öffnung von Bildungswegen in der Sekundarstufe II und die Wahrung von Standards. *Zeitschrift für Pädagogik*, 50(5), 679-700.
- Köller, O., Knigge, M., & Tesch, B. (Hrsg.) (2010). *Sprachliche Kompetenzen im Ländervergleich*. Münster: Waxmann.
- Kühn, S. M. (2013). Vergleichbarkeit beim Mittleren Schulabschluss? Ein Überblick über die Vielfalt schulstrukturell möglicher Bildungswege und Prüfungsverfahren in den deutschen Ländern. *Die Deutsche Schule*, 105(1), 87-101.
- Langer, I., & Schulz von Thun, F. (2007). *Messung komplexer Merkmale in Psychologie und Pädagogik. Ratingverfahren*. Münster: Waxmann.
- Leuders, T. (2004). Selbstständiges Lernen und Leistungsbewertung. *Der Mathematikunterricht*, 3, 63-79.
- Maag Merki, K. (2010). Theoretische und empirische Analysen der Effektivität von Bildungsstandards, standardbezogenen Lernstandserhebungen und zentralen Abschlussprüfungen. In H. Altrichter (Hrsg.), *Handbuch Neue Steuerung im Schulsystem* (S. 145-169). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Mayring, P. (2000). *Qualitative Inhaltsanalyse. Grundlagen und Techniken*. Weinheim: Deutscher Studienverlag.
- Neubrand, J., & Neubrand, M. (2010). *Mathematikdidaktische Analysen der zentralen Prüfungen 2008 in Mathematik am Ende der Klasse 10 in Mathematik* (unveröffentlichtes Manuskript). Vechta/Oldenburg.

- Neumann, M., Nagy, G., Trautwein, U., & Lüdtke, O. (2009). Vergleichbarkeit von Abiturleistungen. Leistungs- und Bewertungsunterschiede zwischen Hamburger und Baden-Württemberger Abiturienten und die Rolle zentraler Abiturprüfungen. *Zeitschrift für Erziehungswissenschaft*, 12(4), 691-714.
- Pant, H. A., Vock, M., Pöhlmann, C., & Köller, O. (2008a). Offenheit für Innovationen. Befunde aus einer Studie zur Rezeption der Bildungsstandards bei Lehrkräften und Zusammenhänge mit Schülerleistungen. *Zeitschrift für Pädagogik*, 54(6), 827-845.
- Pant, H. A., Vock, M., Pöhlmann, C., & Köller, O. (2008b). Eine modellbasierte Erfassung der Auseinandersetzung von Lehrkräften mit den länderübergreifenden Bildungsstandards. In E. M. Lankes (Hrsg.), *Pädagogische Professionalität als Gegenstand empirischer Forschung* (S. 245-260). Münster: Waxmann.
- Prenzel, M. (2010). Geheimnisvoller Transfer? Wie Forschung der Bildungspraxis nützen kann. *Zeitschrift für Erziehungswissenschaft*, 13(1), 21-37.
- Roppelt, A., Katzenbach, M., & Pant, H. A. (2010). *Bereitstellung eines Aufgabenpools für kompetenzbasierte Vergleichsarbeiten im Fach Mathematik in der 8. Jahrgangsstufe im Schuljahr 2009/2010 – Technischer Bericht*. Berlin: Institut zur Qualitätsentwicklung im Bildungswesen, Humboldt-Universität zu Berlin.
- Roppelt, A., Keller, K., Leiß, D., Drüke-Noe, C., & Blum, W. (2008). *Kommentierung der schriftlichen Prüfungsarbeit zum mittleren Schulabschluss 2007 im Fach Mathematik (Berlin) bzw. der Prüfungen am Ende der Jahrgangsstufe 10 Mathematik (Brandenburg) hinsichtlich ihrer Orientierung an den länderübergreifenden Bildungsstandards*. Berlin/Brandenburg: ISQ.
- Sacher, W. (2000). Tests und Klausuren in der Schule – Wie mache ich das? In S.-I. Beutel & W. Vollstädt (Hrsg.), *Leistung ermitteln und bewerten* (S. 63-74). Hamburg: Bergmann + Helbig.
- Schuchart, C. (2007). Schulabschluss und Ausbildungsberuf. Zur Bedeutung der schulartbezogenen Bildungsbiografie. *Zeitschrift für Erziehungswissenschaft*, 10(3), 381-398.
- Seidel, T. (2003). Videobasierte Kodierverfahren in der IPN Videostudie Physik – ein methodischer Überblick. In T. Seidel, M. Prenzel, R. Duit & M. Lehrke (Hrsg.), *Technischer Bericht zur Videostudie „Lehr-Lern-Prozesse im Physikunterricht“* (S. 99-111). Kiel: IPN.
- Sink, C. A., & Stroh, H. R. (2006). Practical Significance: The Use of Effect Sizes in School Counseling Research. *Professional School Counseling*, 9(5), 401-411.
- Steffens, U. (2007). Schulqualitätsdiskussion in Deutschland – ihre Entwicklung im Überblick. In J. van Buer & C. Wagner (Hrsg.), *Qualität von Schule. Ein kritisches Handbuch* (S. 21-51). Frankfurt a. M.: Peter Lang.
- Steffens, U. (2010). Lernstandserhebungen in den deutschen Ländern. *Schulverwaltung, Ausgabe Nordrhein-Westfalen*, 21(11), 290-293.
- Tesch, B. (2010). *Kompetenzorientierte Lernaufgaben im Fremdsprachenunterricht. Konzeptionelle Grundlagen und eine rekonstruktive Fallstudie zur Unterrichtspraxis*. Frankfurt a. M.: Peter Lang.
- Thompson, B. (2002). „Statistical“, „Practical“ and „Clinical“: How Many Kinds of Significance Do Counselors Need to Consider? *Journal of Counseling and Development*, 80(1), 64-71.
- Wacker, A., Maier, U., & Wissinger, J. (2012). Ergebnisorientierte Steuerung – Bildungspolitische Strategien und Verfahren zur Initiierung von Schul- und Unterrichtsreformen. In Dies. (Hrsg.), *Schul- und Unterrichtsreform durch ergebnisorientierte Steuerung. Empirische Befunde und forschungsmethodische Implikationen* (S. 9-34). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Winter, H. (1995). Mathematikunterricht und Allgemeinbildung. *Mitteilungen der Gesellschaft für Didaktik der Mathematik*, 61, 37-46.

- Wirtz, M., & Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität. Methoden zur Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Kategoriensystemen und Ratingskalen*. Göttingen: Hogrefe.
- Zeitler, S. (2010). *Bildungsstandards in der Schule. Eine rekonstruktive Studie zur Implementation der Bildungsstandards*. Frankfurt a. M.: Peter Lang.
- Zeitler, S., Köller, O., & Tesch, B. (2010). Bildungsstandards und ihre Implikation für Qualitätssicherung und Qualitätsentwicklung. In A. Gehrman, U. Hericks & M. Lüders (Hrsg.), *Bildungsstandards und Kompetenzmodelle. Beiträge zu einer aktuellen Diskussion über Schule, Lehrerbildung und Unterricht* (S. 23-26). Bad Heilbrunn: Klinkhardt.
- Zydati, W. (2007). Die Aufgabe von Aufgaben in den zentralen Prüfungen für den „Mittleren Schulabschluss“. Ein Plädoyer für die empirisch-quantitative Erforschung entwicklungsbezogener Lern- und Überprüfungsaufgaben im Englischunterricht. In H. J. Vollmer (Hrsg.), *Synergieeffekte in der Fremdsprachenforschung* (S. 299-312). Frankfurt a. M.: Peter Lang.

Abstract: In order to guarantee quality and comparability, all German Länder are asked to carry out statewide exit examinations for the acquisition of the intermediate leaving certificate on the basis of the educational standards. The authors present the findings of a longitudinal study which analyzed examination tasks in the field of mathematics ($N = 3530$) across all German *Länder* by means of a system of categories based on the educational standards. On the basis of theoretical assumptions posited by implementation research, the study investigates the level of concordance between the educational standards valid for the entire country and the actual assignment practice in the examinations carried out in the individual Länder with regard to the acquisition of the intermediate leaving certificate. It is shown that the examination tasks are highly unbalanced in their reproduction of the three dimensions of the educational standards (central themes, competencies, fields of requirements). Furthermore, the tasks are characterized by a considerable *Länder*-specific heterogeneity. The findings point to a discrepancy between the impact expected in education policy and the effects actually experienced, because the aim of implementing comparable examination requirements throughout the country has obviously hardly been fulfilled yet.

Keywords: Standards, Statewide Exit Exams, Implementation, Task Analysis, Mathematics

Anschrift der Autorinnen

Dr. Svenja Mareike Kühn, Universität Duisburg-Essen, Fakultät für Bildungswissenschaften, Institut für Pädagogik, AG Bildungsforschung, Universitätsstraße 11, 45117 Essen, Deutschland
E-Mail: svenja.kuehn@uni-due.de

Dr. Christina Drüke-Noe, Universität Kassel, Institut für Mathematik, Fachbereich 10 Mathematik und Naturwissenschaften, Heinrich-Plett-Straße 40, 34132 Kassel, Deutschland
E-Mail: drueke-noe@mathematik.uni-kassel.de