



Mathematikleistungen von Schülerinnen und Schülern der gymnasialen Oberstufe in Deutschland: Ein Review und eine Sekundäranalyse der Schulleistungsstudien seit 1995

Tobias Rolfes  · Anke Lindmeier · Aiso Heinze

Eingegangen: 20. April 2020 / Angenommen: 18. November 2020 / Online publiziert: 8. Dezember 2020
© Der/die Autor(en) 2020

Zusammenfassung Die Mathematikleistungen von Schülerinnen und Schülern der gymnasialen Oberstufe sind in den letzten zehn Jahren in Deutschland verstärkt in den Fokus der Bildungsforschung, der Hochschullehre und der Bildungspolitik gerückt. Der vorliegende Artikel hat das Ziel, die vorhandene empirische Befundlage genauer in den Blick zu nehmen und so zu einer Bestandsaufnahme beizutragen. Dazu wird im ersten Teil des Artikels ein systematisches Review über die seit 1995 durchgeführten Schulleistungsuntersuchungen zu Mathematikleistungen in der Oberstufe berichtet. Dabei wurde analysiert, welche Facetten mathematischer Fähigkeiten in diesen Schulleistungsstudien erfasst, welche Aspekte der übergreifenden Bildungsziele vertiefte Allgemeinbildung, Wissenschaftspropädeutik und Studierfähigkeit in den Schulleistungsstudien abgebildet und welche Niveaustufenmodelle zur kriterialen Einordnung der Schülerleistungen entwickelt wurden. Es zeigte sich, dass die verwendeten Instrumente jeweils nur einen besonderen Ausschnitt der mathematischen Fähigkeiten abbildeten und wichtige Aspekte (z. B. prozedural-technische Fertigkeiten) nahezu unberücksichtigt ließen. Auch wurde deutlich, dass ein kriterialer Vergleich der Ergebnisse aus den unterschiedlichen Schulleistungsstudien bisher kaum erfolgte. Deshalb wurden im zweiten Teil des Artikels die Daten und Ergebnisse der Schulleistungsuntersuchungen einer Sekundäranalyse unterzogen. Die Schülerleistungen wurden auf vergleichbare Skalen transformiert und anhand vorhandener Niveaustufenmodelle verglichen. Es zeigte sich, dass der größte Teil der

Zusatzmaterial online Zusätzliche Informationen sind in der Online-Version dieses Artikels (<https://doi.org/10.1007/s13138-020-00180-1>) enthalten.

T. Rolfes (✉) · A. Heinze
IPN – Leibniz-Institut für die Pädagogik der Naturwissenschaften und Mathematik, Kiel,
Deutschland
E-Mail: rolfes@leibniz-ipn.de

A. Lindmeier
Friedrich-Schiller-Universität Jena, Jena, Deutschland

Schülerschaft im Bereich der Allgemeinbildung lebensweltliche Problemstellungen auf der Basis der mathematischen Begriffe und Verfahren der Sekundarstufe I bewältigen konnte. Bezüglich der Wissenschaftspropädeutik und der Studierfähigkeit zeigte die Mehrheit deutliche Defizite, da sie mit fachspezifischen Lerninhalten der Oberstufe (z. B. Analysis, Analytische Geometrie) nur eingeschränkt umgehen konnte. Implikationen für die mathematikdidaktische Forschung zum Mathematikunterricht in der gymnasialen Oberstufe und seinen Bildungszielen werden diskutiert.

Schlüsselwörter Sekundarstufe II · Mathematikleistungen · Schulleistungsuntersuchungen · Bildungsstandards · Allgemeine Hochschulreife

Mathematical Achievements of Upper Secondary School Students in Germany: A Review and a Secondary Analysis of Large Scale Assessment Studies since 1995

Abstract Over the last ten years, students' mathematics achievements in upper secondary school have increasingly become the focus of educational research, university teaching, and educational policy in Germany. This article aims to take a closer look at the existing empirical evidence and, thus, to contribute to an overview. For this reason, the first part of the article reports a systematic review of large scale assessments on students' mathematics achievement in upper secondary school that were conducted since 1995. It was analyzed which facets of mathematical abilities are covered in these large scale assessments, which aspects of the overarching educational goals of in-depth general education, science propaedeutics and scholastic aptitude are reflected in school achievement studies, and which level models have been developed for criteria-based classification of student achievement. It was found that the instruments used each represented only a particular facet of mathematical abilities and almost ignored essential aspects (e.g., procedural-technical skills). It also became clear that a criteria-based comparison of the results from the different large scale assessments has hardly been carried out so far. Therefore, in the second part of the article, the large scale assessments' data and results were subjected to a secondary analysis. Student performances were transformed into comparable scales and compared using existing level models. It was shown that the majority of the students were able to cope with real-life problems based on the mathematical concepts and procedures of the intermediate secondary school leaving certificate. With regard to science propaedeutics and the scholastic aptitude, the majority showed apparent deficits, as they were only able to deal with subject-specific learning contents of upper secondary school (e.g., calculus, analytical geometry) to a limited extent. Implications for mathematics education research on mathematics teaching in upper secondary school and its educational goals are discussed.

Keywords Upper secondary school · Mathematical achievement · Large scale assessments · Educational standards · General university entrance qualification

1 Einleitung

Klagen über das Leistungsniveau der Abiturientinnen und Abiturienten im mathematisch-naturwissenschaftlichen Bereich haben in Deutschland eine gewisse Tradition. Bereits in den 1950er-Jahren bemängelten die Hochschulen schlechte Mathematikkenntnisse bei Studienanfängerinnen und Studienanfängern (Steiner 1984). Anfang der 1980er-Jahre konstatierten mathematisch-naturwissenschaftliche Fachverbände „einen Verfall der Kenntnisse und Fähigkeiten in den mathematischen, naturwissenschaftlichen und technischen Fächern“ (DMV et al. 1982, S. 25), und auch in den 1990er-Jahren stellte die mangelnde „Sicherheit im Umgang mit mathematischen Symbolen und Modellen“ (KMK 1995, S. 24) ein häufig genanntes Monitum dar.

Ähnliche Äußerungen sind in den letzten Jahren wieder verstärkt zu vernehmen. Dabei wird bemängelt, dass die Studienanfängerinnen und Studienanfänger nicht mehr die notwendigen mathematischen Voraussetzungen mitbringen, um erfolgreich ein MINT-Studium zu absolvieren. Besonders bestünden Defizite in grundlegenden prozedural-technischen Fertigkeiten wie Bruchrechnen, Termumformen und Gleichungslösen. Vor allem diese Fertigkeiten werden jedoch von Hochschuldozierenden als essenzielle Eingangsvoraussetzungen für ein MINT-Studium angesehen (Koepf und Kramer 2015; Neumann et al. 2017).

Angesichts der in der Öffentlichkeit intensiv geführten Diskussionen über Ausgestaltungsfragen der gymnasialen Oberstufe (z. B. G8 vs. G9, zentrale vs. dezentrale Abiturprüfungen, Kurssystem vs. Profiloberstufe, Anzahl der Abiturprüfungsfächer) und der Dynamik ihrer Entwicklungen (steigende Abiturientenzahlen, „Noteninflation“) verwundert es, dass sich die empirische Datenlage zur gymnasialen Oberstufe als sehr überschaubar darstellt und ein Bildungsmonitoring für die gymnasiale Oberstufe aktuell nicht geplant ist (Stanat et al. 2016; Neumann und Trautwein 2019). Gerade wegen der limitierten Datenlage für die gymnasiale Oberstufe erscheint uns ein genauerer Blick auf die vorhandenen Ergebnisse angezeigt. Was ist jenseits von anekdotischen Evidenzen über Mathematikleistungen in der gymnasialen Oberstufe in Deutschland bekannt? Welche mathematischen Leistungsniveaus erreichen Abiturientinnen und Abiturienten? Wie entwickeln sich diese Mathematikleistungen im Verlauf der Oberstufe und über Schülergenerationen hinweg? Wie im Folgenden gezeigt wird, sind die vorhandenen Forschungsergebnisse bisher nicht systematisch aufeinander bezogen, zum Teil sehr beschränkt und aus mathematikdidaktischer Perspektive als unzureichend einzuschätzen.

2 Das Fach Mathematik in der gymnasialen Oberstufe in Deutschland

Die Entwicklung der Mathematikleistungen von Schülerinnen und Schülern vollzieht sich jeweils unter den gegebenen schulischen Rahmenbedingungen. Daher erscheint es hilfreich, zunächst grundlegende Entwicklungslinien bezüglich der Bildungsziele und der Ausgestaltung der gymnasialen Oberstufe der letzten Jahrzehnte zu rekapitulieren, um die empirischen Ergebnisse aus unterschiedlichen Zeitpunkten einordnen zu können.

2.1 Übergeordnete Bildungsziele der gymnasialen Oberstufe

Die tradierten übergeordneten Bildungsziele der gymnasialen Oberstufe sind in der Bundesrepublik durch eine Trias an Zielen geprägt: „Vertiefte Allgemeinbildung [...], Wissenschaftspropädeutik und Sicherung von Studierfähigkeit“ (KMK 1995, S. 8). Dabei soll die Allgemeinbildung die Voraussetzungen schaffen für eine „verständige, kritische und selbstdistanzierte Teilhabe am gesellschaftlichen und öffentlichen Leben“ (KMK 1995, S. 72). Die vertiefte Allgemeinbildung umfasst Personal-, Sozial- und Fachkompetenz und beispielsweise auch die Fähigkeit zur Berufs- und Studienwahl (KMK 1995). Eine fächerübergreifende Begriffsbestimmung, wie aus fachinhaltlicher Perspektive eine „Vertiefung“ der Allgemeinbildung für die gymnasiale Oberstufe zu verstehen ist, gibt es bisher nicht (Huber 2009). Baptist und Winter (2001) sehen für die Mathematik eine qualitative und quantitative Vertiefung der drei Grunderfahrungen nach Winter (1995), die bereits in der Sekundarstufe I angebahnt werden, als zentrale allgemeinbildende Aufgabe des Mathematikunterrichts in der Oberstufe. Eine grundlegend andere Perspektive zur Vertiefung der Allgemeinbildung wird von Fischer (2012) eingenommen, der die Kommunikationsfähigkeit mit Expertinnen und Experten und die Entscheidungsfähigkeit als grundlegendes allgemeinbildendes Unterrichtsziel des Mathematikunterrichts in der Oberstufe sieht. Fischers theoretischer Ansatz hat in den curricularen Vorgaben in Deutschland aber bisher wenig Resonanz gefunden. Dagegen wird in den gegenwärtig gültigen *Bildungsstandards für das Fach Mathematik* explizit auf die Grunderfahrungen nach Winter Bezug genommen (KMK 2015).

Das Anliegen des zweiten Aspektes der Trias, der Wissenschaftspropädeutik, beinhaltet nicht das wissenschaftliche Arbeiten selbst oder die Vorwegnahme von Inhalten des Hochschulstudiums, sondern die Einführung in die Denk- und Arbeitsweisen der Wissenschaft (Hahn 2013; KMK 1995). Wissenschaftspropädeutisches Lernen inkludiert auch die Reflexion über Möglichkeiten und Grenzen von Wissenschaft (KMK 1995). Nach Huber (2009) sollte Wissenschaftspropädeutik nicht ausschließlich im studienvorbereitenden Sinne verstanden werden, sondern die Schülerinnen und Schüler sollten auch in einem allgemeinbildenden Sinne auf „Leben und Arbeiten in der verwissenschaftlichen Gesellschaft“ (S. 108) vorbereitet werden.

Die Zielsetzung der Studierfähigkeit stellt ein komplexes Konstrukt dar und adressiert nicht nur kognitive Fähigkeiten, sondern umfasst auch motivationale, ethische und soziale Dimensionen (KMK 1995). Huber (2009) definiert Studierfähigkeit als „Ensemble von Fähigkeiten [...], ein Studium erfolgreich zu beginnen, durchzuführen und abzuschließen“ (S. 108). Während in der wissenschaftlichen Expertise aus dem Jahre 1995 der Aspekt Studierfähigkeit relativ breit verstanden und darunter auch eine fachspezifische Vertiefung durch Leistungskurse subsumiert wird (KMK 1995), definiert die aktuelle *Vereinbarung zur Gestaltung der gymnasialen Oberstufe* die Trias als „vertiefte Allgemeinbildung, allgemeine Studierfähigkeit sowie wissenschaftspropädeutische Bildung“ (KMK 2018, S. 5). Somit wird betont, dass bei der Sicherung der Studierfähigkeit eine *allgemeine* und nicht eine fachspezifische Studierfähigkeit im Zentrum steht.

2.2 Organisatorische und curriculare Ausgestaltung der gymnasialen Oberstufe

Die gleichzeitige Berücksichtigung der Aspekte der Trias ist ein schwieriger Balanceakt (KMK 1995), der im Laufe der Zeit durch organisatorische und curriculare Veränderungen jeweils Neujustierungen erfuhr. So zielte die Einführung des Kurssystems mit flexiblen Wahlmöglichkeiten in der reformierten Oberstufe (KMK 1972) auf die Stärkung der Aspekte Wissenschaftspropädeutik und Studierfähigkeit. Vor allem die fünf- oder sechsständigen Leistungskurse sollten Raum für wissenschaftspropädeutisches Arbeiten bieten.

Die großen Wahlmöglichkeiten in der reformierten Oberstufe stießen jedoch auf Kritik. Durch engere Beleg- und Einbringungsverpflichtungen, die Einführung von vierstündigen Leistungskursen und des fünften Abiturfaches wurde in den 1990er-Jahre wieder eine stärkere Kanonisierung zugunsten der Kernfächer (Mathematik, Deutsch, Fremdsprache) forciert (z. B. Husumer Beschlüsse, vgl. Übersicht in Neumann 2010). Diese Reformen waren unter anderem von dem Ziel getragen, einer zu frühen Spezialisierung entgegenzuwirken und den allgemeinbildenden Anspruch durch Standardisierung von Bildungsinhalten zu stärken. Außerdem sollte durch ein gefestigtes Wissen und Können in den Kernfächern die Studierfähigkeit erhöht werden.

Darüber hinaus sollten curriculare Anpassungen für den Mathematikunterricht der gymnasialen Oberstufe den Aspekt der vertieften Allgemeinbildung stärken. So betonen die 2002 überarbeiteten *Einheitlichen Prüfungsanforderungen in der Abiturprüfung* für das Fach Mathematik insbesondere den anwendungsbezogenen Charakter von Mathematik (KMK 2002). Das mathematische Modellieren zur Lösung realitätsnaher Probleme stellte einen zentralen Aspekt des festgelegten Kompetenzspektrums dar. Dies markiert bereits den Beginn der Kompetenzorientierung mit seinem Bestreben, verstärkt Anwendungs- und Verständnisorientierung im Mathematikunterricht zu etablieren. Es mündete in Fortführung der Bildungsstandards für den Mittleren Schulabschluss (KMK 2004) schließlich in die *Bildungsstandards im Fach Mathematik für die Allgemeine Hochschulreife* (KMK 2015). Das Ziel, dass in der gymnasialen Oberstufe fachspezifische Inhalte der Oberstufenmathematik in lebensweltlichen Problemstellungen verständnisvoll angewendet werden sollen, wird hier in noch stärkerem Maße als bereits in den Prüfungsanforderungen von 2002 betont. So ist zum Beispiel der Integralbegriff nicht nur im Kontext von Flächenberechnungen zu thematisieren, sondern „auch in seiner anwendungsbezogenen Bedeutung als rekonstruierter Bestand [...] verpflichtend [Hervorhebung der Verf.] vorgesehen“ (Stanat et al. 2016, S. 43). In den Bildungsstandards wird explizit festgestellt, dass „Aufgaben mit Anwendungen aus der Lebenswelt [...] die gleiche Wichtigkeit und Wertigkeit wie innermathematische Aufgaben [haben]“ (KMK 2015, S. 12). In Fortführung der Bildungsstandards für den Mittleren Schulabschluss (KMK 2004) definieren die Bildungsstandards für die Allgemeine Hochschulreife mathematische Kompetenzen prozessbezogen anhand von sechs allgemeinen mathematischen Kompetenzen (K1-K6) und inhaltsbezogen entlang von fünf Leitideen (L1-L5). Dazu repräsentieren jeweils drei Anforderungsbereiche „unterschiedliche

kognitive Ansprüche von kompetenzbezogenen mathematischen Aktivitäten“ (KMK 2015, S. 12).

Zusammenfassend kann festgestellt werden, dass die Trias vertiefte Allgemeinbildung, (allgemeine) Studierfähigkeit und Wissenschaftspropädeutik durchgehend tragende Ziele der gymnasialen Oberstufe waren und noch immer sind. Eine Ausbalancierung der drei Aspekte wurde zumeist *implizit* durch konkrete organisatorische und curriculare Veränderungen vorgenommen. Dieses ist auch damit begründet, dass die Trias bisher eher einen Präambelcharakter einnimmt. Eine Operationalisierung, welche Lerninhalte für den Mathematikunterricht aus der Trias abgeleitet werden können, ist aus bildungstheoretischer Perspektive bisher erst in Ansätzen erfolgt. Zudem sind die drei Aspekte der Trias miteinander verwoben, sodass eine Zuordnung einzelner Maßnahmen oder mathematischer Inhalte zu einem Aspekt zumeist nur mit Einschränkungen möglich ist. Trotz dieser Schwierigkeiten ziehen wir die Trias in diesem Artikel als grobes Schema zur Einordnung der in den Schulleistungsstudien erfassten Mathematikleistungen heran, da sie rahmengebend für die Ziele der Oberstufe ist.

3 Forschungsfragen

Angesichts der skizzierten Entwicklungen des Faches Mathematik in der Sekundarstufe II stellt sich die Frage, inwieweit es der gymnasialen Oberstufe gelingt, die anvisierten Bildungsziele zu vermitteln und dem Anspruch der Trias gerecht zu werden. Als Annäherung an diese Frage erscheint eine Bestandsaufnahme bisheriger Ergebnisse sinnvoll, da bisher kein bundesweites Bildungsmonitoring für die Sekundarstufe II etabliert oder vorgesehen ist. So wurden seit 1995 in der Bundesrepublik und in einzelnen Bundesländern Schulleistungsstudien (SLS) durchgeführt, unter anderem um institutionelle Rahmenbedingungen oder Veränderungen zu evaluieren. Eine Synthese der SLS mit speziellem Fokus auf die Mathematikleistungen ist allerdings bisher nicht erfolgt und ist aufgrund der unterschiedlichen Auswertungen und Ergebnisdarstellungen selbst im Falle gleicher Testinstrumente auch nicht trivial. Diese Bestandsaufnahme ist für die Beurteilung hilfreich, welche (möglicherweise limitierten) Forschungserkenntnisse überhaupt vorhanden sind und welche zukünftigen Forschungsbedarfe zum Mathematikunterricht in der gymnasialen Oberstufe bestehen.

Das Vorgehen bei unserem Review und der darauf aufbauenden Sekundäranalyse von Daten wird von den folgenden Fragestellungen geleitet:

1. Welche SLS haben seit 1995 in Deutschland in der gymnasialen Oberstufe stattgefunden?
2. Wie wurde die Mathematikleistung in den SLS erfasst?
 - a. Welche Konstrukte und Instrumente wurden für die Messung der Mathematikleistungen verwendet?
 - b. In welchem Verhältnis stehen diese Konstrukte und Instrumente zur Trias der Ziele der gymnasialen Oberstufe?

- c. Welche kriterialen Niveaustufenmodelle wurden für die Konstrukte und Instrumente entwickelt?
3. Was ist für die jeweiligen Konstrukte und Instrumente der SLS über das Leistungsniveau und die Leistungsentwicklung in Mathematik von Schülerinnen und Schülern in der gymnasialen Oberstufe bekannt?
 - a. Welche Leistungsniveaus wurden in den jeweiligen Konstrukten und Instrumenten am Ende der gymnasialen Oberstufe erreicht?
 - b. Wie entwickelten sich die Mathematikleistungen in den jeweiligen Konstrukten und Instrumenten innerhalb der gymnasialen Oberstufe?
 - c. Wie veränderten sich die Leistungsniveaus von Abschlusskohorten in den jeweiligen Konstrukten und Instrumenten seit 1995?

Die auf der Grundlage der Forschungsfragen 2 und 3 gewonnenen Erkenntnisse werden schließlich in der Diskussion im Hinblick auf die Bildungsziele des Mathematikunterrichts in der gymnasialen Oberstufe interpretiert und bewertet. Außerdem wird erörtert, welche Forschungsdesiderate für die mathematikdidaktische Forschung sich aus den vorhandenen (bzw. fehlenden) empirischen Ergebnissen ableiten lassen.

4 Review der Schulleistungstudien (Forschungsfragen 1 und 2)

4.1 Methode

Für diesen Überblicksartikel wurden SLS mit Erhebungszeitpunkten ab 1995 berücksichtigt, die Mathematikleistungen von Schülerinnen und Schülern der gymnasialen Oberstufe in Deutschland erfasst haben (z. B. an Allgemeinbildenden oder Beruflichen Gymnasien, Schulen mit mehreren Bildungsgängen wie Gesamtschulen). Darüber hinaus sollten die SLS den Anspruch haben, repräsentative Stichproben für die entsprechende Schülerpopulation in Deutschland oder in einzelnen Bundesländern zu untersuchen.

Ergebnisse von SLS werden und wurden in unterschiedlicher Form und an unterschiedlichen Stellen veröffentlicht (z. B. in Sammelbänden, Journalartikeln, technischen Berichten). Um nicht wesentliche Ergebnisse unberücksichtigt zu lassen, war daher eine gezielte Suche nach Publikationen erforderlich, die auf Daten der SLS basieren und wesentliche Informationen für die formulierten Fragestellungen enthalten. Als Kriterium der Qualitätssicherung wurden nur Publikationen berücksichtigt, die entweder in einer begutachteten wissenschaftlichen Zeitschrift oder als Buch- oder Onlineveröffentlichung eines Wissenschaftsverlags (ausgenommen Tagungsbände), einer staatlichen Bildungsadministration oder eines Forschungsinstituts erschienen sind.

Die SLS und die relevanten Publikationen wurden identifiziert, indem zuerst systematisch die Verzeichnisse der Wissenschaftsverlage (Springer, Waxmann, Franzbecker und WTM), die wissenschaftlichen Zeitschriften (Journal für Mathematik-Didaktik, ZDM, Zeitschrift für Erziehungswissenschaften) und die Publikationslisten der Institutionen der staatlichen Bildungsadministration und der Forschungs-

stitute nach entsprechenden Veröffentlichungen durchsucht wurden. Anschließend wurde bei Google Scholar nach Einträgen¹ gesucht, die den Kriterien entsprachen. Außerdem wurde nach dem Schneeballprinzip Referenzen in bereits identifizierten relevanten Publikationen nachgegangen, die auf weitere SLS und entsprechende Veröffentlichungen verwiesen.

4.2 Ergebnisse

4.2.1 Identifizierte Schulleistungsstudien

Insgesamt wurden 13 SLS identifiziert, welche die Auswahlkriterien erfüllten (vgl. Tab. 1). Den zentralen Ausgangspunkt bildete die *Third International Mathematics and Science Study* (TIMSS, Baumert et al. 2000a, 2000b) aus den Jahren 1995 und 1996. Die Zielsetzung von TIMSS war der internationale Vergleich der Abschlussjahrgänge der Sekundarstufe II. Außerdem sollte ein Vergleich der Schülerleistungen mit den Zielsetzungen der nationalen Lehrpläne erfolgen. Die TIMSS-Untersuchung bestand aus zwei Teiluntersuchungen in zwei unterschiedlichen Populationen. Die erste Teiluntersuchung wurde mit Schülerinnen und Schülern aus der gesamten Sekundarstufe II (d.h. auch an beruflichen Schulen) durchgeführt. Die zweite Teiluntersuchung beschränkte sich auf Schülerinnen und Schüler der gymnasialen Oberstufe.

Auf der Grundlage der TIMSS-Instrumente wurden in den darauffolgenden Jahren weitere Oberstufenstudien in einzelnen Bundesländern durchgeführt. In der Längsschnittstudie *Bildungsverläufe und psychosoziale Entwicklung im Jugend- und jungen Erwachsenenalter* (BIJU, Baumert et al. 2003; Köller et al. 2000) wurden Mathematikleistungen in den Jahren 1995, 1997 und 1998 in vier Bundesländern im Längsschnitt erfasst. In Baden-Württemberg wurde im Jahre 2002 mit den TIMSS-Instrumenten die Studie *Transformation des Sekundarschulsystems und akademische Karrieren* (TOSCA, Nagy et al. 2007; Watermann et al. 2004) durchgeführt, um das Leistungsniveau von Allgemeinbildenden und Beruflichen Gymnasien zu vergleichen. Auf TOSCA aufbauend folgte im Jahre 2006 mit *TOSCA-Repeat* (TOSCA-R, Nagy et al. 2010; Neumann und Nagy 2010) eine erneute Erhebung in Baden-Württemberg, um die nach den Husumer Beschlüssen der KMK neugeordnete gymnasiale Oberstufe zu evaluieren.

In Hamburg wurde in zwei Längsschnittstudien die Lernentwicklung in der gymnasialen Oberstufe mit TIMSS-Instrumenten untersucht. Die Studie *Aspekte der Lernausgangslage und der Lernentwicklung* (LAU, Nagy et al. 2007; Lehmann et al. 2012a, 2012b) testete in den Jahren 2002 und 2005 die Jahrgangsstufen 11 und 13. In der Anschlussstudie *Kompetenzen und Einstellungen von Schülerinnen und Schülern* (KESS, Vieluf et al. 2011, 2014) wurde ein Schülerjahrgang bis zum Abitur im Jahre 2011 (bei G8) oder im Jahre 2012 (bei G9) untersucht.

¹ Suchbegriff: (Schulleistungsstudie OR Large-Scale-Assessment OR Oberstufenstudie) AND (Mathematik OR Mathematikkompetenz OR „mathematische Kompetenz“ OR Mathematikleistung) AND (Oberstufe OR „Sekundarstufe II“ OR „Sekundarbereich II“ OR Abitur).

Tab. 1 Schulleistungsstudien zur Mathematikleistung in der gymnasialen Oberstufe seit 1995 (geordnet nach Zeitpunkt)

	Studie	Erhebungszeitpunkt	Jgst	BL	Schulformen	Konstrukt	Anzahl Items	Itemquelle	N (real.)
1	TIMSS	Frühjahr 1995 & 1996	12 & 13 (G9) 12 (G8)	Alle	AG, IGS, BG	MGB	44 ^a	TIMSS	1629 ^{bc}
						VUM	68 ^a	TIMSS	3657 ^c
2a	BIJU 12	Frühjahr 1997	12 (G8 & G9)	BE, MV, NW, SH	AG, IGS	VUM	24	TIMSS (19) & BIJU (5)	6256
2b	BIJU 13	Frühjahr 1998	13 (G9)	BE, NW, SH	AG, IGS	VUM	24	TIMSS	670
3	TOSCA	Frühjahr 2002	13 (G9)	BW	AG, BG	MGB	19	TIMSS	1177
						VUM	68 ^a	TIMSS	4730
4a	LAU 11	Herbst 2002	11 (G9)	HH	AG, BG, IGS, AufG	MGB	28	TIMSS (17) & LAU (9)	5631
						BWF	61 ^a	LAU	5634
4b	LAU 13	Frühjahr 2005	13 (G9)	HH	AG, IGS, BG, AufG	MGB	19	TIMSS	1309
						BWF VUM	17 68 ^a	LAU TIMSS	5317 5317
5	TOSCA-R	Frühjahr 2006	13 (G9)	BW	AG, BG	MGB	19	TIMSS	1260
						VUM	68 ^a	TIMSS	5048
6	ZAS	Frühjahr 2007 & 2008 & 2009	13 (G9)	HB, HE	AG, IGS	VUM	15	TIMSS	2191
7	TOSCA-Sachsen	Frühjahr 2009 & 2010	12 (G8)	SN	AG	VUM	12	TIMSS	1215
8a	KESS 11	Herbst 2009	11 (G8) 11 (G9)	HH	AG, BG, IGS, AufG	MGB	28	TIMSS (17) & LAU (9)	2390
						BWF	59 ^a	LAU (37) & KESS (22)	7886
9	NEPS-TH	2010 & 2011	12 (G8)	TH	AG	MGB	35 ^a	NEPS	2251

Tab. 1 (Fortsetzung)

	Studie	Erhebungszeitpunkt	Jgst	BL	Schulformen	Konstrukt	Anzahl Items	Itemquelle	N (real.)
8b	KESS 12/13	Frühjahr 2011 & 2012	12 (G8) 13 (G9)	HH	AG, IGS, BG, AufG	MGB	19	TIMSS	1650
						BWF	17	LAU	6593
						VUM	68 ^a	TIMSS	6593
10	NEPS- BW	Frühjahr 2011 & 2012 & 2013	12 (G8) 13 (G9)	BW	AG	MGB	20	NEPS	4883
11	LISA-6	Frühjahr 2013	13 (G9)	SH	AG, BG	MGB	20	NEPS	3195
12	NEPS- SC4	2. Hj. 2013	12 (G8 & G9)	Alle	AG, IGS	MGB	29 ^a	NEPS	5733
13	NEPS- SC3	2. Hj. 2017	12 (G8 & G9)	Alle	AG, IGS	MGB	30 ^a	NEPS	3785

Anmerkungen: **Bundesländer:** BE Berlin, BW Baden-Württemberg, HB Bremen, HH Hamburg, MV Mecklenburg-Vorpommern, NW Nordrhein-Westfalen, SH Schleswig-Holstein, SN Sachsen, ST Sachsen-Anhalt, TH Thüringen, **Schulformen:** AG Allgemeinbildendes Gymnasium, AufG Aufbaugymnasium, BG Berufliches Gymnasium, IGS Gesamtschule,

Konstrukte: BWF Basiswissen und -fertigkeiten, MGB Mathematische Grundbildung, VUM Voruniversitäre Mathematik.

^a Multi-Matrix-Design (d. h., die Items wurden nicht von allen Testpersonen bearbeitet)

^b Hier nur Testpersonen aufgeführt, welche die Allgemeine Hochschulreife anstrebten

^c Realisierte Stichprobenanzahlen wurden anhand des nationalen TIMSS-Datensatzes ermittelt

Ebenfalls mit Hilfe von TIMSS-Instrumenten wurden Abschlussjahrgänge zwischen 2007 und 2009 in Bremen und Hessen in der *Zentralabiturstudie* (ZAS, Maag Merki 2012) evaluiert. Auch die Studie *TOSCA-Sachsen* (Wagner et al. 2014) in den Jahren 2009 und 2010 basierte auf TIMSS-Items.

Im Rahmen des *Nationalen Bildungspanels* (*National Educational Panel Study*, NEPS, Blossfeld et al. 2011) erfolgten bis dato vier Studien in der gymnasialen Oberstufe. Originäre NEPS-Erhebungen wurden in der Jahrgangsstufe 12 im Jahre 2013 mit der *Startkohorte 4* (NEPS-SC4, Fischer et al. 2017) und im Jahre 2017 mit der *Startkohorte 3* (NEPS-SC3, Petersen et al. 2020) administriert. Ergänzungstudien wurde in den Jahren 2010 und 2011 mit *NEPS-Thüringen* (NEPS-TH, Kock et al. in Vorbereitung) und in den Jahren 2011 bis 2013 mit *NEPS-Baden-Württemberg* (NEPS-BW, Duchhardt 2015) vorgenommen, um organisatorische Strukturveränderungen in den beiden Bundesländern zu evaluieren. Außerdem wurden beim sechsten Erhebungszeitpunkt der Studie *Lesen in der Sekundarstufe* (LISA-6, Kampa et al. 2016) im Jahr 2013 im Abschlussjahr der Allgemeinbildenden und Beruflichen Gymnasien in Schleswig-Holstein Mathematikleistungen mit Hilfe von NEPS-Items erhoben.

Lediglich TIMSS, NEPS-SC4 und NEPS-SC3 wiesen eine repräsentative Stichprobe für das gesamte Bundesgebiet auf. Die übrigen 10 Studien beschränkten sich jeweils auf repräsentative Stichproben in einzelnen Bundesländern.

4.2.2 In den SLS verwendete Konstrukte und Instrumente zur Messung der Mathematikleistungen

Für die Messung der Mathematikleistungen wurden den identifizierten SLS im Wesentlichen drei Konstrukte zugrunde gelegt, die unten ausführlicher vorgestellt werden. Für das Konstrukt der mathematischen Grundbildung, welches in diesem Artikel im Sinne des Literacy-Ansatzes verstanden wird, liegen die beiden verschiedenen Operationalisierungen aus TIMSS sowie NEPS vor. Beide Grundbildungstests waren mit curricularen Inhalten der Sekundarstufe I lösbar. Auch das LAU/KESS-Instrument beschränkte sich auf curriculare Inhalte der Sekundarstufe I, erfasste jedoch ein anderes Konstrukt, welches als *curriculares Basiswissen und curriculare Basisfertigkeiten* bezeichnet werden kann. Im Rahmen von TIMSS wurde schließlich noch das Konstrukt *voruniversitäre Mathematik* gemessen. Der zugehörige Test adressierte curriculare Inhalte der gymnasialen Oberstufe.

TIMSS: Mathematische Grundbildung Der Grundbildungstest bei TIMSS hatte den Anspruch, einen Kompromiss zwischen dem angelsächsischen Literacy-Konzept und den curricularen Inhalten der Mittelstufe herzustellen (Baumert et al. 1999). Dementsprechend waren fast alle Items des Grundbildungstests in einen lebensweltlichen Kontext eingebunden, und abstrakte, rein innermathematisch orientierte Items waren nahezu nicht enthalten (vgl. Baumert et al. 1999). Inhaltlich berücksichtigte der TIMSS-Grundbildungstest insbesondere die drei Gebiete „(1) Zahlen und Zahlverständnis [...] sowie Proportionalität, (2) algebraische Terme, lineare Gleichungen und Graphen sowie (3) Messen und Schätzen“ (Baumert et al. 1999, S. 10). Als zweite Dimension wurden vier unterschiedliche Anforderungsarten (Wissen, Beherrschung von Routineverfahren, Beherrschung von komplexen Verfahren, Lösung von anwendungsbezogenen Aufgaben und innermathematischen Problemen) definiert, die in den Items zu den jeweiligen Inhaltsgebieten abgebildet werden sollten. Bezogen auf die Trias der gymnasialen Oberstufenziele adressiert der TIMSS-Grundbildungstest somit vor allem den Aspekt der vertieften Allgemeinbildung, da von den Testpersonen gefordert ist, Problemstellungen aus dem Alltagskontext mit mathematischen Mitteln der Sekundarstufe I zu lösen. Aus heutiger Perspektive ist der TIMSS-Grundbildungstest allerdings inhaltlich zu eng konstruiert, da zum Beispiel der Bereich *Daten und Zufall* im Test unterrepräsentiert ist.

Zur kriteriumsorientierten Interpretation der TIMSS-Ergebnisse entwickelten Klieme et al. (2000) anhand der deutschen Stichprobe mit dem Verfahren des *Scale-Anchoring* (vgl. Beaton und Allen 1992) ein Stufenmodell mit vier Niveaustufen

Tab. 2 Niveaustufen aus TIMSS zur mathematischen Grundbildung (Klieme et al. 2000, S. 133)

Scorewert	Niveau	Inhaltliche Charakterisierung
>600	Stufe IV	Mathematisches Argumentieren (insbesondere mit graphischen Darstellungen)
501–600	Stufe III (Mindestniveau Sek. I)	Bildung von Modellen und Verknüpfen von Operationen
401–500	Stufe II	Anwendung von einfachen Routinen
≤400	Stufe I	Alltagsbezogene Schlussfolgerungen

100 g einer Speise haben 300 Kalorien. Wie viele Kalorien haben dann 30 g derselben Speise?

- a) 90 b) 100 c) 900 d) 1000 e) 9000

Abb. 1 Item D7 aus TIMSS-Mathematische Grundbildung auf der Stufe II mit Itemschwierigkeit 451 (Baumert et al. 1999, S. 25)

Eine Schulklasse plant einen Busausflug in einen Wildpark. Ein Bus für bis zu 45 Personen kostet 600 Centros (Geldeinheiten), und die Eintrittskarten kosten 30 Centros pro Person.

Wenn jede Person 50 Centros für den gesamten Ausflug (Bus und Eintritt) bezahlt, wie viele Personen müssen dann mindestens teilnehmen, damit die Kosten gedeckt sind?

- a) 12 b) 20 c) 30 d) 45

Abb. 2 Item A5 aus TIMSS-Mathematische Grundbildung auf der Stufe III mit Itemschwierigkeit 555 (Baumert et al. 1999, S. 35)

(vgl. Tab. 2). So beinhaltet die Stufe II die Anwendung von einfachen Standardroutinen wie beispielsweise den Dreisatz (vgl. Abb. 1), während in Stufe III Modelle gebildet und Operationen verknüpft werden müssen (vgl. Abb. 2). Wie das Itembeispiel aus Abb. 2 und weitere veröffentlichte Items (Baumert et al. 1999) zeigen, stellt Stufe III Anforderungen dar, die am Ende der Sekundarstufe I erwartet werden. Dementsprechend erscheint es plausibel, das Erreichen von Stufe III als Mindestniveau für den Mittleren Schulabschluss anzusehen (vgl. Watermann et al. 2004).

NEPS: Mathematische Grundbildung Die Items des NEPS wurden ähnlich wie Items des *Programme for International Student Assessment* (PISA) auf der Grundlage eines Literacy-Konzeptes entwickelt (Ehmke et al. 2009, Schnittjer und Duchhardt 2015). Dabei wurden die NEPS-Items entlang der vier Inhaltsbereiche *Quantität, Raum und Form, Veränderung und Beziehung und Daten und Zufall* sowie der sechs prozessbezogenen Kompetenzen der Bildungsstandards entwickelt. Ein exemplarisches Item für die Klassenstufe 12 aus dem Bereich Veränderung und Beziehung ist in Abb. 3 zu sehen. Im Vergleich zum TIMSS-Test zur mathematischen Grundbildung war der Inhaltsbereich Daten und Zufall bei NEPS deutlich stärker repräsentiert und spiegelt somit die stärkere Betonung der Stochastik in den letzten 25 Jahren wider. Bezogen auf die Trias der gymnasialen Oberstufenziele adressiert

Im Sommer hat der Zoo von Astadt normalerweise mehr Besucher als im Winter. Aber in diesem Jahr wurde im Oktober ein Braunbär geboren, der mit seiner Niedlichkeit sehr viele Besucher anzog.

Welches der folgenden Diagramme stellt die Anzahl der Besucher im Zoo von Astadt korrekt dar?

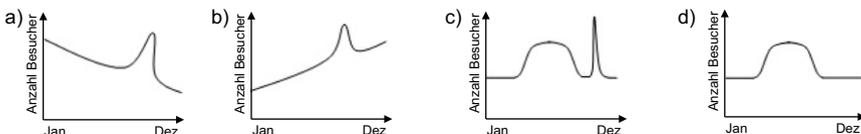


Abb. 3 Exemplarisches Item aus NEPS für die Erhebung in Jahrgangsstufe 12 (Schnittjer und Duchhardt 2015, S. 5)

Tab. 3 Niveaustufenmodell aus LISA-6 auf der Grundlage der NEPS-Items nach Kampa et al. (2016)

Scorewert	Niveau	Inhaltliche Charakterisierung
>675	Stufe IV	Komplexe mathematische Prozesse durchführen, reflektieren und beurteilen
561–675	Stufe III	Mehrschrittige mathematische Prozesse durchführen
461–560	Stufe II (Mindestniveau Sek. I)	Wenigschrittige naheliegende mathematische Prozesse durchführen
290–460	Stufe I	Einfache mathematische Standardprozesse durchführen
<290	–	Verfehlen der Anforderungen der Sek I

Anmerkung. Das Niveaustufenmodell von LISA-6 (Kampa et al. 2016) beinhaltet für jede Stufe sechs inhaltliche Charakterisierungen entlang der prozessbezogenen Kompetenzen der Bildungsstandards. Die hier aufgeführte summative inhaltliche Charakterisierung für jede Stufe ist eine Zusammenfassung der Verfasser bzw. Verfasserin dieses Artikels

auch der NEPS-Grundbildungstest schwerpunktmäßig den Aspekt der vertieften Allgemeinbildung, wobei er noch konsequenter am Literacy-Konzept orientiert ist als der TIMSS-Grundbildungstest.

Ein allgemeines Niveaustufenmodell für die NEPS-Items existiert nicht. Auf der Basis der LISA-6-Daten wurde durch Expertinnen und Experten aus der Mathematikdidaktik und der Bildungsadministration mit Hilfe des *Standard-Setting-Verfahrens* (vgl. Cizek 2012) ein Niveaustufenmodell (vgl. Tab. 3) entwickelt (vgl. detaillierte Beschreibung in Kampa et al. 2016). Dabei wurden von den Expertinnen und Experten Leistungen ab der Stufe II im Einklang mit Zielen der Bildungsstandards für den Mittleren Schulabschluss angesehen. In einer Linkingstudie (Kampa et al. 2018) mit den in LISA-6 verwendeten NEPS-Items und Items aus dem IQB-Bildungstrend wurden die LISA-6-Daten auf das IQB-Stufenmodell der Bildungsstandards für den Mittleren Schulabschluss transformiert. Hier zeigte sich, dass die im Stufenmodell von LISA-6 verwendete Definition von Mindestniveau (hier Stufe II) bedeutet, dass die Testpersonen mindestens den oberen Bereich der *Regelstandards plus* der Bildungsstandards für den Mittleren Schulabschluss erreichten. Gemessen am Verständnis der Bildungsstandards stellt die Stufe II also eine eher ambitionierte Beschreibung von Mindestniveau am Ende der Sekundarstufe I dar.

Beim Vergleich der Niveaustufenmodelle von TIMSS und LISA-6 auf Grundlage der inhaltlichen Charakterisierungen fällt auf, dass in beiden Modellen die Komplexität der mathematischen Prozesse mit höheren Niveaustufen zunimmt. Während im Stufenmodell von TIMSS die Beherrschung von einfachen Routinen die Stufe II charakterisiert, umfasst die Durchführung von einfachen mathematischen Standardprozessen die Stufe I im LISA-6-Stufenmodell. Von daher erscheint es plausibel, dass in TIMSS die Stufe III und bei LISA-6 die Stufe II als Mindestniveau für den mittleren Schulabschluss definiert wurden, da die Schülerinnen und Schüler erst ab diesen Stufen elementare mathematische Routinen und Standardprozesse mit ausreichender Sicherheit beherrschen.

LAU/KESS: Basiswissen und Basisfertigkeiten Der LAU/KESS-Test umfasste Items zum curricularen Basiswissen und zu curricularen Basisfertigkeiten und wurde nicht auf Basis des Literacy-Ansatzes gestaltet. Er bestand daher zum größten

Vereinfachen Sie so weit wie möglich: $5a + 11b + 7a$

- a) $11b + 12a$ b) $5a + 18ab$ c) $12a^2 + 11b$ d) $23ab$

Abb. 4 Item aus dem LAU/KESS-Test mit empirisch geringer Schwierigkeit in LAU11 (Lehmann et al. 2012a, S. 50)

Gegeben ist die quadratische Funktion $f(x) = 0,8x^2 + 3$. Von welcher der folgenden Funktionen hat der Graph mit dem Graphen von f genau zwei Schnittpunkte?

- a) $g(x) = 2x^2 - 1$ b) $g(x) = 1,5x^2 + 4$ c) $g(x) = -x^2 + 3$ d) $0,6x^2 + 2$

Abb. 5 Item aus dem LAU/KESS-Test mit empirisch hoher Schwierigkeit in LAU11 (Lehmann et al. 2012a, S. 50)

Teil aus innermathematisch orientierten Items zu den Themengebieten Arithmetik, Algebra und Geometrie (vgl. Abb. 4 und Abb. 5). Items zur Stochastik waren kaum vertreten. Wie auch die Grundbildungstests von TIMSS und NEPS kann der LAU/KESS-Test dem Aspekt der vertieften Allgemeinbildung zugeordnet werden, wenn unter vertiefter Allgemeinbildung nicht ausschließlich lebensweltrelevante Anforderungen, sondern auch eine Konsolidierung der innermathematischen Lerninhalte der Sekundarstufe I verstanden wird. Zudem haben die Inhalte des LAU/KESS-Tests Relevanz für den Aspekt der Studierfähigkeit, da algebraische Grundfertigkeiten wie der Umgang mit Termen und quadratischen Gleichungen (vgl. Abb. 4 und Abb. 5) von MINT-Hochschullehrenden als notwendige Voraussetzungen für die Aufnahme eines MINT-Studiums angesehen werden (Koeppf und Kramer 2015; Neumann et al. 2017). Ein Niveaustufenmodell wurde für den LAU/KESS-Test bisher nicht entwickelt.

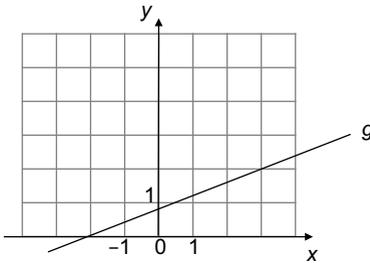
TIMSS: Voruniversitäre Mathematik Im TIMSS-Testteil zur voruniversitären Mathematik (*Advanced Mathematics*) stand die Anwendung von fachimmanenten Konzepten, Operationen und Gesetzmäßigkeiten im Vordergrund. Hier spielte der Literacy-Ansatz im Gegensatz zum TIMSS-Grundbildungstest nur eine untergeordnete Rolle (Klieme 2000). Um eine internationale Vergleichbarkeit der Testergebnisse in voruniversitärer Mathematik zu ermöglichen, wurde eine „transnationale curriculare Validität“ (Baumert et al. 1999, S. 13) zwischen den teilnehmenden Ländern angestrebt. Die drei Inhaltsgebiete *Analysis*, *Geometrie* und *Zahlen/Gleichungen/Funktionen* bildeten den Schwerpunkt des Tests. Die zwei weiteren Inhaltsgebiete *Wahrscheinlichkeit/Statistik* und *Aussagenlogik/Beweise* nahmen nur eine untergeordnete Rolle ein. Der größte Teil der Testitems erforderte Fähigkeiten in Inhaltsgebieten der Oberstufenmathematik (z. B. Grenzwertbildung, Differentiation und Integration, Vektorrechnung). Ein Rating von Personen aus der Mathematikdidaktik in Deutschland kam jedoch zu dem Ergebnis, dass zum damaligen Zeitpunkt etwa 40% der Items des voruniversitären Mathematiktests nicht über den Unterrichtsstoff der Sekundarstufe I hinausgingen (Klieme 2000). Eine weitere Beurteilung von Expertinnen und Experten aus Deutschland ergab, dass 80% der Items als curricular valide für Deutschland eingestuft wurden (vgl. Baumert et al. 2000c). Allerdings zeigte ei-

Wie viele Punkte mit ganzzahligen Koordinaten existieren auf dem Graphen der Funktion $y = \frac{12}{x}, x > 0$?

- a) 2 b) 4 c) 6 d) unendlich viele

Abb. 6 Item I9 aus TIMSS-Voruniversitäre Mathematik auf der Stufe II mit Itemschwierigkeit 499 (Foy und Arora 2009, S. 11)

Die Gerade g in der Figur ist der Graph von $y = f(x)$.



$\int_{-2}^3 f(x) dx$ ist gleich

- a) 3 b) 4 c) 4,5 d) 5 e) 5,5

Abb. 7 Item K6 aus TIMSS-Voruniversitäre Mathematik auf der Stufe III mit Itemschwierigkeit 537 (Baumert et al. 1999, S. 76)

ne Sensitivitätsanalyse, dass internationale Unterschiede bezüglich der curricularen Validität kaum einen Einfluss auf die Länderrangfolge hatten (Baumert et al. 2000c). Auch eine Studie von Kahnert (2014) weist darauf hin, dass der TIMSS-Test zur voruniversitären Mathematik zur validen Erfassung von Ergebnissen des Mathematikunterrichts in der gymnasialen Oberstufe geeignet ist. In einer Untersuchung wurden Leistungsdaten aus dem Zentralabitur 2011 in Nordrhein-Westfalen mit denen aus zusätzlich bearbeiteten TIMSS-Items gemeinsam skaliert. Ein eindimensionales Modell wies eine deutlich bessere Passung auf als ein zweidimensionales Modell, in dem die Zentralabituraufgaben und die TIMSS-Items als zwei unterschiedliche Dimensionen behandelt wurden (Kahnert 2014). Dies bedeutet, dass der TIMSS-Test voruniversitäre Mathematik zum Abitur vergleichbare Anforderungen stellt. Um diese Einschätzung nochmals abzusichern, haben wir die Items des TIMSS-Tests zur voruniversitären Mathematik inhaltlich analysiert. Es zeigte sich, dass der Test schwerpunktmäßig die Bildungsziele Wissenschaftspropädeutik und Studierfähigkeit adressiert. Das Item I9 (Abb. 6) erfordert beispielsweise, Wissen über die Teilbarkeit von Zahlen und die Ganzzahligkeit von Koordinaten zu verknüpfen. In Item K6 (Abb. 7) muss die Grundvorstellung vom Integral als (orientierten) Flächeninhalt aktiviert und adäquat angewendet werden. In Item J19 (Abb. 8) soll eine geometrische Situation mit Mitteln der analytischen Geometrie mathematisiert und ein Beweis geführt werden. Die Fähigkeit, zwei mathematische Begriffe in zunächst ungewohnter Weise zu verknüpfen (Item I9), mit grundlegenden mathematischen Symbolen verständlich umzugehen (Item K6) und erlernte mathematische Begriffe

Im unteren Viereck $ABCD$ schneiden sich die Diagonalen \overline{AC} und \overline{BD} im Punkt E .

BEWEISEN Sie, dass E der Mittelpunkt von \overline{AC} und \overline{BD} ist.

Schreiben Sie alle Ihre Arbeitsschritte auf.

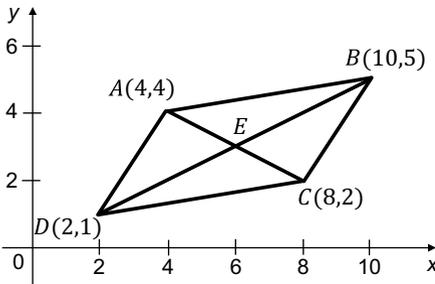


Abb. 8 Item J19 aus TIMSS-Voruniversitäre Mathematik auf der Stufe IV mit Itemschwierigkeit 614 für eine teilweise korrekte und 732 für eine vollständig korrekte Antwort (Foy und Arora 2009, S. 26)

und Verfahren zum Beweisen von Aussagen einzusetzen (Item J19) sind zentrale Aspekte der Studierfähigkeit und der Wissenschaftspropädeutik. Insgesamt kann der TIMSS-Test zur voruniversitären Mathematik trotz der herausgestellten Defizite (beschränkte curriculare Validität für Deutschland, Unterrepräsentanz der Stochastik) Ergebnisse zu den Mathematikleistungen in den Bereichen der Studierfähigkeit und der Wissenschaftspropädeutik liefern.

Für den Test zur voruniversitären Mathematik entwickelte Klieme (2000) ebenfalls mit Hilfe des Scale-Anchoring-Verfahrens ein vierstufiges Modell (Tab. 4). Items der Stufe II konnten noch mit Kenntnissen aus der Sekundarstufe I bewältigt werden (vgl. Abb. 6). Erst bei Items ab der Stufe III waren Kenntnisse der gymnasialen Oberstufe erforderlich (vgl. Abb. 7). Somit kann man erst ab der dritten Stufe davon sprechen, dass die Fähigkeit gegeben ist, typische Aufgabenstellungen der Oberstufenmathematik zu bewältigen (vgl. Klieme 2000).

5 Sekundäranalyse der empirischen Daten bzw. Ergebnisse (Forschungsfrage 3)

Nachdem in der obigen Bestandsaufnahme die Studien sowie zugrundeliegende Konstrukte und verwendete Tests gegenübergestellt wurden, soll nun eine Synthese der Befunde Antwort auf die Frage geben, was über das Leistungsniveau und die Leistungsentwicklung in Mathematik von Schülerinnen und Schülern in der gymnasialen Oberstufe bekannt ist. Bei der Synthese der Befunde aus den unterschiedlichen SLS ergab sich allerdings die Herausforderung, dass die Ergebnisse nicht direkt vergleichbar waren. So wurden Ergebnisse der SLS zum Teil nur für spezielle Subpopulationen (z. B. nur Allgemeinbildendes Gymnasium) oder nur als Veränderungswert zu einer Vergleichskohorte berichtet. Außerdem bezogen sich die dargestellten Werte, wie oben ausgeführt, auf unterschiedliche Konstrukte und darüber hinaus wurden teils unterschiedliche Skalen verwendet. Da die meisten Studien

Tab. 4 TIMSS-Niveaustufenmodell zur voruniversitären Mathematik (Klieme 2000, S. 87 ff)

Scorewert	Niveau	Inhaltliche Charakterisierung
>600	Stufe IV	Selbstständiges Lösen mathematischer Probleme auf Oberstufenniveau
501–600	Stufe III (Mindestniveau Sek. II)	Anwendung von Lerninhalten der Oberstufe im Rahmen typischer Standardaufgaben
401–500	Stufe II	Anwendung einfacher mathematischer Begriffe und Regeln, die kein Verständnis von Konzepten der Oberstufenmathematik voraussetzen
≤400	Stufe I	Elementares Schlussfolgern

TIMSS- oder NEPS-Instrumente verwendeten, ist es jedoch möglich, mit Hilfe einer elaborierten Datenaufbereitung die notwendigen Studienvergleiche mit Blick auf die Konstrukte mathematische Grundbildung (TIMSS und NEPS) sowie voruniversitäre Mathematik (TIMSS) und unter Nutzung der jeweiligen Stufenmodelle vorzunehmen. Für das Konstrukt curriculares Basiswissen und curriculare Basisfertigkeiten reichten die publizierten Testwerte aus LAU und KESS nicht aus, um interpretierbare Ergebnisse für das erreichte Niveau am Ende der gymnasialen Oberstufe berichten zu können. Allerdings konnten aus diesen Studien Befunde zur längsschnittlichen Entwicklung abgeleitet werden.

5.1 Methoden zur Aufbereitung der Daten

In diesem Abschnitt werden die Methoden aus Platzgründen nur überblicksartig dargestellt. Im Online-Supplement sind sie für die interessierten Lesenden ausführlich, auch mit Bezug zu weiterführender Literatur, dargelegt, und die durch die Datenaufbereitung gewonnenen Werte für die folgenden Sekundäranalysen sind vollständig angegeben.

Im ersten Schritt mussten die Ergebnisse aus Studien mit TIMSS-Instrumenten in eine gemeinsame Metrik (TIMSS-Metrik) transformiert und anhand der TIMSS-Niveaustufenmodelle eingeordnet werden. Für TIMSS, TOSCA und LAU-13 wurden bereits die originären Auswertungen in der TIMSS-Metrik berichtet. Eine Niveaustufenverteilung für die jeweilige Gesamtpopulation dieser Studien konnte somit durch elementare Berechnungen (z.B. gewichteter Mittelwert) ermittelt werden. Für die TOSCA-R-Ergebnisse musste eine Verteilungstransformation vorgenommen werden. Dazu wurden die TOSCA-R-Ergebnisse unter der Annahme normalverteilter latenter Personenfähigkeit in die TIMSS-Niveaustufen überführt. Für KESS12/13 lagen Ergebnisse nur für Subgruppen und als Veränderungswerte (Cohens d) zu LAU 13 vor. Hier erfolgte durch gewichtete Verteilungsaggregation und anschließende Verteilungstransformation eine Zuordnung zu den TIMSS-Niveaustufen. Für BIJU, ZAS und TOSCA-SN dagegen war die Transformation in die TIMSS-Metrik nicht durch elementare Berechnungen möglich. In den Primäranalysen dieser Studien wurden für jede Testperson Summenscores gebildet, indem für jedes korrekt beantwortete Item ein Punkt vergeben wurde. Für ZAS und TOSCA-SN lagen daher die Stichproben-Mittelwerte und die Stichproben-Standardabweichungen der

manifesten Summenscores vor. Hier wurde die latente Fähigkeitsverteilung auf der TIMSS-Skala geschätzt, indem eine Normalverteilung der latenten Fähigkeiten angenommen wurde und mit Hilfe der bei TIMSS berichteten Itemschwierigkeiten die latenten Fähigkeitswerte auf der TIMSS-Skala ermittelt wurden. Dieses Vorgehen wird detailliert im Online-Supplement (Kap. A) dargestellt. Für BIJU lagen nur mittlere manifeste Summenscores vor. Daher wurde eine Standardabweichung von 89 Punkten auf der TIMSS-Skala wie bei der deutschen TIMSS-Stichprobe (Baumert et al. 2000c) angenommen und ansonsten analog zur im Online-Supplement (Kap. A) beschriebenen Berechnung vorgegangen. Die detaillierten numerischen Ergebnisse für die Studien mit TIMSS-Instrumenten und die ermittelte Verteilung auf die Niveaustufen finden sich ebenfalls im Online-Supplement (Kap. B).

Analog mussten die Studien, die NEPS-Items verwendeten, auf eine gemeinsame Metrik (LISA-6) und das in LISA-6 entwickelte Niveaustufenmodell gebracht werden. Für die LISA-6-Daten wurde bereits in der Primäranalyse die Verteilung auf die Niveaustufen berichtet. Um diese für die vier NEPS-Studien (NEPS-BW, NEPS-TH, NEPS-SC3, NEPS-SC4) zu ermitteln, wurden die Itemscores in den von NEPS zur Verfügung gestellten Scientific-Use-Files² Rasch-skaliert. Durch die Festsetzung von Itemschwierigkeiten gemäß LISA-6 sind die Ergebnisse kompatibel zur LISA-6-Metrik. Die NEPS-Tests sind jedoch nicht vollständig identisch, sodass darauf aufbauend die übrigen Itemschwierigkeiten mit einer simultanen Mehrgruppen-Skalierung geschätzt werden mussten. Auf der Grundlage dieses Schätzmodells wurden für jede Testperson 15 *Plausible Values* gezogen und diese den LISA-6-Niveaustufen zugeordnet. Abschließend wurde die prozentuale Verteilung der Plausible Values auf die LISA-6-Niveaustufen separat für die vier NEPS-Studien ermittelt.

Effektstärken werden in diesem Artikel aus zwei Gründen in Logit berichtet: Zum einen wurden in NEPS längsschnittliche Zuwächse originär auf der Logit-Skala ermittelt. Zum anderen hat die Logit-Skala gegenüber klassischen Effektstärkemaßen wie Cohens d den Vorteil, dass sie nicht von der Varianz in der Stichprobe abhängt. Zur Veranschaulichung von Effekten auf der Logit-Skala stellt Abb. 9 den Zusammenhang zwischen der Fähigkeit³ auf der Logit-Skala und der Lösungswahrscheinlichkeit für eine Aufgabe mit mittlerer Itemschwierigkeit dar. Wie dort zu sehen ist, entspricht eine Veränderung um 0,5 Logit näherungsweise einer Zunahme der Lösungswahrscheinlichkeit von 10 Prozentpunkten.

² Diese Arbeit nutzt Daten des Nationalen Bildungspanels (NEPS): Zusatzstudie Thüringen, <https://doi.org/10.5157/NEPS:TH:2.0.0>; Zusatzstudie Baden-Württemberg, <https://doi.org/10.5157/NEPS:BW:3.2.0>; Startkohorte Klasse 9, <https://doi.org/10.5157/NEPS:SC4:10.0.0>; Startkohorte Klasse 5, <https://doi.org/10.5157/NEPS:SC3:9.0.0>. Die Daten des NEPS wurden von 2008 bis 2013 als Teil des Rahmenprogramms zur Förderung der empirischen Bildungsforschung erhoben, welches vom Bundesministerium für Bildung und Forschung (BMBF) finanziert wurde. Seit 2014 wird NEPS vom Leibniz-Institut für Bildungsverläufe e. V. (IfBi) an der Otto-Friedrich-Universität Bamberg in Kooperation mit einem deutschlandweiten Netzwerk weitergeführt.

³ Ein Fähigkeitswert von 0 bedeutet auf dieser Skala, dass die Person in einer Stichprobe mittlere Fähigkeiten hat.

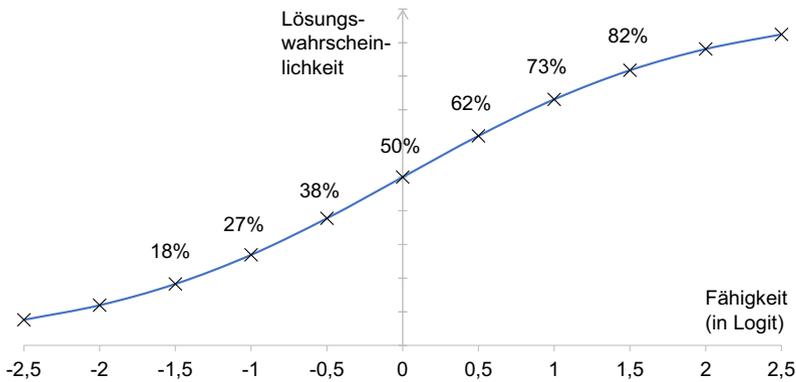


Abb. 9 Zusammenhang zwischen der Logit-Skala und der Lösungswahrscheinlichkeit anhand einer Itemfunktion für eine Aufgabe, die im Schnitt eine Lösungswahrscheinlichkeit von 50 % aufweist

5.2 Ergebnisse

Im Folgenden berichten wir die Ergebnisse der Sekundäranalyse zur Beantwortung der Forschungsfrage 3 auf Basis der aufbereiteten Daten. Dabei nutzen wir die vorgestellten Stufenmodelle (vgl. Tab. 2, 3 und 4), um die Fragen des Leistungsniveaus und der Leistungsveränderung kriterial zu beantworten.

5.2.1 Niveaus bezüglich mathematischer Grundbildung

In der bundesweiten TIMS-Studie erreichten von den Testpersonen, welche die allgemeine Hochschulreife anstrebten, 5 % Stufe I, 25 % Stufe II, 38 % Stufe III und 32 % Stufe IV im Test zur Mathematischen Grundbildung (Watermann und Baumert 2000). Damit belegten 70 % dieser Schülerinnen und Schüler am Ende der gymnasialen Oberstufe mindestens die Stufe III (vgl. Tab. 2). Entsprechend verfehlte fast ein Drittel am Ende der gymnasialen Oberstufe das für die Sekundarstufe I angesetzte Mindestniveau in mathematischer Grundbildung.

Ein durch die Aufbereitung der Daten möglicher Vergleich dieser bundesweiten Daten mit den Studien aus Baden-Württemberg (TOSCA und TOSCA-R) und Hamburg (LAU-13) zeigte, dass in Baden-Württemberg über 80 % mindestens TIMSS-Stufe III erreichte, während dieser Anteil in Hamburg bei 56 % lag (Abb. 10).

In den Studien auf Basis der Grundbildungsitems aus NEPS erreichten gemäß der in dieser Arbeit durchgeführten Sekundäranalyse zwischen 61 % (NEPS-SC3) und 91 % (NEPS-BW) der jeweiligen Gesamtstichprobe ein Leistungsniveau mindestens auf LISA-6-Stufe II (vgl. Tab. 3), das den Zielen der Allgemeinbildung für den mittleren Schulabschluss genügt (vgl. Abb. 11). Anders formuliert verfehlten zwischen 9 und 39 % der Schülerinnen und Schüler noch am Ende der gymnasialen Oberstufe die Ziele des Mittleren Schulabschlusses. Die sichtbaren Disparitäten zwischen den Studien sind zu einem Teil durch unterschiedliche Untersuchungspopulationen zu erklären. Während in NEPS-TH und NEPS-BW ausschließlich Allgemeinbildende

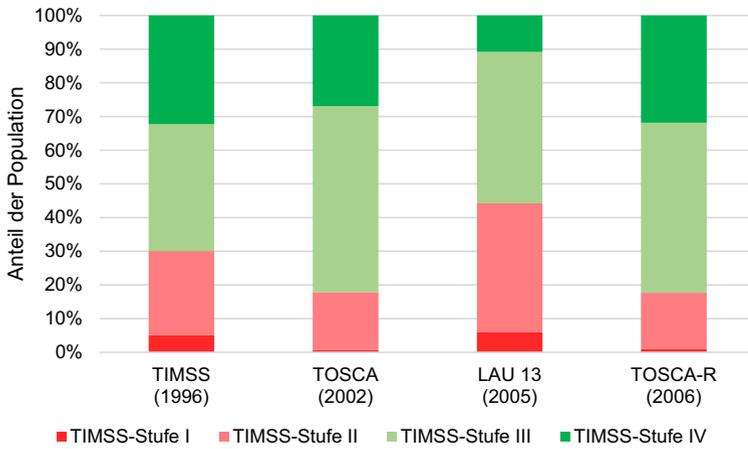


Abb. 10 Ergebnisse zur mathematischen Grundbildung in Bezug zum TIMSS-Niveaustufenmodell (vgl. Tab. 2)

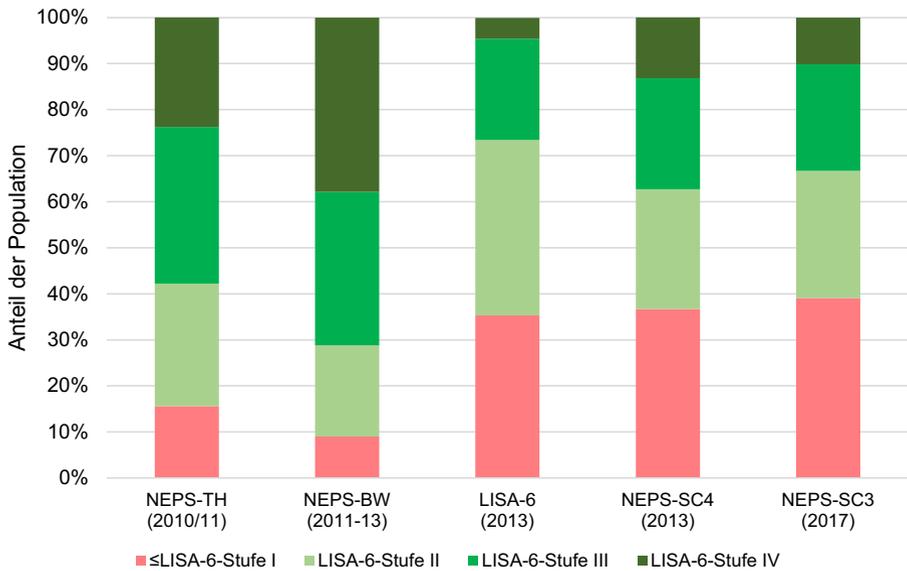


Abb. 11 Ergebnisse zur mathematischen Grundbildung in Bezug zum LISA-6-Niveaustufenmodell (vgl. Tab. 3)

Gymnasien einbezogen wurden, nahmen an LISA-6 auch Berufliche Gymnasien und an NEPS-SC4 und NEPS-SC3 auch Integrierte Gesamtschulen teil.

Insgesamt kann also mit Hilfe der Ergebnisse zur mathematischen Grundbildung aus den SLS festgestellt werden, dass der Großteil der Schülerinnen und Schüler am Ende der gymnasialen Oberstufe die Bildungsziele des mittleren Schulabschlusses in Bezug auf den Trias-Aspekt Vertiefte Allgemeinbildung erreichte. Gemäß der

Stufenmodelle von TIMSS und LISA-6 verfehlte allerdings ein nicht unerheblicher Teil auch noch am Ende der gymnasialen Oberstufe diese definierten Mindestniveaus der Sekundarstufe I.

5.2.2 Niveaus bezüglich voruniversitärer Mathematik

Im TIMSS-Test zur voruniversitären Mathematik befanden sich in Deutschland circa 30 % des Abschlussjahrgangs 1996 der gymnasialen Oberstufe auf den Stufen III und IV (vgl. Tab. 4). Damit zeigten nach dem Stufenmodell von Klieme (2000) lediglich 30 % der Schülerinnen und Schüler, dass sie mindestens die grundlegenden Lerninhalte der gymnasialen Oberstufe beherrschen. Entsprechend verfehlten 70 % der Absolventinnen und Absolventen der gymnasialen Oberstufe die Mindestanforderungen an voruniversitärer Mathematik. Ein durch die Sekundäranalyse möglicher Vergleich dieser Ergebnisse mit Befunden aus darauffolgenden SLS in einzelnen Bundesländern zeigte, dass es auch in den Leistungen bezüglich voruniversitärer Mathematik bedeutende Unterschiede zwischen den Bundesländern gab (vgl. Abb. 12). Die besten Leistungen zeigten Schülerinnen und Schüler in TOSCA-R in Baden-Württemberg, wo näherungsweise 45 % des Abschlussjahrgangs mindestens Stufe III und damit die Mindestniveaus in voruniversitärer Mathematik erreichten. Die geringsten Leistungen wurden in den BIJU-Studien und in den Stadtstaaten Hamburg und Bremen gemessen, wo lediglich teilweise nur rund 10 % der Schülerschaft mindestens Stufe III erreichten.

Ein bedeutender Unterschied zeigte sich erwartungsgemäß bei den Leistungen in Abhängigkeit von der Kursform (Leistungskurs versus Grundkurs). Bei TIMSS erreichten in den Leistungskursen 56 % der Testpersonen Stufe III oder IV. Zwischen den Bundesländern variierten die Schülerleistungen auch bei ausschließlicher Berücksichtigung der Leistungskurse (vgl. Abb. 13). Die höchsten Werte wurden

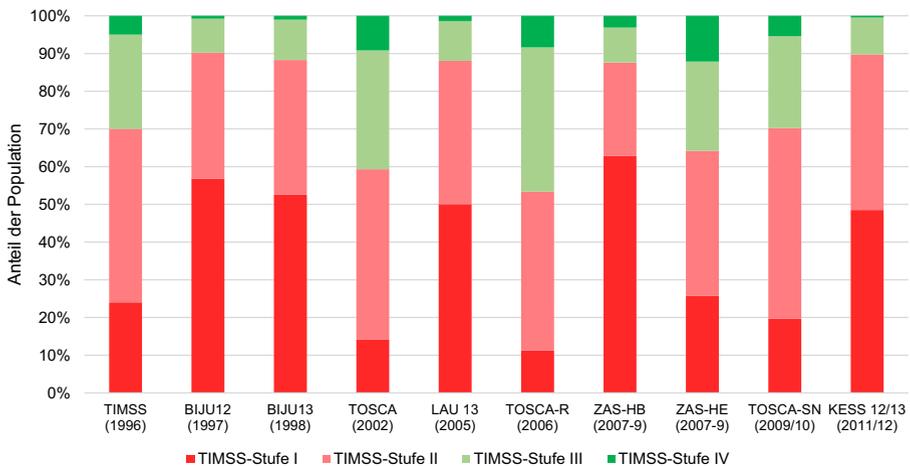


Abb. 12 Ergebnisse zur voruniversitären Mathematik in Bezug zum TIMSS-Niveaustufenmodell (vgl. Tab. 4) (Anmerkung: die Stichprobe bei ZAS-HB bestand für die Grundkurse nur aus Schülerinnen und Schülern, die in Mathematik schriftlich geprüft wurden)

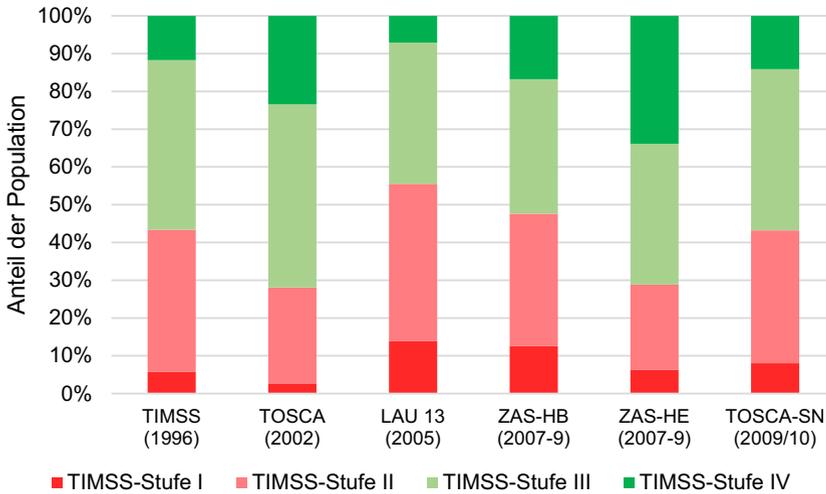


Abb. 13 Schülerleistungen in den Leistungskursen in Bezug zum TIMSS-Niveaustufenmodell (vgl. Tab. 4) (Anmerkung: Zu BIJU liegen keine kursdifferenzierenden Ergebnisse vor; bei TOSCA-R und KESS wurde zum Erhebungszeitpunkt nicht kursdifferenziert unterrichtet)

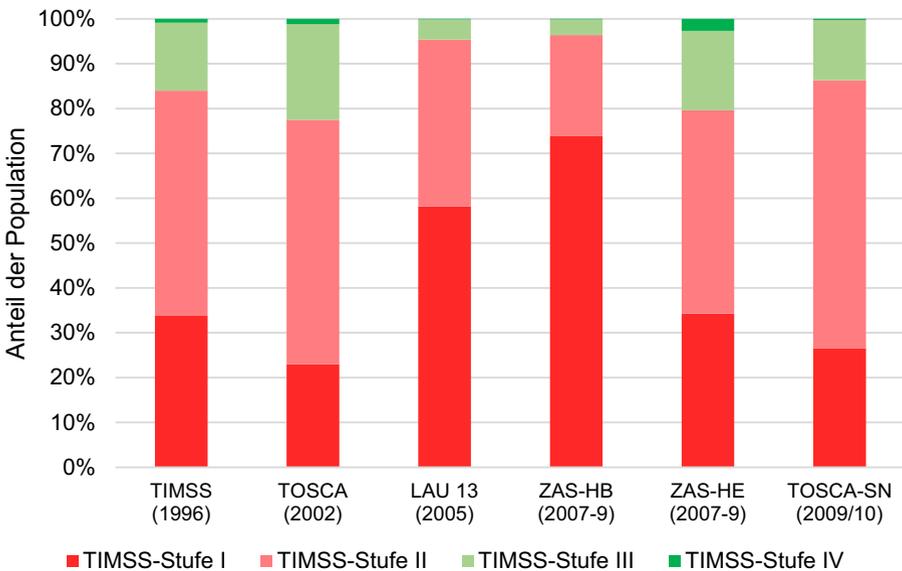


Abb. 14 Schülerleistungen in den Grundkursen in Bezug zum TIMSS-Niveaustufenmodell (vgl. Tab. 4) (Anmerkung: Zu BIJU liegen keine kursdifferenzierenden Ergebnisse vor; bei TOSCA-R und KESS wurde zum Erhebungszeitpunkt nicht kursdifferenziert unterrichtet; die Stichprobe bei ZAS-HB bestand für die Grundkurse nur aus Schülerinnen und Schülern, die in Mathematik schriftlich geprüft wurden)

in Baden-Württemberg und Hessen gemessen, wo rund 70% die Mindestziele in voruniversitärer Mathematik erreichten. In den Stadtstaaten Bremen und Hamburg erreichten hingegen etwa nur die Hälfte der Schülerinnen und Schüler in Leistungskursen die definierten Mindestanforderungen in voruniversitärer Mathematik.

In den Grundkursen zeigten sich erwartungsgemäß geringere Leistungen als in den Leistungskursen, die Unterschiede fielen aber ebenso deutlich aus (vgl. Abb. 14). Bei TIMSS 1996 erreichten lediglich 16% der Schülerinnen und Schüler in Grundkursen mindestens die Stufen III und IV. Auch hier schwankten die Werte deutlich über die Studien hinweg. In den Grundkursen der Stadtstaaten Bremen und Hamburg erreichten lediglich um die 5% der Schülerinnen und Schüler die Mindestanforderungen der Oberstufe nach Klieme (2000), wohingegen in Baden-Württemberg und Hessen um die 20% diese Mindestanforderungen erreichten.

5.2.3 Leistungszuwachs im Verlauf der gymnasialen Oberstufe

Zur Beurteilung der Frage, in welchem Ausmaß sich die Mathematikleistungen im Verlauf der gymnasialen Oberstufe verändern, wurden in der Sekundäranalyse die Ergebnisse der (echten) Längsschnittstudien LAU, KESS, NEPS-SC4 und NEPS-SC3 betrachtet (vgl. Abb. 15). Es können damit für die Konstrukte Basiswissen und Basisfertigkeiten sowie mathematische Grundbildung Befunde dargestellt werden.

In der LAU-Studie in Hamburg wurden zum Beginn der 11. Jahrgangsstufe und gegen Ende der 13. Jahrgangsstufe die Mathematikleistungen mit dem Instrument curriculares Basiswissen und curriculare Basisfertigkeiten erhoben. Die Schülerinnen und Schüler verbesserten sich innerhalb dieses Zeitraums um 0,71 Logit⁴ (List et al. 2018). Wird der zeitliche Abstand zwischen den Erhebungen (31 Monate) umgerechnet auf ein durchschnittliches jährliches Wachstum, entspricht dies einem Zuwachs von 0,27 Logit. Der KESS-Jahrgang wurde am Ende der 10. Klasse, am Beginn der 11. Jahrgangsstufe und am Ende der 12. Jahrgangsstufe bei einem G8-Abitur oder am Ende der 13. Jahrgangsstufe bei einem G9-Abitur getestet. Eine Schätzung des Leistungszuwachses im curricularen Basiswissen und in den curricularen Basisfertigkeiten der Schülerinnen und Schüler von KESS 10/11 zu KESS 12/13 ergab eine durchschnittliche Verbesserung um 0,48 Logit⁵ (Feddermann et al. 2019), was unter Berücksichtigung des Zeitabstandes in einen jährlichen Zuwachs von etwa 0,21 Logit umgerechnet werden kann.

Mit Blick auf die Zuwächse in Bezug auf die mathematische Grundbildung ergeben sich ähnliche Größenordnungen. Die Startkohorte 4 von NEPS wurde im Jahre 2010 in der Klassenstufe 9 und im Jahre 2013 in der Jahrgangsstufe 12 getestet. Die anhand der Linking-Konstante geschätzte Leistungssteigerung betrug 0,50 Logit (Fischer et al. 2017), was einem durchschnittlichen jährlichen Zuwachs von 0,17 Logit entsprach. Die Startkohorte 3 wurde in NEPS in 2014 und 2017 in den Jahrgangsstufen 9 und 12 getestet. Der auf der Basis der Linking-Konstante geschätzte Zuwachs betrug 0,60 Logit (Petersen et al. 2020), was einem jährlichen Zuwachs von 0,20 Logit entsprach.

⁴ Befunde auf Basis der Linking-Items.

⁵ Befunde auf Basis der Linking-Items.

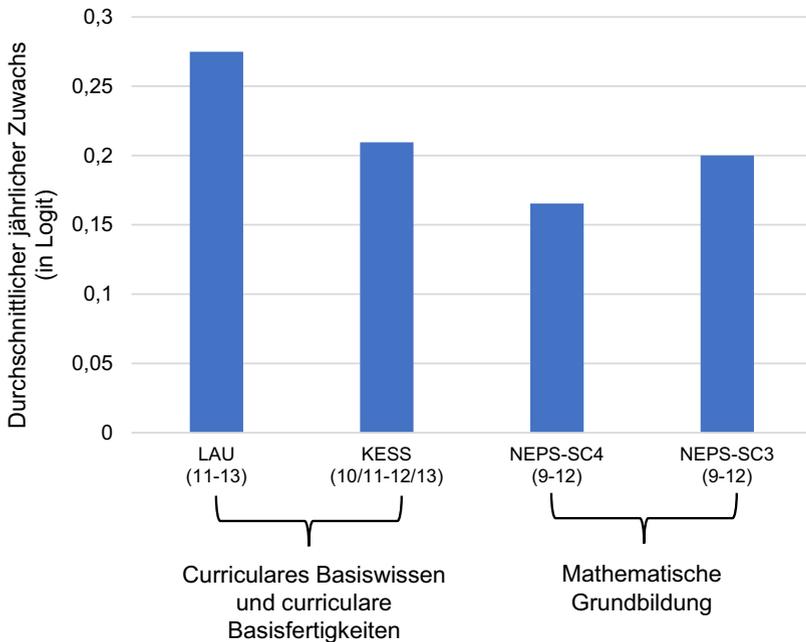


Abb. 15 Durchschnittlicher jährlicher Leistungszuwachs in Längsschnittstudien

Bei den Längsschnittstudien zeigten sich allerdings trotz der ähnlichen Größenordnungen für die Gesamtskalen auch deutliche Messvarianzen, das heißt, der Lernzuwachs erfolgte nicht homogen über alle Items. Differenzierte Analysen, etwa in Bezug auf mathematische Inhaltsbereiche erlauben wie dargestellt die verwendeten Instrumente nicht, allerdings lässt die Betrachtung einzelner Itembeispiele vermuten, dass die Leistungsentwicklung in Teilbereichen sehr dynamisch sein kann. So verhielt sich beispielsweise das Item THMA10 sowohl bei LAU als auch bei KESS als sehr sensitiv für den Mathematikunterricht in der gymnasialen Oberstufe. Die Differenz der Itemschwierigkeiten dieses Items zwischen dem Anfang und dem Ende der gymnasialen Oberstufe betrug 3,2 Logit (LAU) bzw. 2,0 Logit (KESS). Dieses Item war eines der wenigen Stochastik-Items und erforderte die Berechnung einer Pfadwahrscheinlichkeit bei einem dreistufigen Zufallsexperiment. Dieser Inhalt ist originärer Inhalt des Stochastikunterrichts der gymnasialen Oberstufe, und der deutliche Anstieg der Lösungswahrscheinlichkeit spiegelt dies vermutlich.

Im Gegensatz zum Item THMA10 zeigte sich zum Beispiel bei Item THMA09 weder bei LAU noch bei KESS eine bedeutende Veränderung der Lösungswahrscheinlichkeiten im Verlaufe der gymnasialen Oberstufe. Die LAU-Stichprobe zeigte am Ende der gymnasialen Oberstufe sogar eine geringere Fähigkeit bei diesem Item als am Beginn der gymnasialen Oberstufe ($-0,5$ Logit), während sich der KESS-Jahrgang bei diesem Item leicht verbesserte ($+0,3$ Logit). Das Item THMA09 erforderte Wissen um den Kongruenzbegriff und die Kongruenzsätze im Dreieck. Dieses Thema wird in der Sekundarstufe I unterrichtet und im Mathematikunterricht der

gymnasialen Oberstufe im Prinzip weder aufgegriffen noch benötigt, was die beobachtete Stagnation erklären kann.

5.2.4 Vergleich der Leistungsniveaus verschiedener Abschlusskohorten

Ein Kohortenvergleich ist auf Grundlage der vorliegenden SLS wie in Tab. 1 ersichtlich lediglich mit den Ergebnissen von TOSCA und TOSCA-R in Baden-Württemberg, mit den LAU und KESS-Studien in Hamburg und den NEPS-Daten aus der 12. Jahrgangsstufe möglich⁶. Hier zeigte sich, dass in Baden-Württemberg die Leistungsstände in mathematischer Grundbildung und voruniversitärer Mathematik von 2002 zu 2006 leicht zugenommen hatten. In Hamburg nahmen von 2005 zu 2011/12 die Leistungsstände des Abschlussjahrgangs in voruniversitärer Mathematik dagegen leicht ab (vgl. Abb. 12). Ein Vergleich der bundesweiten NEPS-Daten von 2013 und 2017 ergab so gut wie keinen Unterschied in der Leistung der beiden Kohorten der 12. Jahrgangsstufe (vgl. Abb. 11). Der Mangel an Daten erlaubt keine Aussagen über längere Vergleichszeiträume.

6 Diskussion

Die Ergebnisse der berichteten Sekundäranalysen auf Basis des systematischen Reviews zu SLS in der gymnasialen Oberstufe zeigen auf der einen Seite, dass es zwischen den Bundesländern deutliche Disparitäten in Bezug auf die erzielten Mathematikleistungen gibt. Dies reiht sich zwar in die Befunde zu den Leistungen in der Sekundarstufe I ein, die Disparitäten konnten aber durch die Sekundäranalyse erstmals auch für die gymnasiale Oberstufe abgeschätzt werden. Auf der anderen Seite zeichnet sich für alle Bundesländer das Bild, dass die Mehrheit der Schülerinnen und Schüler am Ende der gymnasialen Oberstufe Anforderungen im Bereich der Allgemeinbildung (gemessen durch die TIMSS- und NEPS-Tests zur mathematischen Grundbildung) erfüllten, die für den Mittleren Schulabschluss erwartet werden. In einigen Bundesländern zeigte aber immer noch ein erheblicher Anteil der Schülerinnen und Schüler am Ende der Sekundarstufe II deutliche Schwierigkeiten in diesem Bereich, das heißt im Umgang mit lebensweltlichen mathematischen Anforderungen zu Mittelstufeninhalten.

Im Bereich der Wissenschaftspropädeutik und Studierfähigkeit (gemessen durch den TIMSS-Test zur voruniversitären Mathematik) zeigten die Abiturientinnen und Abiturienten unbefriedigende Leistungen. Nur eine Minderheit verfügte am Ende der Sekundarstufe II über Fähigkeiten zu fachspezifischen Lerninhalten der gymnasialen Oberstufe (z. B. verständiger Umgang mit elementaren Grundvorstellungen zum Integral). Lediglich in den Leistungskursen stellte sich die Situation positiver dar. Hier zeigte ein großer Teil der Schülerinnen und Schüler Fähigkeiten im Umgang

⁶ An den TIMSS-Untersuchungen in der Sekundarstufe II (*TIMSS Advanced*) in den Jahren 2008 und 2015 hat Deutschland (nach der einmaligen Teilnahme 1995/96) nicht teilgenommen, sodass für Deutschland keine repräsentativen TIMSS-Trenddaten vorliegen. Auch die anderen SLS sind wegen regionaler Beschränkungen oder nicht-vergleichbarer Instrumente nicht für den Kohortenvergleich geeignet.

mit Inhalten der Oberstufenmathematik, die man aus wissenschaftspropädeutischer Sicht und aus der Perspektive der Sicherung der Studierfähigkeit insbesondere im Bereich der MINT-Fächer mindestens erwarten sollte (vgl. Neumann et al. 2017).

Der Leistungszuwachs im Verlauf der gymnasialen Oberstufe lag in den untersuchten Längsschnittstudien bei etwa durchschnittlich 0,2 Logit pro Jahr. Dieser Zuwachs bezog sich auf Tests zur mathematischen Grundbildung (NEPS) und zum curricularen Basiswissen und den curricularen Basisfertigkeiten (LAU/KESS). Diese Ergebnisse zeigen, dass Schülerinnen und Schüler im Verlauf der gymnasialen Oberstufe auch bei allgemeinbildenden Lerninhalten der Sekundarstufe I noch dazu lernen. Allerdings verdeutlichten die Einzelitemanalysen von LAU und KESS auch, dass die gemessene Zunahme sehr stark von den verwendeten Linking-Items abhing und deswegen differenzierter betrachtet werden müsste.

Bezüglich der Leistungsentwicklung von Abschlusskohorten lässt sich anhand der Daten aus Baden-Württemberg und Hamburg vermuten, dass die Leistungen der Abiturientinnen und Abiturienten ungeachtet struktureller Veränderungen innerhalb der Bundesländer im Zeitverlauf relativ stabil sind. Diese Beobachtung würde in gewissem Kontrast zur aktuellen Diskussion stehen. Allerdings ist die Befundlage für diese Fragestellung insgesamt sehr dünn, so dass keine über eine Vermutung hinausgehende abschließende Beurteilung erfolgen kann.

6.1 Zielsetzungen des Mathematikunterrichts in der Oberstufe

Die Ergebnisse verdeutlichen, dass die Diskussion über die Wirkung des Mathematikunterrichts in der gymnasialen Oberstufe in den letzten Jahrzehnten nicht an Aktualität verloren hat. Selbst in Bundesländern mit leistungsstarker Oberstufe zeigt die Mehrheit der Schülerinnen und Schüler Schwierigkeiten im Umgang mit wissenschaftspropädeutischen und studienvorbereitenden Lerninhalten der gymnasialen Oberstufe. In den Grundkursen ist am Ende der gymnasialen Oberstufe ein erheblicher Teil der Schülerinnen und Schüler immer noch nicht den allgemeinbildenden Leistungsanforderungen gewachsen, die für den Mittleren Schulabschlusses erwartet werden. Die bisherigen Ergebnisse geben keinen empirischen Beleg dafür, dass sich der Anteil der leistungsschwachen Schülerinnen und Schüler in den Bundesländern, welche die Kursdifferenzierung abgeschafft haben, bedeutend verringert hat.

Die Diskussion um Struktur und Inhalte der gymnasialen Oberstufe ist in Deutschland häufig von der Zielsetzung der *Allgemeinen Hochschulreife* geprägt (vgl. Neumann und Trautwein 2019). Somit sollen idealerweise alle Abiturientinnen und Abiturienten die Voraussetzungen für alle Hochschulstudiengänge erfüllen. In der aktuell gültigen Vereinbarung zur Gestaltung der Oberstufe ist dieses noch einmal dadurch betont worden, dass es in der Trias beim Aspekt der Studierfähigkeit um eine allgemeine und nicht um eine fachspezifische Studierfähigkeit geht. Das heißt, dass auch Schülerinnen und Schüler der Grundkurse in Mathematik eigentlich in der Lage sein sollten, ein MINT-Studium aufzunehmen. Eine Konsequenz aus diesem Anspruch ist, dass in den Mathematik-Grundkursen im Grunde genommen die gleichen Fachinhalte wie in den Leistungskursen unterrichtet werden, wenn auch im geringeren zeitlichen Umfang und in einer reduzierten Tiefe (vgl. KMK 2002, 2015).

Angesichts der empirischen Daten ist zumindest für die MINT-Studienfächer fraglich, inwieweit das Abitur dem Anspruch als Allgemeine Hochschulreife jemals gerecht geworden ist. Bereits im Jahre 1996 bei TIMSS ließen die Leistungen eines Teils der Schülerinnen und Schüler aus den Grundkursen nicht erwarten, dass sie erfolgreich ein MINT-Studium durchlaufen könnten. Die empirischen Daten der späteren SLS deuten darauf hin, dass – trotz einzelner Bemühungen der Bildungspolitik, die Leistungsstandards in Mathematik in der gymnasialen Oberstufe zu sichern (z. B. Bildungsstandards für die Allgemeine Hochschulreife, Mathematik als verpflichtendes vierstündiges Profulfach auf erhöhtem Niveau, zentrale Abiturprüfungen) – weiterhin ein nicht unerheblicher Teil der Schülerinnen und Schüler zwar erfolgreich das Abitur ablegt, aber nicht über das notwendige mathematische Wissen und Können für ein MINT-Studium verfügt. Vor dem Hintergrund der empirischen Datenlage stellt sich daher die Frage, ob diese Zielsetzung, die auf der Grundlage von fachlichen und fachdidaktischen Überlegungen festgelegt wurde (Stanat et al. 2016), zu hoch angesetzt ist (vgl. Köller 2019). Nach Ansicht von Baptist und Winter (2001) liegt dabei auch ein Missverständnis von „Wissenschaftsorientierung“ und „Allgemeiner Hochschulreife“ vor: Eine sichere Beherrschung von grundlegenden Fähigkeiten sei sicherlich eine bessere Studienvorbereitung als sehr spezifische Fähigkeiten in einzelnen Fachgebieten (Baptist und Winter 2001).

Für den Mathematikunterricht in der gymnasialen Oberstufe würde sich dann die Frage stellen, welche grundlegenden mathematischen Fähigkeiten vermittelt werden sollten. Handelt es sich hierbei eher um die Beherrschung von Integrationsregeln oder um elementare algebraische Grundfertigkeiten? Wie die Studie *Mathematische Lernvoraussetzungen für MINT-Studiengänge* (MaLeMINT, Neumann et al. 2017) zeigte, erwarten die Hochschuldozierenden der MINT-Fächer nicht unbedingt die Beherrschung von elaborierten fachspezifischen Unterrichtsinhalten der gymnasialen Oberstufe (z. B. partielle Integration), sondern vor allem die sichere Beherrschung elementarer algebraischer Grundfertigkeiten der Sekundarstufe I (z. B. Termumformen, Lösen linearer und quadratischer Gleichungen). Vor diesem Hintergrund wäre eine Sicherung dieser Grundfertigkeiten in der gymnasialen Oberstufe für einen Teil der Schülerschaft möglicherweise sogar eine bessere Studienvorbereitung als die intensive Behandlung originärer Oberstoffinhalte.

Außerdem könnte ein stärker zieldifferenter Mathematikunterricht den Begabungen und Interessen der Schülerinnen und Schüler möglicherweise in größerem Maße gerecht werden. Schülerinnen und Schüler mit einem mathematisch-naturwissenschaftlichen Begabungs- und Interessenschwerpunkt könnten dabei die Möglichkeit erhalten, wie klassischerweise in den Leistungskursen geschehen, sich zeitlich und inhaltlich vertieft mit voruniversitärer Mathematik zu beschäftigen und auf ein MINT-Studium vorzubereiten. Die Daten zeigen, dass die differenzierte gymnasiale Oberstufe dies leisten kann. Schülerinnen und Schüler mit einem sprachlichen, künstlerischen oder gesellschaftswissenschaftlichen Begabungs- und Interessensprofil könnten im Mathematikunterricht dagegen verstärkt die Möglichkeit erhalten, sich auf lebensweltliche Anforderungen vorzubereiten, für deren Bewältigung vertieftes anwendungsbezogenes mathematisches Wissen und Können zu fachspezifischen Inhalten der Sekundarstufe I hilfreich ist. Dieses hätte allerdings zur Konsequenz, dass die voruniversitäre Mathematik für diese Schülerschaft im Unterricht nicht mehr die

herausgehobene Rolle einnimmt, die sie heute für alle Schülerinnen und Schüler der gymnasialen Oberstufe gemäß der Bildungsstandards hat bzw. haben sollte. Entsprechend der Daten wurde dieser generelle Anspruch aber bisher auch mehrheitlich nicht erreicht.

6.2 Prozedural-technische Fertigkeiten

Die berichteten Ergebnisse der SLS basierten fast ausschließlich auf den Instrumenten aus TIMSS oder NEPS. Rückschlüsse auf prozedural-technische Fertigkeiten (z.B. Gleichungslösen) der Absolventinnen und Absolventen lassen diese Instrumente nur eingeschränkt zu. Der LAU/KESS-Test zum Basiswissen und zu den Basisfertigkeiten beinhaltet zwar diese spezifische inhaltliche Facette. Die Testhefte, die in den Abschlussjahrgängen eingesetzt wurden, enthielten hierzu aber nur wenige Aufgaben. Daher ist aus den empirischen Daten der SLS nahezu nicht zu entnehmen, in welcher Weise am Ende der gymnasialen Oberstufe innermathematische prozedural-technische Fertigkeiten beherrscht werden und ob heute Abiturientinnen und Abiturienten gerade in diesem Bereich geringere Fähigkeiten aufweisen als früher.

Aufgrund der eingeschränkten inhaltlichen Breite der eingesetzten Instrumente, die durch beschränkte Testzeiten und andere Rahmenbedingungen zustande kam, muss somit auch weiterhin offen bleiben, wie sich die angestrebte verstärkte Kompetenzorientierung auf die prozedural-technischen Fertigkeiten der Schülerinnen und Schüler in der gymnasialen Oberstufe ausgewirkt hat. Aufgrund der spärlichen Datentlage zur gymnasialen Oberstufe allgemein und SLS speziell können noch nicht einmal Aussagen darüber gemacht werden, ob und in welcher Weise überhaupt ein kompetenzorientierter Mathematikunterricht in der gymnasialen Oberstufe erfolgt. Da Veränderungsprozesse im Schulsystem träge sind, ist nicht auszuschließen, dass das implementierte vom intendierten Curriculum abweicht und sich die inhaltliche Gestaltung in den letzten 25 Jahren weniger verändert hat als man aufgrund der organisatorischen Veränderungen vorschnell annehmen könnte.

6.3 Limitationen

Die im Artikel berichteten Ergebnisse unterliegen Limitationen. Aus theoretischer Perspektive stellt sich die Frage, inwieweit die eingesetzten Tests Konstruktvalidität beanspruchen können. Testinstrumente decken naturgemäß nur Facetten eines komplexen Spektrums an mathematischen Leistungen ab. Daher können einzelne Instrumente nicht den Anspruch haben, mathematische Leistungen in toto zu erfassen. In dem vorliegenden Artikel wurde daher versucht, durch Itembeispiele und eine Beschreibung der Konstrukte und Instrumente transparent zu machen, welche Facette mathematischer Leistungen jeweils getestet wurde. Prinzipiell können daher zunächst nur Aussagen zu den Facetten getätigt werden, die auch gemessen wurden. In welcher Weise Schülerinnen und Schüler beispielsweise mit fachspezifischen Inhalten der Oberstufe in lebensweltlichen Zusammenhängen (z.B. Ableitung als lokale Änderungsrate, Integral als rekonstruierter Bestand) umgehen können, lässt sich aus den eingesetzten Instrumenten nicht ableiten. Die Tests zur mathematischen

Grundbildung (NEPS und TIMSS) beschränken sich vornehmlich auf Lerninhalte der Sekundarstufe I, wohingegen der TIMSS-Test zur voruniversitären Mathematik zwar originäre Oberstufeninhalte umfasst, diese sich aber vornehmlich auf traditionelle innermathematische Aufgabenstellungen konzentrieren. Begrenzt wird diese Einschränkung allerdings durch die Tatsache, dass Facetten mathematischer Leistungen in empirischen Studien häufig hohe Korrelationen aufweisen. Unter der Voraussetzung hoher Korrelationen können auf Gruppenebene auch mit der Erfassung nur einer Facette begrenzte Aussagen über die gemessene Facette hinaus getätigt werden.

Zudem ist die Setzung von Mindestniveaus bei Instrumenten normativer Natur und keine empirische Frage. Die dem Artikel zugrunde gelegten Mindesterwartungen in den Niveaustufenmodellen wurden von Expertinnen und Experten explizit in Hinblick auf die gymnasiale Oberstufe formuliert. Nichtsdestotrotz würden andere normative Erwartungen zu anderen Ergebnissen und Schlussfolgerungen führen.

Methodisch wurde die Zuordnung der Schülerleistungen aus einigen SLS zu den Leistungsstufen anhand berichteter Summenscores (z. B. BIJU, ZAS, TOSCA-Sachsen) berechnet (vgl. Online-Supplement, Kap. A). Somit sind Abweichungen der berichteten Verteilung auf die Leistungsstufen von der tatsächlichen Populationsverteilung möglich. Allerdings sind substantielle Abweichungen von den hier berichteten Ergebnissen nicht zu erwarten, sondern eher Unterschiede im Bereich von wenigen Prozentpunkten.

Die Verlinkung der NEPS- bzw. LAU-Studien und damit der Vergleich der Ergebnisse aus verschiedenen Studien ist nur möglich, wenn Linking-Items genutzt werden. Dabei treten für einzelne dieser Items deutliche Differenzen in der Itemschwierigkeit auf, was die vorausgesetzte Messinvarianz verletzt. Das bedeutet, dass die ermittelten Unterschiedswerte von den verwendeten Linking-Items und deren Sensitivität für längsschnittliche oder querschnittliche Gruppenunterschiede abhängig sind. Für zukünftige Studien ist entsprechend auf eine ausreichende Anzahl an messinvarianten Linking-Items zu achten.

Beim Vergleich der Studien muss zudem berücksichtigt werden, dass die Anzahl der eingesetzten Items in den SLS nicht unerheblich schwankt (vgl. Tab. 1), was sich auf die Validität und die Reliabilität auswirken kann. Außerdem muss offen bleiben, inwiefern sich Aussagen zu heutigen Schülerleistungen in voruniversitärer Mathematik aus den verwendeten Daten schlussfolgern lassen, da die KESS-Studie aus Hamburg im Jahre 2012 die jüngste SLS ist, die explizit fachspezifische Lerninhalte der gymnasialen Oberstufe testete. Daher bieten die Ergebnisse aus den SLS bisher nur spezifische Momentaufnahmen zu den mathematischen Leistungen in der gymnasialen Oberstufe.

6.4 Forschungsdesiderate

Wie bereits angedeutet, kann die Forschungslage zum Stand des mathematischen Wissens und Könnens in der gymnasialen Oberstufe in Deutschland als defizitär angesehen werden (vgl. auch Neumann und Trautwein 2019). Im Lichte der wiederkehrenden Diskussionen um die Leistungsfähigkeit der gymnasialen Oberstufe ist das Unwissen über die aktuelle Lage besonders bedauerlich. Diese Ausgangssituati-

on wird sich voraussichtlich jedoch in naher Zukunft nicht ändern. Ein Bildungsmonitoring, das ähnlich wie der Bildungstrend für die Primarstufe und Sekundarstufe I aufzeigen kann, ob die Bildungsstandards für die Allgemeine Hochschulreife erreicht werden, ist weiterhin nicht geplant (vgl. Neumann und Trautwein 2019; Stanat et al. 2016). Auch eine erneute Teilnahme an einer internationalen Schulvergleichsstudie in der Oberstufe (z. B. TIMSS Advanced) ist nicht zu erwarten.

Angesichts dieser Rahmenbedingungen erscheinen die Bedingungen für eine intensivere Beforschung der gymnasialen Oberstufe auf der Systemebene zurzeit nicht besonders günstig zu sein. Dennoch sollen im Folgenden einige Ansätze für wichtige Forschungsthemen formuliert werden. So wäre es aus mathematikdidaktischer Perspektive wünschenswert, die Mathematikleistungen in der Oberstufe in einer größeren Breite zu erfassen. Hierzu bedarf es auch einer noch intensiveren bildungstheoretischen Erörterung, welche spezifischen Lerninhalte sich für den Mathematikunterricht der Oberstufe aus der Trias der Bildungsziele vertiefte Allgemeinbildung, Wissenschaftspropädeutik und Studierfähigkeit ableiten lassen. Kurze Instrumente, welche die mathematische Grundbildung fokussieren (z. B. NEPS-TH, NEPS-BW, LISA-6), liefern zwar ein Globalmaß, das für bestimmte bildungswissenschaftliche Fragen geeignet ist. Schlussfolgerungen auf vorhandenes Wissen und erworbene Fähigkeiten in den zentralen Oberstufeninhalten Analysis, Analytische Geometrie und Stochastik sind anhand dieser beschränkten Tests aber nur sehr eingeschränkt bzw. gar nicht möglich. Es wäre daher wünschenswert, valide und reliable Instrumente für die Oberstufe zu entwickeln und einzusetzen, die differenzierte Aussagen sowohl zu den mathematischen Leistungen hinsichtlich der Lerninhalte der Sekundarstufe I als auch der Sekundarstufe II und zur Trias der übergreifenden Bildungsziele ermöglichen.

Es stellt sich außerdem die Frage, warum so große Leistungsdisparitäten zwischen Bundesländern (z. B. Hamburg und Baden-Württemberg) oder Zweigen der gymnasialen Oberstufe (Allgemeinbildendes Gymnasium und nicht-technisches Berufliches Gymnasium/Gesamtschule) bestehen. Allgemeiner stellt sich die Frage, welche individuellen und unterrichtlichen Bedingungsfaktoren zu einer Leistungsentwicklung und zu einem kumulativen Wissensaufbau in der gymnasialen Oberstufe führen. Um dem nachzugehen, wären daher in stärkerem Maße Längsschnittstudien wünschenswert. Mit einem experimentellen Anteil in Substichproben böten sie zudem die Möglichkeit, mehr über Wirkungszusammenhänge zu erfahren und über rein deskriptive Befunde hinauszugehen. Beobachtungs- oder Befragungsstudien ermöglichen darüber hinaus Einblicke in die tatsächliche Unterrichtspraxis der gymnasialen Oberstufe.

Vor dem Hintergrund der bisherigen empirischen Ergebnisse stellt sich zudem die bildungstheoretische Frage, welche Mindestniveaus im Mathematikunterricht der gymnasialen Oberstufe erreicht werden sollten. Die in der Sekundäranalyse zugrunde gelegten normativen Mindestniveaus basierten auf Einschätzungen von Expertinnen und Experten und erschienen den Verfassenden dieses Artikels in ihrer grundsätzlichen Ausrichtung plausibel. Wünschenswert wäre aber eine noch intensivere wissenschaftliche Auseinandersetzung mit diesem Thema, indem beispielsweise der Katalog von Basiskompetenzen (Drüke-Noe et al. 2011) auf die gymnasiale Oberstufe erweitert und die Diskussion um mathematisches Grundwissen und -kөн-

nen in der Sekundarstufe II (Bruder et al. 2015; Feldt-Caesar 2017; Pinkernell et al. 2015) fortgeführt wird. Außerdem bietet die Rückkehr einiger Bundesländer zum niveaudifferenzierten Kurssystem Anlass, auch aus mathematikdidaktischer Perspektive die bildungspolitische Frage wieder aufzunehmen (vgl. bereits Heymann 1996; KMK 1995), in welchem Ausmaß eine Adjustierung der Anforderungen bzw. eine Spezialisierung auf mathematische Begabungs- und Interessenprofile in der gymnasialen Oberstufe ermöglicht und gefördert werden sollte.

7 Fazit

Das Review konnte aufzeigen, dass seit 1995 zwar einige Schulleistungsstudien zu Mathematikleistungen in der gymnasialen Oberstufe durchgeführt wurden, diese aber zumeist auf einzelne Bundesländer beschränkt waren. Außerdem bildeten die verwendeten Instrumente jeweils nur einen besonderen Ausschnitt der Mathematikleistungen ab und ließen wichtige Aspekte (z. B. prozedural-technische Fertigkeiten) nahezu unberücksichtigt. Die Sekundäranalyse verdeutlichte, dass ausnahmslos in *allen* untersuchten SLS große Teile der Schülerschaft im Bereich der Wissensschaffspropädeutik und Studierfähigkeit Bildungsziele der gymnasialen Oberstufe in voruniversitärer Mathematik nicht erreichten. Lediglich in den Leistungskursen stellte sich die Situation günstiger dar. Trotzdem muss auch festgestellt werden, dass die derzeitige Datenlage nicht ausreicht, um für Deutschland umfassende deskriptive Aussagen über den Leistungsstand der Schülerinnen und Schüler in der gymnasialen Oberstufe zu tätigen.

Darüber hinaus bleibt aus empirischer Sicht bisher unklar, wie aktuell Mathematikunterricht in der Oberstufe überhaupt gestaltet wird und was günstige unterrichtliche Bedingungen für eine positive Leistungsentwicklung in Mathematik in der gymnasialen Oberstufe sind. Dies ist vor dem Hintergrund der Kritik von Hochschulseite an den mathematischen Fähigkeiten, mit denen Studienanfängerinnen und Studienanfänger ins Studium starten, und der damit verbundenen Spekulationen über Ursachen und vermutete Veränderungen in den letzten Jahren eine problematische Situation. Empfehlungen zum Mathematikunterricht in der gymnasialen Oberstufe (z. B. Mathematik-Kommission Übergang Schule-Hochschule et al. 2019) können bisher fast ausschließlich auf Erfahrungswissen und theoretische Annahmen zurückgreifen. Daher sind empirische Erkenntnisse über den tatsächlichen Leistungsstand und über Wirkmechanismen notwendig, um evidenzbasierte Lösungen für aktuelle Probleme des Mathematikunterrichts der gymnasialen Oberstufe generieren zu können.

Danksagung Wir danken Nele Kampa, Katharina Maag-Merki und Wolfgang Wagner für die weitergehenden Informationen (z. B. verwendete Items) zu den von ihnen durchgeführten Studien. Außerdem sind wir Gabriel Nagy und Alexander Robitzsch für methodische Ratschläge bezüglich der Sekundäranalysen zu Dank verpflichtet.

Funding Open Access funding enabled and organized by Projekt DEAL.

Open Access Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in

jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

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

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

Literatur

- Baptist, P., & Winter, H. (2001). Überlegungen zur Weiterentwicklung des Mathematikunterrichts in der Oberstufe des Gymnasiums. In H.-E. Tenorth (Hrsg.), *Kerncurriculum Oberstufe. Mathematik, Deutsch, Englisch; Expertisen* (S. 54–76). Weinheim: Beltz.
- Baumert, J., Bos, W., Klieme, E., Lehmann, R., Lehrke, M., Hosenfeld, I., Watermann, R., et al. (1999). *Testaufgaben zu TIMSS/III: Mathematisch-naturwissenschaftliche Grundbildung und voruniversitäre Mathematik und Physik der Abschlussklassen der Sekundarstufe II (Population 3)*. Berlin: Max-Planck-Institut für Bildungsforschung.
- Baumert, J., Bos, W., & Lehmann, R. (Hrsg.). (2000a). *TIMSS/III Dritte Internationale Mathematik- und Naturwissenschaftsstudie – Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn* (Bd. 1: Mathematische und naturwissenschaftliche Grundbildung am Ende der Pflichtschulzeit). Opladen: Leske + Budrich.
- Baumert, J., Bos, W., & Lehmann, R. (Hrsg.). (2000b). *TIMSS/III Dritte Internationale Mathematik- und Naturwissenschaftsstudie – Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn* (Bd. 2: Mathematische und physikalische Kompetenzen am Ende der gymnasialen Oberstufe). Opladen: Leske + Budrich.
- Baumert, J., Bos, W., & Watermann, R. (2000c). Fachleistungen im voruniversitären Mathematik- und Physikunterricht im internationalen Vergleich. In J. Baumert, W. Bos & R. Lehmann (Hrsg.), *TIMSS/III Dritte Internationale Mathematik- und Naturwissenschaftsstudie – Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn* (Bd. 2: Mathematische und physikalische Kompetenzen am Ende der gymnasialen Oberstufe, S. 129–180). Opladen: Leske + Budrich.
- Baumert, J., Gruehn, S., Heyn, S., Köller, O., Schnabel, K.-U., Leven, I., et al. (2003). *Bildungsverläufe und psychosoziale Entwicklung im Jugend- und jungen Erwachsenenalter (BIJU). Dokumentation der eingesetzten Items und Skalen, Band 2, Längsschnitt I, Welle 5*. Berlin: Max-Planck-Institut für Bildungsforschung.
- Beaton, A.E., & Allen, N.L. (1992). Interpreting scales through scale anchoring. *Journal of Educational and Behavioral Statistics*, 17(2), 191–204.
- Blossfeld, H.-P., Roßbach, H.-G. & von Maurice, J. (Hrsg.) (2011). Education as a lifelong process—The German National Educational Panel Study (NEPS). *Zeitschrift für Erziehungswissenschaft: Sonderheft 14*. <https://doi.org/10.1007/s11618-011-0179-2>.
- Bruder, R., Feldt-Caesar, N., Pallack, A., Pinkernell, G., & Wynands, A. (2015). Mathematisches Grundwissen und Grundkönnen in der Sekundarstufe II. In W. Blum, S. Vogel, C. Drüke-Noe & A. Roppelt (Hrsg.), *Bildungsstandards aktuell: Mathematik in der Sekundarstufe II* (S. 108–124). Braunschweig: Bildungshaus Schulbuchverlage.
- Cizek, G.J. (Hrsg.). (2012). *Setting performance standards. Foundations, methods, and innovations* (2. Aufl.). New York: Routledge.
- DMV (Deutsche Mathematiker-Vereinigung), DPG, GDCh, MNU, & VDB (1982). Rettet die mathematisch-naturwissenschaftliche Bildung. *Physikalische Blätter*, 38(1), 25.
- Drüke-Noe, C., Möller, G., Pallack, A., Schmidt, S., Schmidt, U., Sommer, N., et al. (2011). *Basiskompetenzen Mathematik. Für Alltag und Berufseinstieg am Ende der allgemeinen Schulpflicht*. Berlin: Cornelsen.
- Duchhardt, C. (2015). *NEPS technical report for mathematics: scaling results for the additional study Baden-Wuerttemberg*. Bamberg: Leibniz Institute for Educational Trajectories.

- Ehmke, T., Duchhardt, C., Geiser, H., Grüßing, M., Heinze, A., & Marschick, F. (2009). Kompetenzentwicklung über die Lebensspanne – Erhebung von mathematischer Kompetenz im Nationalen Bildungspanel. In A. Heinze & M. Grüßing (Hrsg.), *Mathematiklernen vom Kindergarten bis zum Studium. Kontinuität und Kohärenz als Herausforderung beim Mathematiklernen* (S. 313–327). Münster: Waxmann.
- Feddermann, M., Guill, K., List, M. K., Matthießen, R., Ömeroğulları, M., Stallasch, S. E., et al. (2019). *KESS – Skalierung der Leistungstests*. Kiel: IPN – Leibniz-Institut für die Pädagogik der Naturwissenschaften und Mathematik.
- Feldt-Caesar, N. (2017). *Konzeptualisierung und Diagnose von mathematischem Grundwissen und Grundkönnen*. Wiesbaden: Springer.
- Fischer, R. (2012). Fächerorientierte Allgemeinbildung: Entscheidungskompetenz und Kommunikationsfähigkeit mit ExpertInnen. In R. Fischer, U. Greiner & H. Bastel (Hrsg.), *Domänen fächerorientierter Allgemeinbildung* (S. 9–17). Linz: Trauner.
- Fischer, L., Rohm, T., & Gnams, T. (2017). *NEPS technical report for mathematics: Scaling results of starting cohort 4 for grade 12*. Bamberg: Leibniz Institute for Educational Trajectories.
- Foy, P., & Arora, A. (2009). *TIMSS advanced 2008 user guide for the international database. Released Items. Advanced mathematics*. Chestnut Hill: Boston College.
- Hahn, S. (2013). Wissenschaftspropädeutik in der gymnasialen Oberstufe. In D. Bosse, F. Eberle & B. Schneider-Taylor (Hrsg.), *Standardisierung in der gymnasialen Oberstufe* (S. 161–174). Wiesbaden: Springer.
- Heymann, H. W. (1996). Mathematikunterricht in der Gymnasialen Oberstufe. *Zeitschrift für Pädagogik*, 42(4), 541–556.
- Huber, L. (2009). Von „basalen Fähigkeiten“ bis „vertiefte Allgemeinbildung“: Was sollen Abiturientinnen und Abiturienten für das Studium mitbringen? In D. Bosse (Hrsg.), *Gymnasiale Bildung zwischen Kompetenzorientierung und Kulturarbeit* (S. 107–124). Wiesbaden: VS.
- Kahnert, J. (2014). *Das Zentralabitur im Fach Mathematik. Eine empirische Analyse von Abitur- und TIMSS-Daten im Vergleich*. Münster: Waxmann.
- Kampa, N., Köller, O., Schmidt, F. T. C., & Leucht, M. (2016). Kompetenzen im Fach Mathematik. In M. Leucht, N. Kampa & O. Köller (Hrsg.), *Fachleistungen von Abiturienten. Vergleich allgemeinbildender und beruflicher Gymnasien in Schleswig-Holstein: LISA 6* (S. 119–143). Münster: Waxmann.
- Kampa, N., Hinz, H., Haag, N., & Köller, O. (2018). Standardbezogene Kompetenzen im Fach Mathematik am Ende der gymnasialen Oberstufe. *Zeitschrift für Erziehungswissenschaft*, 21(1), 121–141.
- Klieme, E. (2000). Fachleistungen im voruniversitären Mathematik- und Physikunterricht: Theoretische Grundlagen, Kompetenzstufen und Unterrichtsschwerpunkte. In J. Baumert, W. Bos & R. Lehmann (Hrsg.), *TIMSS/III Dritte Internationale Mathematik- und Naturwissenschaftsstudie – Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn* (Bd. 2: Mathematische und physikalische Kompetenzen am Ende der gymnasialen Oberstufe, S. 57–128). Opladen: Leske + Budrich.
- Klieme, E., Baumert, J., Köller, O., & Bos, W. (2000). Mathematische und naturwissenschaftliche Grundbildung: Konzeptuelle Grundlagen und die Erfassung und Skalierung von Kompetenzen. In J. Baumert, W. Bos & R. Lehmann (Hrsg.), *TIMSS/III Dritte Internationale Mathematik- und Naturwissenschaftsstudie – Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn*. (Bd. 1: Mathematische und naturwissenschaftliche Grundbildung am Ende der Pflichtschulzeit, S. 85–133). Opladen: Leske + Budrich.
- KMK (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland) (1972). *Beschlüsse der Kultusministerkonferenz. Vereinbarungen zur Neugestaltung der gymnasialen Oberstufe in der Sekundarstufe II*. Bonn: KMK. https://www.kmk.org/fileadmin/Dateien/pdf/Bildung/AllgBildung/176_Vereinb_Gestalt_Gym_Ob_Sek_II-1972_01.pdf
- KMK (1995). *Weiterentwicklung der Prinzipien der gymnasialen Oberstufe und des Abiturs. Abschlussbericht der von der Kultusministerkonferenz eingesetzten Expertenkommission*. Kiel: Schmidt & Klau-nig.
- KMK (2002). *Einheitliche Prüfungsanforderungen in der Abiturprüfung Mathematik. (Beschluss der Kultusministerkonferenz vom 01.12.1989 i.d.F. vom 24.05.2002)*. Bonn: KMK.
- KMK (2004). *Beschlüsse der Kultusministerkonferenz. Bildungsstandards im Fach Mathematik für den Mittleren Schulabschluss. Beschluss vom 04.12.2003*. München: Wolters Kluwer.
- KMK (2015). *Bildungsstandards im Fach Mathematik für die Allgemeine Hochschulreife (Beschluss der Kultusministerkonferenz vom 18.10.2012)*. Köln: Wolters Kluwer.
- KMK (2018). *Vereinbarung zur Gestaltung der gymnasialen Oberstufe und der Abiturprüfung. (Beschluss der Kultusministerkonferenz vom 07.07.1972 i.d.F. vom 15.02.2018)*. Berlin: KMK.

- Kock, A.-L., Petersen, L. A., & Litteck, K. (in Vorbereitung). *NEPS technical report for mathematics: Scaling results of the additional study Thuringia*. Bamberg: Leibniz Institute for Educational Trajectories.
- Koepf, W., & Kramer, J. (2015). Der Beitrag der Bildungsstandards zum Übergang Sekundarstufe II – Universität. In W. Blum, S. Vogel, C. Drüke-Noe & A. Roppelt (Hrsg.), *Bildungsstandards aktuell: Mathematik in der Sekundarstufe II* (S. 125–131). Braunschweig: Bildungshaus Schulbuchverlage.
- Köller, O. (2019). Was nützen Ziele, wenn man sie nicht erreicht? Schulische Leistungen am Ende der gymnasialen Oberstufe. In L. Haag & K. Zierer (Hrsg.), *Jahrbuch für Allgemeine Didaktik 2018: Thementeil: Unterrichten wir das „Richtige“? – Die Fragen nach zeitgemäßen Bildungsinhalten in der Schule* (S. 105–122). Baltmannsweiler: Schneider Hohengehren.
- Köller, O., Baumert, J., & Schnabel, K. (2000). Zum Zusammenspiel von schulischem Interesse und Lernen im Fach Mathematik: Längsschnittanalysen in den Sekundarstufen I und II. In U. Schiefele & K.-P. Wild (Hrsg.), *Interesse und Lernmotivation. Untersuchungen zu Entwicklung, Förderung und Wirkung* (S. 163–181). Münster: Waxmann.
- Lehmann, R. H., Hunger, S., Ivanov, S., Gänsfuß, R., & Hoffmann, E. (2012a). LAU 11. Aspekte der Lernausgangslage und der Lernentwicklung – Klassenstufe 11 – Ergebnisse einer längsschnittlichen Untersuchung in Hamburg. In Behörde für Schule & Berufsbildung (Hrsg.), *LAU – Aspekte der Lernausgangslage und der Lernentwicklung*. Münster: Waxmann.
- Lehmann, R. H., Vieluf, U., Nikolova, R., & Ivanov, S. (2012b). LAU 13 – Aspekte der Lernausgangslage und der Lernentwicklung – Klassenstufe 13 – Erster Bericht. In Behörde für Schule & Berufsbildung (Hrsg.), *LAU – Aspekte der Lernausgangslage und der Lernentwicklung* (S. 151–231). Münster: Waxmann.
- List, M. K., Nagy, G., Guill, K., & Köller, O. (2018). *LAU – Skalierung der Leistungstests*. Kiel: IPN – Leibniz-Institut für die Pädagogik der Naturwissenschaften und Mathematik.
- Maag Merki, K. (2012). Die Leistungen der Gymnasiastinnen und Gynmasiasten in Mathematik und Englisch. In K. Maag Merki (Hrsg.), *Zentralabitur* (S. 263–292). Wiesbaden: VS.
- Mathematik-Kommission Übergang Schule–Hochschule, Deutsche Mathematiker-Vereinigung, Gesellschaft für Didaktik der Mathematik & Verband zur Förderung des MINT-Unterrichts (2019). *Mathematik: 19 Maßnahmen für einen konstruktiven Übergang Schule – Hochschule. Mitteilungen der Gesellschaft für Didaktik der Mathematik*, 107, 6–7.
- Nagy, G., Neumann, M., Becker, M., Watermann, R., Köller, O., & Lüdtkke, O. (2007). Mathematikleistungen am Ende der Sekundarstufe II. In U. Trautwein, O. Köller, R. Lehmann & O. Lüdtkke (Hrsg.), *Schulleistungen von Abiturienten. Regionale, schulformbezogene und soziale Disparitäten* (S. 71–112). Münster: Waxmann.
- Nagy, G., Neumann, M., Trautwein, U., & Lüdtkke, O. (2010). Voruniversitäre Mathematikleistungen vor und nach der Neuordnung der gymnasialen Oberstufe in Baden-Württemberg. In U. Trautwein, M. Neumann, G. Nagy, O. Lüdtkke & K. Maaz (Hrsg.), *Schulleistungen von Abiturienten. Die neu geordnete gymnasiale Oberstufe auf dem Prüfstand* (S. 147–180). Wiesbaden: VS.
- Neumann, M. (2010). Innovation und Restauration – Die (Rück-?)Reform der gymnasialen Oberstufe in Baden-Württemberg. In U. Trautwein, M. Neumann, G. Nagy, O. Lüdtkke & K. Maaz (Hrsg.), *Schulleistungen von Abiturienten. Die neu geordnete gymnasiale Oberstufe auf dem Prüfstand* (S. 37–90). Wiesbaden: VS.
- Neumann, M., & Nagy, G. (2010). Mathematische und naturwissenschaftliche Grundbildung vor und nach der Neuordnung der gymnasialen Oberstufe in Baden-Württemberg. In U. Trautwein, M. Neumann, G. Nagy, O. Lüdtkke & K. Maaz (Hrsg.), *Schulleistungen von Abiturienten. Die neu geordnete gymnasiale Oberstufe auf dem Prüfstand* (S. 215–241). Wiesbaden: VS.
- Neumann, M., & Trautwein, U. (2019). Sekundarbereich II und der Erwerb der Hochschulzugangsberechtigung. In O. Köller, M. Hasselhorn, F. W. Hesse, K. Maaz, J. Schrader, H. Solga, et al. (Hrsg.), *Das Bildungswesen in Deutschland. Bestand und Potenziale* (S. 533–564).
- Neumann, I., Pigge, C., & Heinze, A. (2017). *Welche mathematischen Lernvoraussetzungen erwarten Hochschullehrende für ein MINT-Studium?* Kiel: IPN – Leibniz-Institut für die Pädagogik der Naturwissenschaften und Mathematik.
- Petersen, L. A., Litteck, K., & Rohenroth, D. (2020). *NEPS technical report for mathematics: scaling results of starting cohort 3 for grade 12*. Bamberg: Leibniz Institute for Educational Trajectories.
- Pinkernell, G., Elschenbroich, H.-J., Heintz, G., Körner, H., Langlotz, H., & Pallack, A. (2015). *Grundlegendes Wissen und Können am Ende der Sekundarstufe II: Zentrale Begriffe und Verfahren beherrschen und verstehen*. Düsseldorf: MNU.
- Schnittjer, I., & Duchhardt, C. (2015). *Mathematical competence: framework and exemplary test items*. Bamberg: Leibniz Institute for Educational Trajectories.

- Stanat, P., Becker-Mrotzek, M., Blum, W., & Tesch, B. (2016). Vergleichbarkeit in der Vielfalt. Bildungsstandards der Kultusministerkonferenz für die Allgemeine Hochschulreife. In J. Kramer, M. Neumann & U. Trautwein (Hrsg.), *Abitur und Matura im Wandel* (S. 29–58). Wiesbaden: Springer.
- Steiner, H.-G. (1984). Mathematisch-naturwissenschaftliche Bildung – Kritisch-konstruktive Fragen und Bemerkungen zum Aufruf einiger Fachverbände. In M. Reiß & H.-G. Steiner (Hrsg.), *Mathematikkenntnisse – Leistungsmessung – Studierfähigkeit* (S. 5–58). Köln: Aulis.
- Vieluf, U., Ivanov, S., & Nikolova, R. (Hrsg.). (2011). *KESS 10/11. Kompetenzen und Einstellungen von Schülerinnen und Schülern an Hamburger Schulen am Ende der Sekundarstufe I und zu Beginn der gymnasialen Oberstufe*. Münster: Waxmann.
- Vieluf, U., Ivanov, S., & Nikolova, R. (2014). *KESS 12/13. Kompetenzen und Einstellungen von Schülerinnen und Schülern an Hamburger Schulen am Ende der gymnasialen Oberstufe*. Hamburg: Autoren. <https://bildungsserver.hamburg.de/contentblob/4396048/6b49c68061321ae400aaa4f7250ebe9f/data/kess12-13.pdf>
- Wagner, W., Rose, N., Dicke, A.-L., Neumann, M., & Trautwein, U. (2014). Alle alles lehren – Schulleistungen in Englisch, Mathematik und den Naturwissenschaften vor und nach der Neuordnung der gymnasialen Oberstufe in Sachsen. *Zeitschrift für Erziehungswissenschaft*, 17(2), 345–369.
- Watermann, R., & Baumert, J. (2000). Mathematische und naturwissenschaftliche Grundbildung beim Übergang von der Schule in den Beruf. In J. Baumert, W. Bos & R. Lehmann (Hrsg.), *TIMSS/III Dritte Internationale Mathematik- und Naturwissenschaftsstudie – Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn*. (Bd. 1: Mathematische und naturwissenschaftliche Grundbildung am Ende der Pflichtschulzeit, S. 199–259). Opladen: Leske + Budrich.
- Watermann, R., Nagy, G., & Köller, O. (2004). Mathematikleistungen in allgemein bildenden und beruflichen Gymnasien. In O. Köller, R. Watermann, U. Trautwein & O. Lüdtke (Hrsg.), *Wege zur Hochschulreife in Baden-Württemberg. TOSCA – Eine Untersuchung an allgemein bildenden und beruflichen Gymnasien* (S. 205–326). Opladen: Leske + Budrich.
- Winter, H. (1995). Mathematikunterricht und Allgemeinbildung. *Mitteilungen der Gesellschaft für Didaktik der Mathematik*, 61(4), 37–45.