

Denise Depping, Markus Lücken, Frank Musekamp  
& Franziska Thonke

## **Kompetenzstände Hamburger Schüler\*innen vor und während der Corona-Pandemie**

---

### **Zusammenfassung**

*Im Mittelpunkt dieses Beitrages steht die Gegenüberstellung der Kompetenzstände vergleichbarer Schüler\*innen-Kohorten in den Bereichen Deutsch-Leseverstehen und Mathematik vor und während der Corona-Pandemie. Als Ersatz für die verpflichtenden und jährlich durchgeführten KERMIT-Testungen (Kompetenzen ermitteln) wurden den Hamburger Schulen die Testinstrumente LERNSTAND und FLIP (Flexible Lernstandserhebung in Pandemiezeiten) zur Messung der Lese- und Mathematikkompetenz zur Verfügung gestellt. FLIP und LERNSTAND ermöglichen es, die Kompetenzstände der Schüler\*innen ausgewählter Klassenstufen des Schuljahres 2020/21 mit denen des Schuljahres 2019/20 zu vergleichen und auf diese Weise abzuschätzen, ob die Corona-bedingten Veränderungen im Unterrichtsgeschehen zu geringeren Kompetenzständen und/oder einer Vergrößerung sozialer Disparitäten geführt haben. Im Beitrag werden die LERNSTAND-Ergebnisse in Jahrgang 4 und 5 berichtet und mit den Ergebnissen aus KERMIT 3 und KERMIT 5 des Jahres 2019 verglichen.*

*Schlüsselwörter: Kompetenzmessung, Corona-bedingte Schulschließungen, Kohortenvergleich, datengestützte Schulentwicklung*

### **Alternative Pupils' Competence Measurement in Hamburg during the Corona Pandemic**

#### **Abstract**

*This article focuses on measures of basic competencies in reading (German) and mathematics, comparing data from learner cohorts before and during the COVID-19 pandemic. As a response to the conditions posed by the pandemic, the test instruments LERNSTAND and FLIP were provided for Hamburg schools as an alternative to the mandatory annual KERMIT tests. FLIP and LERNSTAND allow for a comparison between competence levels of students in selected grades in the academic year of 2020/21 with groups in the academic year of 2019/20. The objective was to assess whether COVID-19-related changes in teaching practices (e.g., distance learning) have*

*led to lower competence levels and/or an increase in social disparities. In the article, LERNSTAND results in grades 4 and 5 from 2020 are reported and compared with the results from KERMIT 3 and KERMIT 5 from 2019.*

*Keywords: competence measurement, Corona-related school closings, cohort comparisons, evidence-based school development*

## 1 Einleitung

Aufgrund des weltweiten Ausbruchs von Covid-19 im Jahr 2020 wurden in vielen Bereichen des gesellschaftlichen Lebens Maßnahmen ergriffen, die die Ausweitung der Pandemie bremsen sollten. Der Bildungssektor war in den meisten Ländern beispielweise durch Schulschließungen und Fernunterricht betroffen, wobei der zeitliche Umfang der Schulschließungen weltweit variierte (vgl. den Beitrag von Blum & Dobrotić in diesem Band). In der öffentlichen Debatte wurden zum Teil erhebliche Lernrückstände der Schüler\*innen und vor allem auch eine Verschärfung bestehender sozialer Disparitäten angenommen. Belastbare empirische Befunde liegen indes kaum vor. Eine erste empirische Untersuchung wurde vor kurzem von Engzell und Kollegen veröffentlicht (Engzell, Frey & Verhagen, 2020). In ihrer Untersuchung von 350.000 niederländischen Schüler\*innen kommen sie zu dem Ergebnis, dass der Corona-bedingte achtwöchige Lockdown im niederländischen Schulsystem zu Einbußen in der Lernentwicklung geführt hat. Am stärksten sind sie bei Schüler\*innen aus Familien mit einem eher bildungsfernen Hintergrund. Diese Schüler\*innen zeigen die geringste Entwicklung in den Bereichen Leseverstehen, Rechtschreibung und Mathematik.

In dem vorliegenden Beitrag soll der Frage nachgegangen werden, inwieweit die Schulschließungen und der hohe Anteil an Fernunterricht vor den Sommerferien mit den Lernständen der Hamburger Schüler\*innen zusammenhängen. Die Analysen werden für die Gesamtgruppen der Schüler\*innen und getrennt nach dem Grad der sozialen Belastung der Schulen durchgeführt. Mit diesem Vorgehen soll geklärt werden, ob Schüler\*innen an Schulen mit höherer sozialer Belastung in anderem Maße von den Corona-bedingten Änderungen im Unterrichtsgeschehen betroffen sind als Schüler\*innen, die eine Schule mit geringerer sozialer Belastung besuchen. Zur Beantwortung dieser Fragen werden die Ergebnisse aus Kompetenzmessungen analysiert, die vom Institut für Bildungsmonitoring und Qualitätsentwicklung (IfBQ) in Hamburg regelmäßig durchgeführt und den Schulen zurückgemeldet werden. Der folgende Abschnitt gibt einen Überblick über die regulären Kompetenzmessungen in Hamburg.

## 2 Kompetenzmessungen in Hamburg (KERMIT)

Im Rahmen der Gesamtstrategie der Kultusministerkonferenz (KMK) zum Bildungsmonitoring werden in Deutschland regelhaft Verfahren zur Beobachtung und Optimierung von Bildungsprozessen durchgeführt (KMK, 2015). Neben der Teilnahme an internationalen Schulleistungsstudien umfasst die Gesamtstrategie unter anderem die regelmäßige Durchführung von Ländervergleichen und die Durchführung von Vergleichsarbeiten in den Jahrgangsstufen 3 und 8 (VERA), an denen nahezu alle Bundesländer teilnehmen. Diese bundesweiten Erhebungen werden vom Institut zur Qualitätsentwicklung im Bildungswesen (IQB) entwickelt und dann von den einzelnen Bundesländern durchgeführt.

In Hamburg werden die Vergleichsarbeiten in der dritten und achten Jahrgangsstufe unter dem Namen KERMIT 3 und KERMIT 8 (Kompetenzen *ermitteln*) durchgeführt. Sie sind Teil eines kohärenten Systems zur Unterstützung der Hamburger Schulen bei der datengestützten Schul- und Unterrichtsentwicklung, das in Deutschland einzigartig ist. Jährlich werden in sechs Jahrgangsstufen die Kompetenzen aller Hamburger Schüler\*innen in den Fächern Deutsch und Mathematik untersucht. Ab der Sekundarstufe I werden zusätzlich die Kompetenzen in Englisch und – mit Ausnahme der Jahrgangsstufe 8 – auch in den Naturwissenschaften erhoben.

In der Primarstufe und in der achten Jahrgangsstufe werden die KERMIT-Erhebungen von Lehrkräften durchgeführt. Für alle anderen KERMIT-Erhebungen werden externe, geschulte Testleitungen im Auftrag des IfBQ zur Testdurchführung an die Schulen entsandt. Die Eingabe bzw. Kodierung der Testaufgaben erfolgt für alle KERMIT-Erhebungen am IfBQ. Tabelle 1 fasst die geplanten Testzeiträume des Jahres 2020 für die KERMIT-Erhebungen und ihren Durchführungstatus zusammen. Für weiterführende Informationen zu den KERMIT-Testungen sei auf den Beitrag von Lücken und Kolleg\*innen verwiesen (Lücken et al., 2014).

Tab. 1: Übersicht über die Hamburger KERMIT-Erhebungen im Jahr 2020

Erhebung	(geplanter) Zeitraum	Durchführungstatus
KERMIT 9	Februar 2020	stattgefunden
KERMIT 8	Februar 2020	stattgefunden
KERMIT 2	April/Mai 2020	entfallen
KERMIT 3	April/Mai 2020	entfallen
KERMIT 5	September 2020	entfallen
KERMIT 7	August/September 2020	entfallen

Quelle: eigene Darstellung

Da durch die Ausbreitung der Corona-Pandemie ab Ende März 2020 bis zu den Sommerferien Ende Juni 2020 der Unterricht in Hamburger Schulen gar nicht bzw. sehr eingeschränkt stattfand, konnten die KERMIT-Erhebungen in der Primarstufe (KERMIT 2 und KERMIT 3) nicht durchgeführt werden. Die Erhebungen KERMIT 5 und 7, die üblicherweise zu Beginn eines Schuljahres geschrieben werden, wurden aufgrund von Risikoabwägungen nicht in der üblichen standardisierten Form mit externen Testleitungen des IfBQ angeboten. Dadurch sollte verhindert werden, dass Testleitungen zu sogenannten Superspreadern werden, die das Virus in den Schulen verteilen.

### **3 Alternative Kompetenzmessungen während der Pandemie**

Um den Schulen dennoch eine aussagekräftige Diagnostik des Lernstandes ihrer Schüler\*innen zu ermöglichen, wurden vom IfBQ alternative Testinstrumente für die Kompetenzmessung in den für die weitere Lernentwicklung zentralen Bereichen Deutsch-Leseverstehen und Mathematik zur Verfügung gestellt.

Für die Entwicklung der alternativen Erhebungen LERNSTAND und FLIP waren drei Überlegungen leitend:

- 1) Die unterstützende Funktion bei der datengestützten Schul- und Unterrichtsentwicklung sollte aufrechterhalten werden. Die Schulen sollten ein Instrument zur Kompetenzmessung nutzen können, das wie die KERMIT-Erhebungen belastbare Hinweise auf Stärken und Schwächen der Lernenden gibt, auf deren Grundlage Anpassungen in der Gestaltung des Unterrichts vorgenommen werden können. Im Gegensatz zu den KERMIT-Erhebungen sollten die Ergebnisse wesentlich schneller (bis zu 14 Tage nach dem Ende eines Testblocks) an die Schulen zurückgemeldet werden, damit die Erkenntnisse zeitnah in die laufende Unterrichtsplanung einfließen konnten. Die Möglichkeit der deutlich schnelleren Rückmeldung der Ergebnisse ergab sich u. a. aus der Konzentration auf die beiden zentralen Testdomänen Mathematik und Deutsch-Leseverstehen, die eine wesentlich schnellere Dateneingabe zuließ.
- 2) Die Schulen sollten die Erhebung mit möglichst großer Flexibilität und möglichst geringem Aufwand durchführen können. Da die Corona-bedingten Veränderungen ungewohnte Belastungen und neue Herausforderungen mit sich bringen, z. B. die Entwicklung neuer Formen der Vermittlung der Lerninhalte und der Bereitstellung von Lernmaterialien (digital, per Post usw.) sowie die Kompensation der ausgefallenen Unterrichtsstunden, sollte der Mehraufwand durch die Lernstandserhebungen möglichst gering gehalten werden. Deshalb war die Teilnahme an FLIP 3 und FLIP 7 für alle Schulen freiwillig. Die Teilnahme an

LERNSTAND 4 und LERNSTAND 5 war hingegen für alle staatlichen Schulen verpflichtend.

- 3) Die Instrumente sollten Hinweise auf mögliche Auswirkungen der Corona-Pandemie auf die Lernstände der Schüler\*innen geben. Vor allem mit den verpflichtenden Kompetenzmessungen LERNSTAND 4 und LERNSTAND 5 hatten die Schulen die Möglichkeit, die Lernstände ihrer Schüler\*innen im Hinblick auf normalerweise erwartbare Kompetenzstände einzuschätzen: In den Ergebnisrückmeldungen wurden die erzielten Punktwerte Kompetenzstufen und Bildungsstandards zugeordnet und mit den Ergebnissen vorheriger KERMIT-Kohorten verglichen.

Im Folgenden werden die Erhebungen LERNSTAND und FLIP genauer beschrieben.

### **3.1 LERNSTAND 4 und LERNSTAND 5**

Wie bereits erwähnt, war die Teilnahme an den Kompetenzmessungen LERNSTAND 4 und LERNSTAND 5 für alle staatlichen Schulen verpflichtend. Die Schulen in privater Trägerschaft konnten selbst entscheiden, ob sie an den Erhebungen teilnehmen. Beide LERNSTAND-Erhebungen wurden in je drei aufeinanderfolgenden Testblöcken durchgeführt. Jede Schule wurde einem der Testblöcke zugeteilt, sodass jeder Block etwa gleich stark mit Schulen besetzt war. Nach Beendigung eines Testblocks wurden die bearbeiteten Testhefte vom IfBQ codiert und ausgewertet. Die Ergebnisse wurden den drei Teilgruppen sukzessive jeweils innerhalb von 14 Tagen zurückgemeldet.

Normalerweise erhalten die Schulen im Rahmen der KERMIT-Rückmeldung faire soziale Vergleichswerte. Sie ergeben sich aus den Leistungsmittelwerten von jeweils acht Schulen, die eine vergleichbare soziale Belastung (Sozialindex) aufweisen. Der Sozialindex ist ein offizielles Maß für die soziale Belastung jeder Schule und mit Ressourcenzuweisung verbunden (Schulte, Hartig & Pietsch, 2014). Insgesamt gibt es sechs Sozialindex-Gruppen. In Sozialindex-Gruppe 1 befinden sich Schulen mit einer sozial stark belasteten Schülerschaft. Schulen der Sozialindex-Gruppe 6 sind kaum sozial belastet. Da die Testauswertung als Konsequenz der gestaffelten Erhebungszeiträume nicht erst nach Abschluss der gesamten LERNSTAND-Erhebung durchgeführt wurde, konnten in den Rückmeldungen keine fairen sozialen Vergleichswerte aus der aktuellen Erhebung zur Verfügung gestellt werden.

Die Erhebung LERNSTAND 4 wurde als Ersatz für die ausgefallene KERMIT-3-Erhebung (April/Mai 2020) durchgeführt. Das heißt, dass die Kinder, die im Frühjahr KERMIT 3 geschrieben hätten, nun ein paar Monate später im Rahmen von LERNSTAND 4 getestet wurden. Die eingesetzten Aufgaben für Deutsch-Lese-

verstehen und Mathematik waren identisch mit den Aufgaben aus KERMIT 3 (2019). Neben den Kompetenzen in Deutsch-Leseverstehen wurden im Testbereich Mathematik die Kompetenzen für die Leitidee 3 (Raum und Form) und die Leitidee 5 (Daten, Häufigkeiten und Wahrscheinlichkeit) überprüft. Die Testzeit betrug insgesamt 100 Minuten (40 Minuten für Deutsch-Leseverstehen und 60 Minuten für Mathematik).

Durch den Einsatz des bereits erprobten Testinstruments KERMIT 3 war es möglich, die Auswertung auf der sogenannten Bista-Metrik (Metrik der Bildungsstandards) vorzunehmen. Dieses Vorgehen erlaubt den Schulen die Interpretation der erzielten Punktwerte in Bezug auf die Kompetenzstufenmodelle der KMK und damit in Bezug auf die vertrauten Bildungsstandards (kriteriale Bezugsnorm; vgl. Lücken et al., 2014; Schulte et al., 2014). Darüber hinaus konnten die Kompetenzen der damaligen Drittklässler\*innen in KERMIT 3 (2019) mit den Kompetenzen der Viertklässler\*innen in LERNSTAND 4 (2020) verglichen werden. Dadurch war es möglich, eventuelle Leistungsunterschiede zwischen den beiden Kohorten sichtbar zu machen.

Neben den mittleren Leistungswerten wurden auf Klassenebene und auf Schulebene die prozentualen Verteilungen auf die Kompetenzstufen I bis V für Deutsch-Leseverstehen und Mathematik (global) zurückgemeldet. Ferner wurden auf Klassenebene die prozentualen Lösungshäufigkeiten für jede Aufgabe berichtet und auf individueller Ebene die erzielten Punktwerte für Deutsch-Leseverstehen und Mathematik (global und getrennt nach Leitideen).

Als Vergleichswerte zu ihren LERNSTAND-4-Ergebnissen wurden den Grundschulen die durchschnittlichen Kompetenzwerte aus KERMIT 3 (2019) von Grundschulen in ähnlicher sozialer Lage zurückgemeldet. Da sich die Testzeiträume zwischen den beiden Erhebungen um drei bis vier Monate unterscheiden, von denen die sechswöchigen Sommerferien abzuziehen sind, sollte dieser Umstand bei der Interpretation der Ergebnisse berücksichtigt werden.

LERNSTAND 5 wurde als Alternative zu der ausgefallenen KERMIT-5-Erhebung in allen fünften Klassen der staatlichen Hamburger Schulen verpflichtend durchgeführt. Analog zu LERNSTAND 4 war die Teilnahme für Schulen in privater Trägerschaft freiwillig. In Mathematik wurden die Kompetenzen in allen mathematischen Leitideen erhoben. Die Testzeit betrug insgesamt 80 Minuten (40 Minuten für Deutsch-Leseverstehen und 40 Minuten für Mathematik). Die Punktwerte waren wie in LERNSTAND 4 auf der Bista-Metrik angegeben. Dadurch war es möglich, die erzielten Punktwerte unterschiedlichen Kompetenzstufen und Bildungsstandards zuzuordnen. Als Referenzwerte wurden die Ergebnisse aus KERMIT 5 (2019) für die jeweilige Schulform angegeben.

Auf Klassen- und Schulebene wurden die mittleren Leistungswerte für Deutsch-Leseverstehen und Mathematik zurückgemeldet. Außerdem wurden die prozentualen Lösungshäufigkeiten für die Lesetexte in Deutsch, die Leitideen in Mathematik und die einzelnen Aufgaben berichtet. Auf individueller Ebene wurden sowohl die Punktwerte für die beiden Testbereiche als auch die Lösungshäufigkeiten für die Lesetexte und die mathematischen Leitideen zurückgemeldet.

Tabelle 2 stellt ausgewählte Merkmale der LERNSTAND-Erhebungen den KERMIT-Erhebungen gegenüber. Der Verpflichtungsgrad (verpflichtend) und die Auswertungsbedingungen (Kodierung durch das IfBQ) waren für alle Erhebungen identisch.

Tab. 2: Vergleich der LERNSTAND-Erhebungen mit den korrespondierenden KERMIT-Erhebungen

Erhebung	LERNSTAND 4	KERMIT 3	LERNSTAND 5	KERMIT 5
Testzeitraum	August – September (2020)	April – Mai (2019)	August – September (2020)	September (2020)
Durchführungsbedingungen	Testdurchführung durch Lehrkräfte	Testdurchführung durch Lehrkräfte	Testdurchführung durch Lehrkräfte	Testdurchführung durch geschulte externe Testleitungen
Vergleichswerte der klassenbezogenen Rückmeldungen	Schule, Schulen mit demselben Sozialindex	Schule, acht Schulen mit ähnlicher Schülerschaft (fairer Vergleich)	Schule, schulform-bezogene Ergebnisse	Schule, acht Schulen mit ähnlicher Schülerschaft (fairer Vergleich), schulform-bezogene Ergebnisse
Testdomänen	Deutsch-Leseverstehen, Mathematik	Deutsch-Leseverstehen, Deutsch-Zuhören, Deutsch-Rechtschreibung <sup>1</sup> , Mathematik	Deutsch-Leseverstehen, Mathematik	Deutsch-Leseverstehen, Deutsch-Rechtschreibung, Mathematik, Englisch-Hörverstehen, Naturwissenschaften

Quelle: eigene Darstellung

### 3.2 FLIP 3 und FLIP 7

In den dritten und siebten Jahrgangsstufen fanden die FLIP-Erhebungen statt, bei denen die Teilnahme allen Hamburger Schulen freigestellt war. Die Testzeiträume für beide Erhebungen waren sehr groß angelegt, damit die Lehrkräfte die Erhebung möglichst flexibel durchführen konnten. Im Anschluss an die Testungen wurden die

<sup>1</sup> Regelhaft wird im Rahmen von KERMIT 3 auch Deutsch-Rechtschreibung durchgeführt.

Antworten der Schüler\*innen von den Lehrkräften in einem Online-Portal eingetragen. Die Ergebnisse standen den Lehrkräften dadurch automatisiert und umgehend zur Verfügung. Die FLIP-7-Testung wurde sowohl als Papier-Bleistift-Version als auch als Online-Version angeboten.

Tabelle 3 gibt einen Überblick über die Unterschiede und Gemeinsamkeiten der KERMIT-2- und KERMIT-7-Erhebungen im Vergleich zu den alternativen FLIP-Erhebungen. Da die Teilnahme an FLIP freiwillig war und die selektive Stichprobe der teilnehmenden Schulen zu verzerrten Ergebnissen führen könnte, werden im Ergebnisteil dieses Beitrages ausschließlich die Ergebnisse für die LERNSTAND-Erhebungen in den Klassenstufen vier und fünf berichtet.

Tab. 3: Vergleich der FLIP-Erhebungen mit den korrespondierenden KERMIT-Erhebungen

Erhebung	FLIP 3	KERMIT 2	FLIP 7	KERMIT 7
Verpflichtungsgrad	freiwillig	verpflichtend	freiwillig	verpflichtend
Testzeitraum	August – Oktober (2020)	April – Mai (2019)	August – Oktober (2020)	August – September (2020)
Durchführungsbedingungen	Testdurchführung durch Lehrkräfte	Testdurchführung durch Lehrkräfte	Testdurchführung durch Lehrkräfte	Testdurchführung durch geschulte externe Testleitungen
Auswertungsbedingungen	Testkodierung durch Lehrkräfte	Testkodierung durch IfBQ	Testkodierung durch Lehrkräfte, in der Online-Version Testkodierung durch IfBQ	Testkodierung durch IfBQ
Vergleichswerte in den klassenbezogenen Rückmeldungen	Schule, frühere Ergebnisse aller Grundschulen	Schule, acht Schulen mit ähnlicher Schülerschaft (fairer Vergleich)	Schule, frühere schulformbezogene Ergebnisse	Schule, acht Schulen mit ähnlicher Schülerschaft (fairer Vergleich), schulformbezogene Ergebnisse
Testdomänen	Deutsch-Leseverstehen, Mathematik	Deutsch-Leseverstehen, Mathematik	Deutsch-Leseverstehen, Mathematik	Deutsch-Leseverstehen, Mathematik, Englisch-Hörverstehen, Englisch-Leseverstehen, Naturwissenschaften

Quelle: eigene Darstellung



Im folgenden Abschnitt werden die Ergebnisse der LERNSTAND-Erhebungen den Ergebnissen der korrespondierenden KERMIT-Kohorten gegenübergestellt. Abschließend werden die Befunde im Hinblick auf mögliche Zusammenhänge mit den Corona-bedingten Änderungen im Unterrichtsgeschehen diskutiert.

#### 4 Ergebnisse LERNSTAND 4 und LERNSTAND 5

Die hier vorgestellten Ergebnisse beziehen sich auf eine Teilgruppe von Hamburger Schüler\*innen, die eine staatliche Schule mit einem zugewiesenen Sozialindex besuchen. Für die Analysen wurden Schulen aus je zwei Sozialindex-Gruppen zusammengefasst (Belastungsgruppe 1 = Sozialindex 1 und 2, Belastungsgruppe 2 = Sozialindex 3 und 4, Belastungsgruppe 3 = Sozialindex 5 und 6).

Tabelle 4 gibt einen Überblick über die Anzahl der Schüler\*innen in den LERNSTAND-4- und LERNSTAND-5-Erhebungen und den korrespondierenden KERMIT-Erhebungen, aufgeteilt nach Schulform und Belastungsgruppen. In Hamburg existieren seit der Schulreform 2010 für die weiterführenden Schulen zwei unterschiedliche Schulformen. Auf den Gymnasien legen die Schüler\*innen das Abitur nach der 12. Klasse ab; an den Stadtteilschulen sind der Erste und Mittlere Schulabschluss möglich sowie das Abitur nach 13 Jahren. In den weiterführenden Schulen ist die Verteilung der Schüler\*innen auf die Belastungsgruppen nach Schulform ungleichmäßig (vgl. Tab. 4).

Tab. 4: Anzahl der Schüler\*innen in den LERNSTAND- und KERMIT-Erhebungen getrennt nach Schulform und Belastungsgruppe

Erhebung	Schulform	Anzahl (N)			
		gesamt	Belastungsgruppe 1	Belastungsgruppe 2	Belastungsgruppe 3
LERNSTAND 4 (2020)	Grundschule	13.480	3.626	4.847	5.007
KERMIT 3 (2019)	Grundschule	13.922	3.631	5.107	5.184
	Gesamt	14.090	3.171	4.431	6.488
LERNSTAND 5 (2020)	Stadtteilschule	6.814	2.873	2.956	985
	Gymnasium	7.276	298	1.475	5.503
	Gesamt	13.716	2.975	4.482	6.259
KERMIT 5 (2019)	Stadtteilschule	6.722	2.697	3.058	967
	Gymnasium	6.994	278	1.424	5.292

Quelle: eigene Darstellung

Kompetenzwerte einer bereits relativ geringen Anzahl an Schüler\*innen können demnach in der Belastungsgruppe 3 der Stadtteilschulen bzw. in der Belastungsgruppe 1 der Gymnasien zu anderen Verteilungen innerhalb dieser Belastungsgruppen führen. Dieser Umstand sollte bei der Interpretation der Ergebnisse berücksichtigt werden.

#### 4.1 LERNSTAND 4

In einer ersten Analyse wurden die Anteile der Schüler\*innen aus LERNSTAND 4 und KERMIT 3 (2019) gegenübergestellt, die in Deutsch-Leseverstehen weniger als zehn und in Mathematik weniger als 20 Aufgaben bearbeitet haben (Dropout).

Bei den regulären KERMIT-Erhebungen wird dieses Kriterium üblicherweise angewendet, um eine möglichst stabile Kompetenzschätzung zu erhalten. Erst wenn eine ausreichend große Anzahl an Aufgaben bearbeitet wurde, ist eine reliable und somit valide Schätzung der Kompetenzen gegeben. Da KERMIT in erster Linie ein Instrument zur Schul- und Unterrichtsentwicklung ist und nicht für systemische Analysen entwickelt wurde, werden diese Dropouts lediglich für die Schulen gekennzeichnet.

In Deutsch-Leseverstehen hat in beiden Kohorten eine insgesamt sehr große Gruppe von Schüler\*innen weniger als zehn Aufgaben bearbeitet. Der Anteil der Dropouts fällt in LERNSTAND 4 höher aus als in KERMIT 3 (2019). Je größer die soziale Belastung der Schulen ist, desto häufiger hören die Schüler\*innen mit der Bearbeitung der Testaufgaben frühzeitig auf (vgl. Tab. 5).

Tab. 5: Anzahl der Schüler\*innen, die weniger als zehn Aufgaben in Deutsch-Leseverstehen bearbeitet haben

Erhebung	Deutsch-Leseverstehen		
	Anzahl Teilnehmer*innen	Dropout (absolut/%)	
LERNSTAND 4 (2020)	Gesamt	13.104	836 (6,4%)
	Belastungsgruppe 1	3.482	377 (10,8%)
	Belastungsgruppe 2	4.706	300 (6,4%)
	Belastungsgruppe 3	4.916	159 (3,2%)
KERMIT 3 (2019)	Gesamt	13.209	698 (5,3%)
	Belastungsgruppe 1	3.391	321 (9,5%)
	Belastungsgruppe 2	4.840	254 (5,2%)
	Belastungsgruppe 3	4.978	123 (2,5%)

Quelle: eigene Darstellung

In Mathematik haben im Vergleich zu Deutsch-Leseverstehen insgesamt deutlich weniger Kinder eine zu geringe Anzahl an Aufgaben bearbeitet. Auch in dieser Testdomäne ist der Anteil dieser Schüler\*innen in LERNSTAND 4 höher als in KERMIT 3 (2019). Analog zu den Befunden in Deutsch-Leseverstehen wurde auch in Mathematik die Testbearbeitung an sozial stärker belasteten Schulen häufiger frühzeitig abgebrochen (vgl. Tab. 6).

Tab. 6: Anzahl der Schüler\*innen, die weniger als 20 Aufgaben in Mathematik bearbeitet haben

Erhebung	Mathematik		
		Anzahl Teilnehmer*innen	Dropout (absolut/%)
LERNSTAND 4 (2020)	Gesamt	13.165	237 (1,8%)
	Belastungsgruppe 1	3.520	121 (3,4%)
	Belastungsgruppe 2	4.713	85 (1,8%)
	Belastungsgruppe 3	4.932	31 (0,6%)
KERMIT 3 (2019)	Gesamt	13.238	126 (1,0%)
	Belastungsgruppe 1	3.401	75 (2,2%)
	Belastungsgruppe 2	4.848	36 (0,7%)
	Belastungsgruppe 3	4.989	15 (0,3%)

Quelle: eigene Darstellung

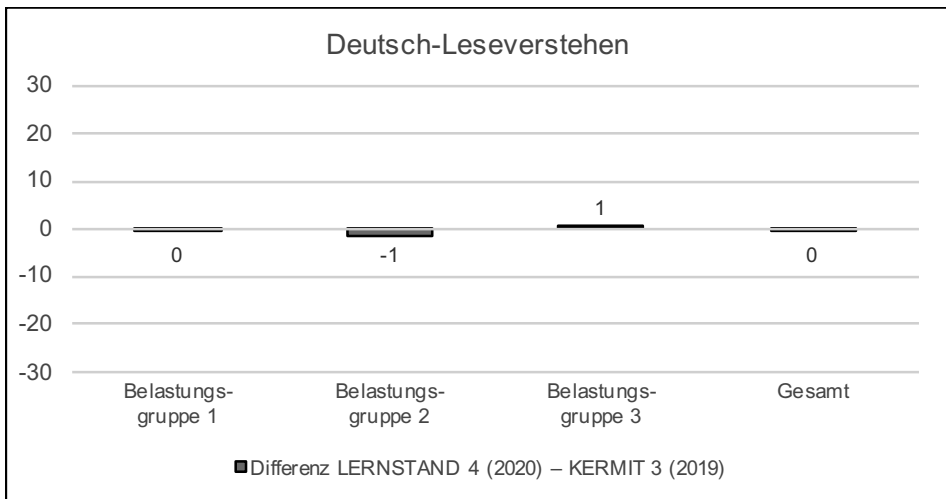
Die Abbildungen 1 und 2 auf der folgenden Seite zeigen die Differenzen zwischen den erreichten Kompetenzwerten in LERNSTAND 4 und KERMIT 3. Negative Werte bedeuten, dass die Schüler\*innen der KERMIT-3-Kohorte aus dem Jahr 2019 höhere Punktwerte erzielt haben als die Schüler\*innen der LERNSTAND-4-Kohorte aus dem Jahr 2020. Die Ergebnisse wurden für jede Belastungsgruppe und für die gesamte Gruppe der Grundschüler\*innen berechnet.

Um die praktische Bedeutsamkeit der Unterschiede zwischen den Kohorten beurteilen zu können, wurden Effektstärken (vgl. Cohen, 1988) berechnet.<sup>2</sup> Effekte kleiner als  $d = .2$  werden als statistisch nicht bedeutsam interpretiert. Erst ab einer Effektstärke von  $d = .2$  kann von einem kleinen Effekt ausgegangen werden, der sich eventuell auf die veränderten schulischen Bedingungen zurückführen ließe. Aus Gründen der Sparsamkeit werden im Ergebnisteil ausschließlich Effekte berichtet, die größer als  $d = .1$  sind, oder die Effektstärke des größten Unterschieds.

2 Nach Cohen (1988) sind Effektstärken ( $d$ ) zwischen  $d = .2$  und  $d = .4$  als kleine Effekte definiert. Effektstärken zwischen  $d = .5$  und  $d = .7$  werden als mittlere Effekte interpretiert und Effektstärken ab  $d = .8$  als große Effekte.

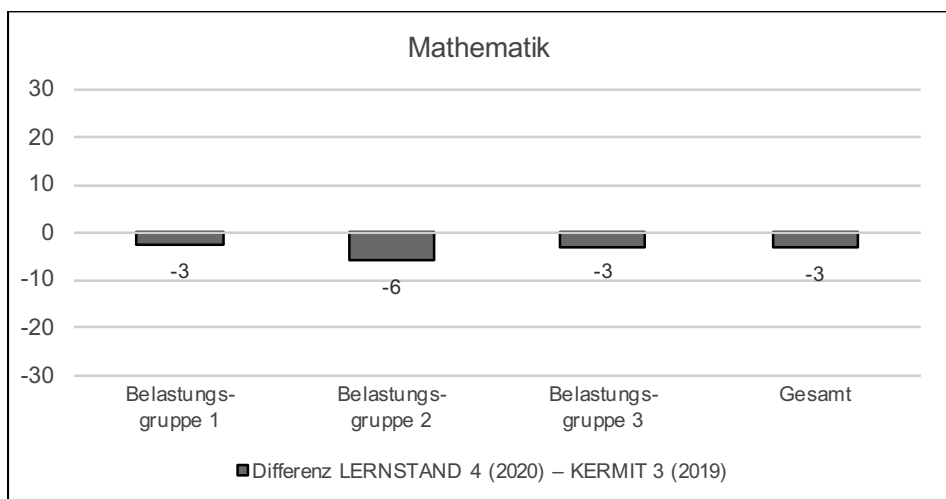
Insgesamt zeigt sich, dass es sowohl in der Gesamtgruppe als auch in den Belastungsgruppen keine bedeutsamen Unterschiede zwischen den durchschnittlich erzielten Punktwerten der beiden Kohorten gibt. Der größte Unterschied in Belastungsgruppe 2 in Mathematik erreicht eine Effektstärke von  $d = .061$ .

Abb. 1: Differenzen der erzielten Punktwerte in Deutsch-Leseverstehen zwischen der LERNSTAND-4- und der KERMIT-3-Kohorte



Quelle: eigene Berechnungen

Abb. 2: Differenzen der erzielten Punktwerte in Mathematik zwischen der LERNSTAND-4- und der KERMIT-3-Kohorte



Quelle: eigene Berechnungen

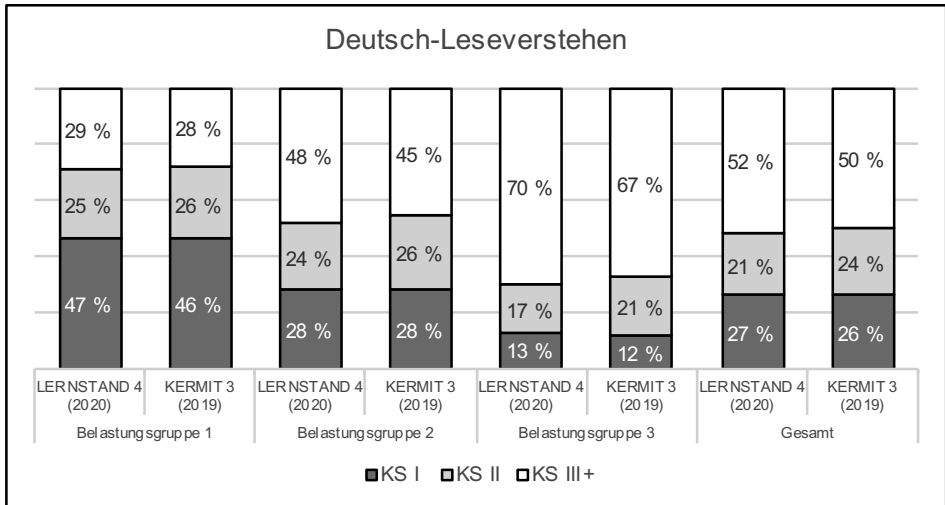
Die Skalierung der Ergebnisse auf der Metrik der Bildungsstandards der KMK ermöglicht die Zuordnung konkreter Punktwerte zu bestimmten Kompetenzstufen. Mit dem vom IQB entwickelten Kompetenzstufenmodell lassen sich die Kompetenzstufen wiederum Mindeststandards und Regelstandards zuordnen, die für das Ende der Grundschulzeit definiert sind. Schüler\*innen mit Punktwerten auf der Kompetenzstufe 2 (KS II) erfüllen die Mindeststandards und diejenigen mit Punktwerten auf der Kompetenzstufe 3 oder höher (KS III+) die Regelstandards. Liegen die Punktwerte auf Kompetenzstufe 1 (KS I), haben die Schüler\*innen die Mindeststandards nicht erreicht.

Die Abbildungen 3 und 4 auf der folgenden Seite stellen die Kompetenzstufenverteilungen von LERNSTAND 4 den Kompetenzstufenverteilungen von KERMIT 3 gegenüber. Die Ergebnisse sind jeweils für die Gesamtgruppe und die Belastungsgruppen ausgewiesen. In beiden Kompetenzbereichen zeigt der Vergleich zwischen den Kohorten, dass die prozentuale Verteilung auf die Kompetenzstufen ungefähr gleich ausfällt. Ähnlich große Abweichungen von bis zu 15 Prozentpunkten zwischen einzelnen Kompetenzstufen lassen sich beispielsweise im Kohortenvergleich verschiedener KERMIT-5-Erhebungen der vergangenen Jahre beobachten (vgl. Depping, 2020; Lücken, 2017). Dieser Befund lässt sich außerdem für Kohortenvergleiche zwischen vergangenen KERMIT-3-Erhebungen replizieren.

Die Unterschiede zwischen den Kompetenzstufenverteilungen in LERNSTAND 4 und KERMIT 3 (2019) sind deshalb als praktisch nicht bedeutsam zu interpretieren und lassen sich vermutlich auch auf normale Leistungsschwankungen zwischen Kohorten verschiedener Jahrgänge zurückführen.

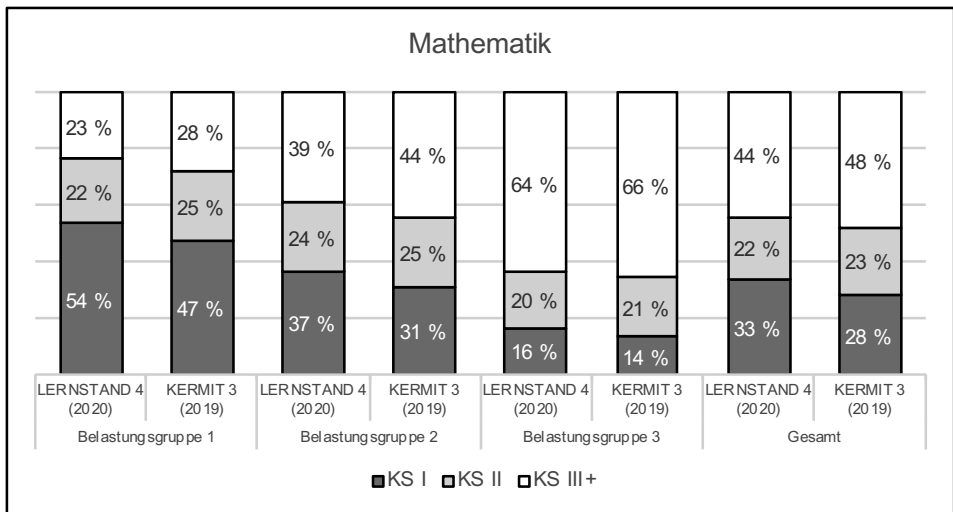
In Mathematik sind im Vergleich zu Deutsch-Leseverstehen etwas größere Unterschiede zwischen den Kohorten zu verzeichnen. Die größten Schwankungen zeigen sich in Mathematik in den Belastungsgruppen 1 und 2. Insgesamt verfehlen in LERNSTAND 4 etwas mehr Schüler\*innen die Mindeststandards, und die Regelstandards werden etwas seltener erreicht. Allerdings ist der Zusammenhang zwischen der sozialen Belastung einer Schule und dem Erreichen der Bildungsstandards wesentlich größer, als es die Kohortenunterschiede im Hinblick auf die Kompetenzstufenverteilungen sind.

Abb. 3: Vergleich der prozentualen Verteilungen auf die Kompetenzstufen in Deutsch-Leseverstehen zwischen LERNSTAND 4 (2020) und KERMIT 3 (2019)



Quelle: eigene Berechnungen

Abb. 4: Vergleich der prozentualen Verteilungen auf die Kompetenzstufen in Mathematik zwischen LERNSTAND 4 (2020) und KERMIT 3 (2019)



Quelle: eigene Berechnungen

## 4.2 LERNSTAND 5

Analog zu LERNSTAND 4 wurde in LERNSTAND 5 zunächst untersucht, wie viele Schüler\*innen nicht genügend Aufgaben für eine Berechnung ihrer Kompetenzen bearbeitet haben. In Deutsch-Leseverstehen betrifft dies analog zu LERNSTAND 4 und KERMIT 3 die Schüler\*innen, die weniger als zehn Aufgaben bearbeitet haben. In Mathematik ist im Gegensatz zur Primarstufe eine prozentuale Bearbeitung von 20 Prozent der Aufgaben erforderlich. Die Analysen für die fünfte Jahrgangsstufe wurden zusätzlich getrennt für Stadtteilschulen (STS) und Gymnasien (GYM) durchgeführt.

In Deutsch-Leseverstehen fallen die Dropouts erneut wesentlich höher aus als in Mathematik (vgl. Tab. 7 und Tab. 8). Der Anteil der Schüler\*innen, die weniger als zehn Aufgaben bearbeitet haben, ist in LERNSTAND 5 höher als in KERMIT 5 (2019). Außerdem zeigt sich auch für die fünfte Jahrgangsstufe, dass der Anteil der Dropouts deutlich höher ist, je größer die soziale Belastung der Schule ist. Aus den Analysen getrennt nach Schulform geht hervor, dass die Testabbrüche vor allem an den Stadtteilschulen stattgefunden haben – insbesondere an Schulen mit hoher so-

Tab. 7: Anzahl der Schüler\*innen, die weniger als zehn Aufgaben in Deutsch-Leseverstehen bearbeitet haben

		Deutsch-Leseverstehen	
		Anzahl Teilnehmer*innen STS / GYM	Dropout (absolut/%) STS / GYM
LERNSTAND 5 (2020)	Gesamt	13.936	760 (5,5 %)
		6.693 / 7.243	734 (11,0 %) / 26 (0,4 %)
	Belastungsgruppe 1	3.096	421 (13,6 %)
		2.801 / 295	420 (15,0 %) / 1 (0,3 %)
	Belastungsgruppe 2	4.386	267 (6,1 %)
		2.914 / 1.472	262 (9,0 %) / 5 (0,3 %)
KERMIT 5 (2019)	Belastungsgruppe 3	6.454	72 (1,1 %)
		978 / 5.476	52 (5,3 %) / 20 (0,4 %)
	Gesamt	13.716	473 (3,4 %)
		6.722 / 6.994	448 (6,7 %) / 25 (0,4 %)
	Belastungsgruppe 1	2957	244 (8,2 %)
		2.697 / 278	241 (8,9 %) / 3 (1,1 %)
	Belastungsgruppe 2	4.482	170 (3,8 %)
		3.058 / 1.424	164 (5,4 %) / 6 (0,4 %)
	Belastungsgruppe 3	6.259	59 (0,9 %)
	967 / 5.292	43 (4,4 %) / 16 (0,3 %)	

Quelle: eigene Darstellung

zialer Belastung. Die Hamburger Gymnasiast\*innen bearbeiten ebenfalls weniger Aufgaben als diejenigen in KERMIT 5 (2019). Jedoch sind die Unterschiede und auch die absoluten Zahlen insgesamt sehr niedrig.

In Mathematik sind die Befunde für die Schüler\*innen, die weniger als 20 Prozent der Aufgaben bearbeitet haben, für beide Kohorten relativ ähnlich mit leicht günstigeren Ergebnissen für LERNSTAND 5.

An den Stadtteilschulen gab es im Vergleich zu den Gymnasien deutlich mehr Schüler\*innen, die zu wenige Aufgaben bearbeitet haben. Außerdem wird an den Stadtteilschulen der Zusammenhang zwischen dem Anteil der Dropouts und der sozialen Belastung sichtbar. An den Gymnasien kann aufgrund der geringen Fallzahlen in den Dropouts keine Aussage zum Zusammenhang mit der sozialen Belastung getroffen werden.

Tab. 8: Anzahl der Schüler\*innen, die weniger als 20 Prozent der Aufgaben in Mathematik bearbeitet haben

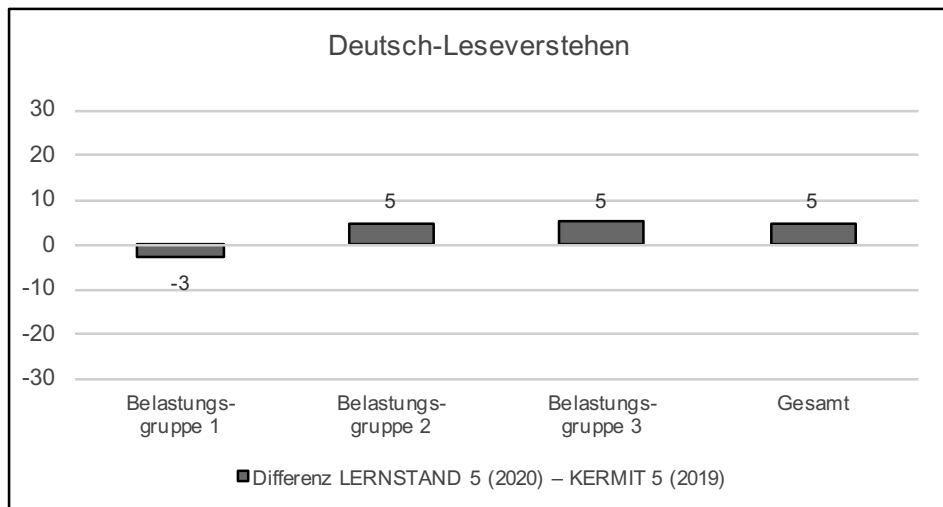
		Mathematik	
		Anzahl Teilnehmer*innen STS / GYM	Dropout (absolut/%) STS / GYM
LERNSTAND 5 (2020)	Gesamt	13.938 6.710 / 7.228	110 (0,8%) 107 (1,6%) / 3 (0,04%)
	Belastungsgruppe 1	3.126 2.828 / 298	59 (1,9%) 59 (2,1%) / 0
	Belastungsgruppe 2	4369 2.905 / 1.464	38 (0,9%) 37 (1,3%) / 1 (0,1%)
	Belastungsgruppe 3	6.443 977 / 5.466	13 (0,2%) 11 (1,1%) / 2 (0,04%)
	Gesamt	13.703 6.709 / 6.994	133 (1,0%) 119 (1,8%) / 14 (0,2%)
	KERMIT 5 (2019)	Belastungsgruppe 1	2.967 2.689 / 278
	Belastungsgruppe 2	4.477 3.053 / 1.424	52 (1,2%) 45 (1,5%) / 7 (0,5%)
	Belastungsgruppe 3	6.259 967 / 5.292	18 (0,3%) 12 (1,2%) / 6 (0,1%)

Quelle: eigene Darstellung



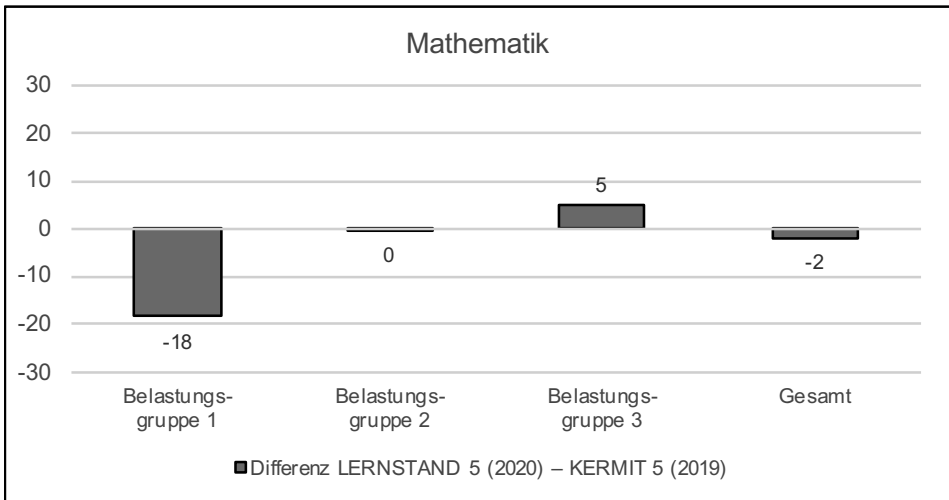
Analog zu LERNSTAND 4 werden die Differenzen zwischen den mittleren Punktwerten in LERNSTAND 5 und KERMIT 5 berichtet sowie die prozentualen Verteilungen auf die Kompetenzstufen. Die Analysen wurden sowohl für die Gesamtgruppe als auch getrennt für jede Schulform durchgeführt. Die Abbildungen 5 und 6 zeigen, dass sich die durchschnittlich erzielten Punktwerte zwischen beiden Kohorten nicht bedeutsam voneinander unterscheiden. Der größte Differenzwert innerhalb der Belastungsgruppe 1 in Mathematik erreicht eine Effektstärke von  $d = .19$  und hat daher keine statistische Bedeutsamkeit.

Abb. 5: Differenzen der erzielten Punktwerte in Deutsch-Leseverstehen zwischen der LERNSTAND-5- und der KERMIT-5-Kohorte



Quelle: eigene Berechnungen

Abb. 6: Differenzen der erzielten Punktwerte in Mathematik zwischen der LERNSTAND-5- und der KERMIT-5-Kohorte



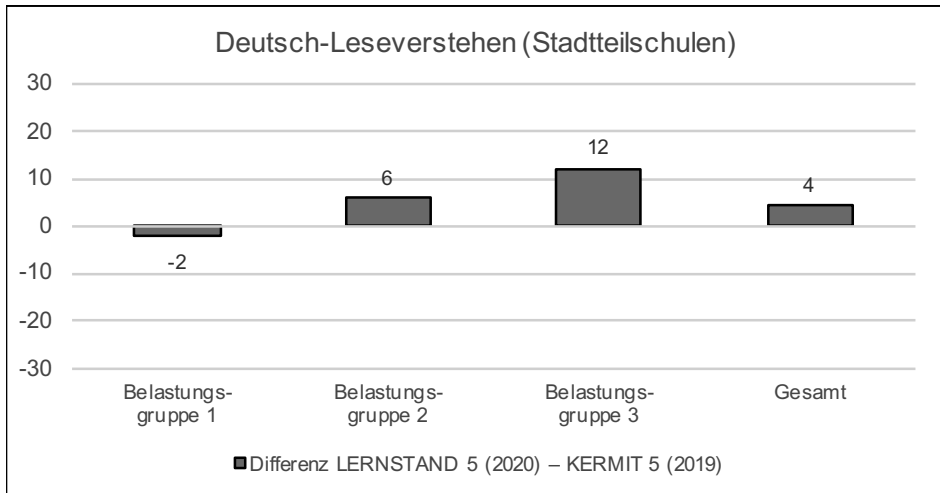
Quelle: eigene Berechnungen

Die Punktwertdifferenzen zwischen den Kohorten für die Stadtteilschulen sind den Abbildungen 7 und 8 auf der folgenden Seite zu entnehmen. Über alle Belastungsgruppen hinweg erzielen die Schüler\*innen der LERNSTAND-5-Kohorte in Mathematik im Mittel 11 Punkte weniger ( $d = .11$ ). In der Belastungsgruppe 3 fallen die Ergebnisse in LERNSTAND 5 für Deutsch-Leseverstehen mit 12 Punkten etwas besser aus ( $d = .11$ ). In Mathematik erreichen die Schüler\*innen in der Belastungsgruppe 1 im Mittel 19 Punkte weniger ( $d = .21$ ) als in der vergleichbaren KERMIT-5-Erhebung (2019).

Auch für die Gymnasien fallen die durchschnittlich erzielten Kompetenzwerte in LERNSTAND 5 (2020) und KERMIT 5 (2019) ähnlich aus (vgl. Abb. 9 und Abb. 10 auf S. 70). In der Belastungsgruppe 1, die mit knapp 300 Schüler\*innen pro Kohorte vergleichsweise klein ist, sind die Unterschiede in beiden Kompetenzbereichen am größten. Allerdings sind die Unterschiede weder in Deutsch-Leseverstehen ( $d = 0.18$ ) noch in Mathematik ( $d = .11$ ) statistisch bedeutsam.

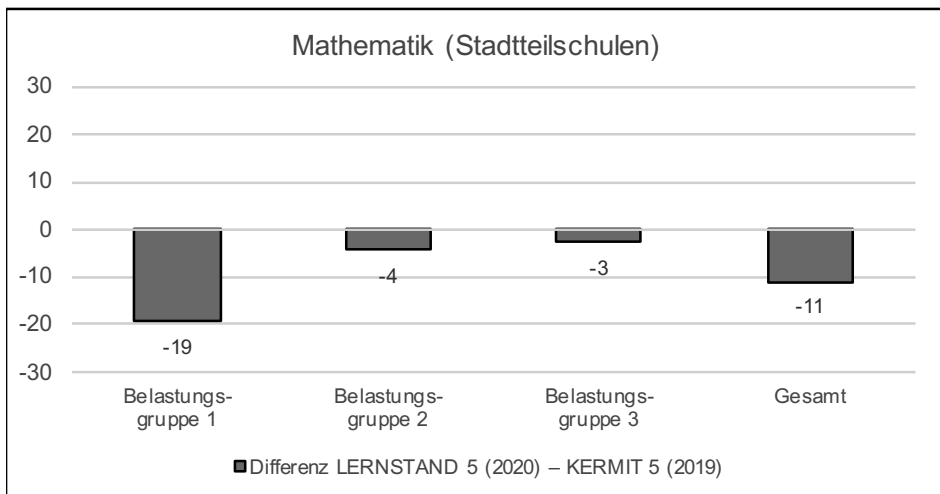
Schulformübergreifend verteilen sich die Kompetenzen der Schüler\*innen im Kohortenvergleich in beiden Kompetenzbereichen annähernd gleich auf die Kompetenzstufen (vgl. Abb. 11 und Abb. 12 auf S. 71). Wie bereits für LERNSTAND 4 lässt sich auch für LERNSTAND 5 feststellen, dass zwar ein starker Zusammenhang zwischen der sozialen Belastung einer Schule und den Kompetenzen ihrer Schülerschaft besteht, nicht aber zwischen dem Zeitpunkt der Erhebung und den Kompetenzverteilungen.

Abb. 7: Differenzen der erzielten Punktwerte in Deutsch-Leseverstehen zwischen der LERNSTAND-5- und der KERMIT-5-Kohorte an Stadtteilschulen



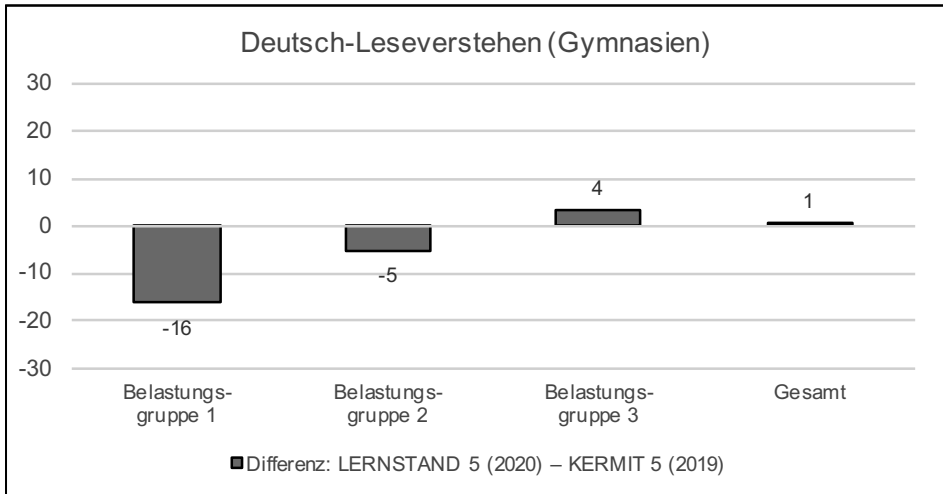
Quelle: eigene Berechnungen

Abb. 8: Differenzen der erzielten Punktwerte in Mathematik zwischen der LERNSTAND-5- und der KERMIT-5-Kohorte an Stadtteilschulen



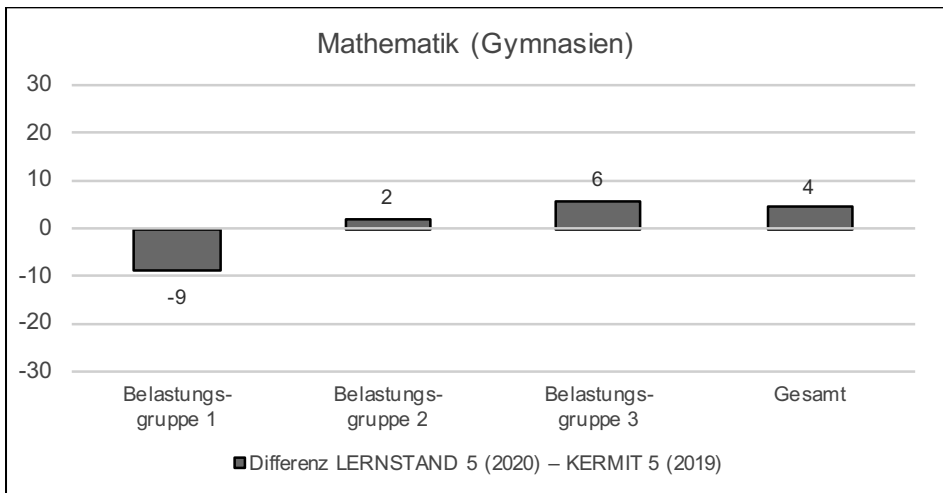
Quelle: eigene Berechnungen

Abb. 9: Differenzen der erzielten Punktwerte in Deutsch-Leseverstehen zwischen der LERNSTAND-5- und der KERMIT-5-Kohorte an Gymnasien



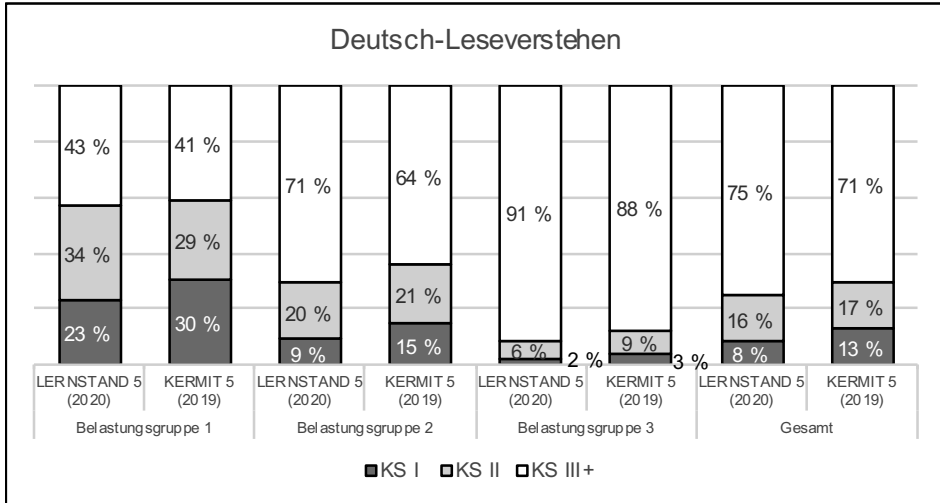
Quelle: eigene Berechnungen

Abb. 10: Differenzen der erzielten Punktwerte in Mathematik zwischen der LERNSTAND-5- und der KERMIT-5-Kohorte an Gymnasien



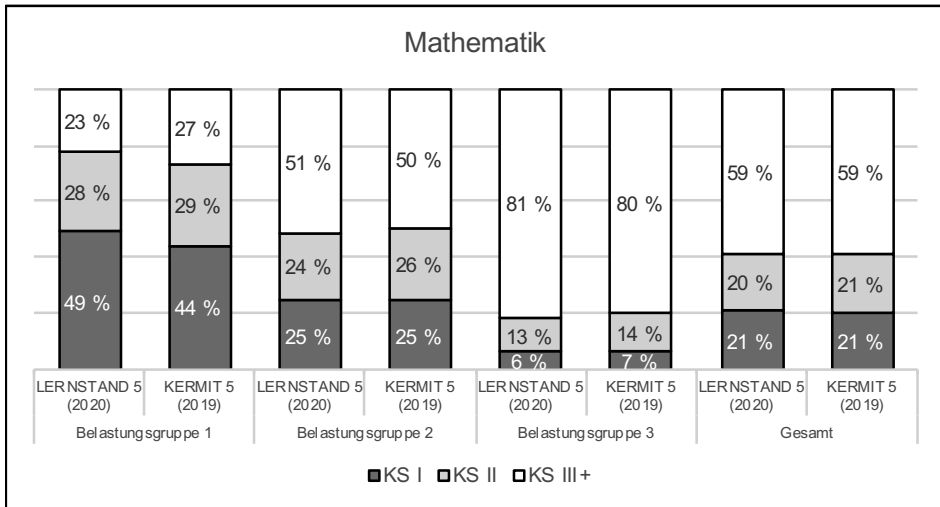
Quelle: eigene Berechnungen

Abb. 11: Vergleich der prozentualen Verteilungen auf die Kompetenzstufen in Deutsch-Leseverstehen zwischen LERNSTAND 5 (2020) und KERMIT 5 (2019)



Quelle: eigene Berechnungen

Abb. 12: Vergleich der prozentualen Verteilungen auf die Kompetenzstufen in Mathematik zwischen LERNSTAND 5 (2020) und KERMIT 5 (2019)

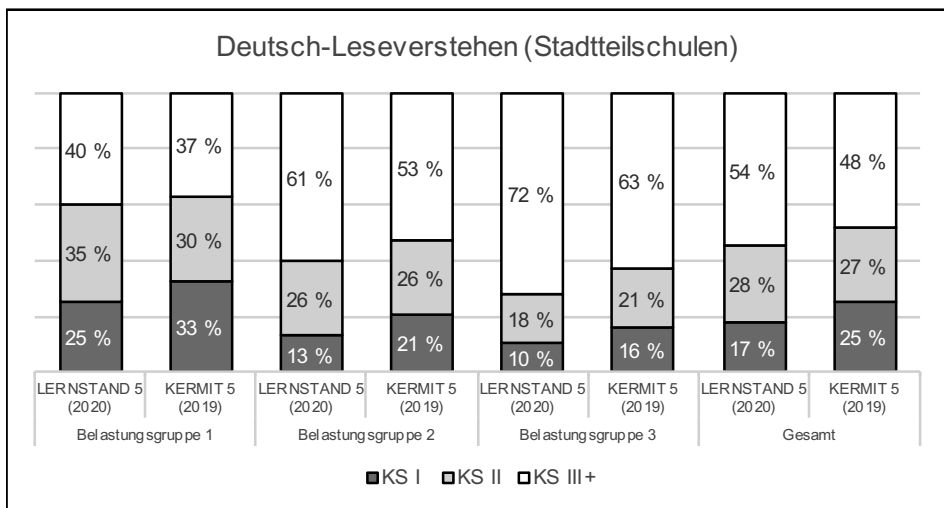


Quelle: eigene Berechnungen

Für die Stadtteilschulen fallen die Kompetenzstufenverteilungen der LERNSTAND-5-Kohorte in Deutsch-Leseverstehen ein wenig günstiger aus. Die größten Schwankungen über alle Kompetenzstufen hinweg sind in Belastungsgruppe 3 zu verzeichnen (vgl. Abb. 13).

In Mathematik verteilen sich die Kompetenzen der Schüler\*innen in beiden Kohorten ungefähr gleich mit den größten Schwankungen über alle Kompetenzstufen hinweg in Belastungsgruppe 1 (vgl. Abb. 14 auf der folgenden Seite).

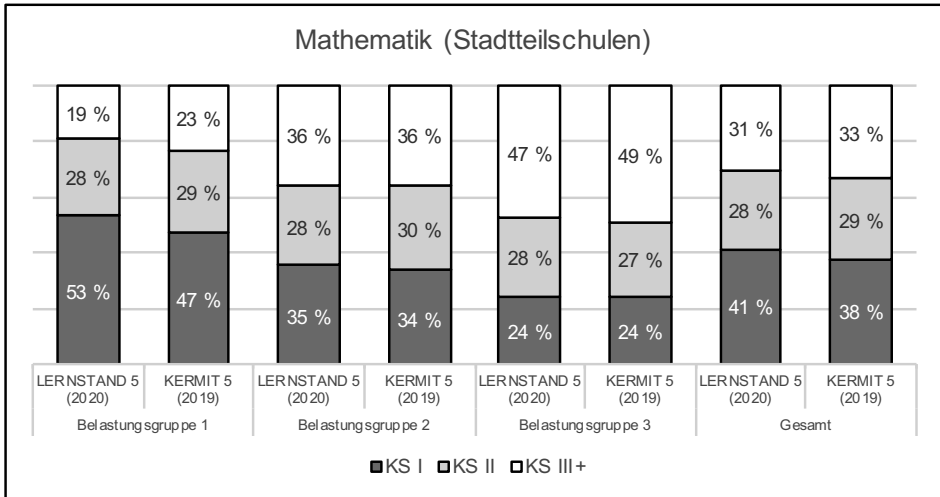
Abb. 13: Vergleich der prozentualen Verteilungen auf die Kompetenzstufen in Deutsch-Leseverstehen zwischen LERNSTAND 5 (2020) und KERMIT 5 (2019) an Stadtteilschulen



Quelle: eigene Berechnungen

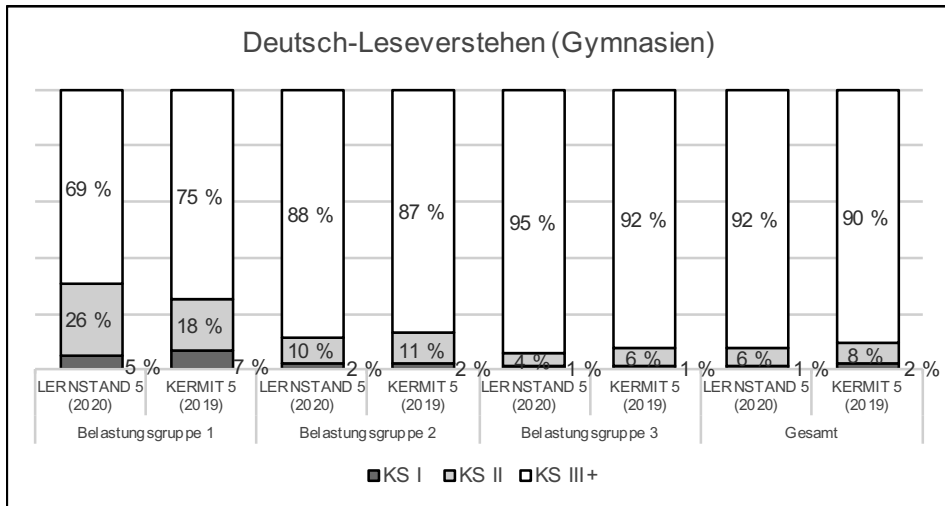
An den Gymnasien verteilen sich die Punktwerte der Schüler\*innen in LERNSTAND 5 und KERMIT 5 (2019) ebenfalls annähernd gleich auf die Kompetenzstufen (vgl. Abb. 15 und Abb. 16 auf S. 73 f.). In der Belastungsgruppe 1 gibt es in beiden Testdomänen die größten Abweichungen. In dieser Belastungsgruppe erreichen etwas mehr Schüler\*innen in KERMIT 5 die Regelstandards; allerdings werden die Mindeststandards auch etwas seltener erreicht. Insgesamt sind die Unterschiede in den prozentualen Verteilungen jedoch als praktisch nicht bedeutsam zu interpretieren, da sie vergleichbar sind mit den üblichen Schwankungen innerhalb der KERMIT-5-Erhebungen vergangener Jahre.

Abb. 14: Vergleich der prozentualen Verteilungen auf die Kompetenzstufen in Mathematik zwischen LERNSTAND 5 (2020) und KERMIT 5 (2019) an Stadtteilschulen



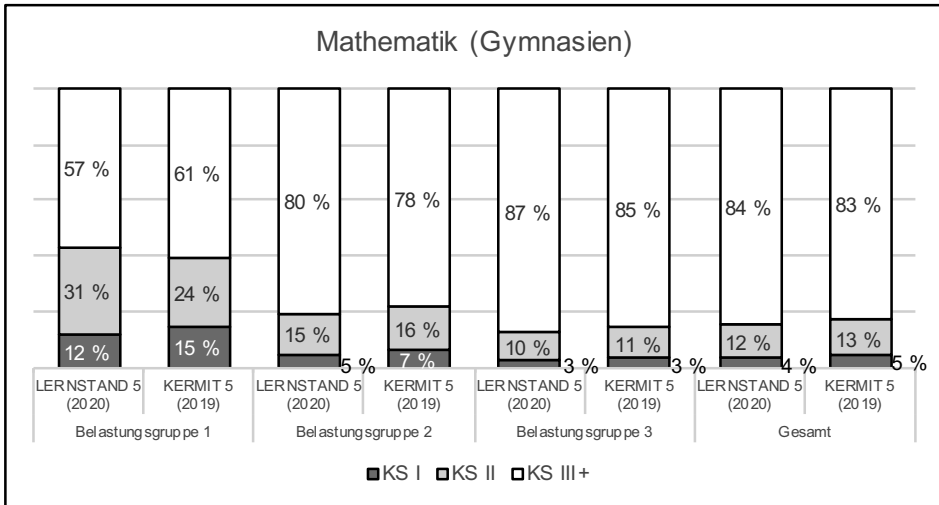
Quelle: eigene Berechnungen

Abb. 15: Vergleich der prozentualen Verteilungen auf die Kompetenzstufen in Deutsch-Leseverstehen zwischen LERNSTAND 5 (2020) und KERMIT 5 (2019) an Gymnasien



Quelle: eigene Berechnungen

Abb. 16: Vergleich der prozentualen Verteilungen auf die Kompetenzstufen in Mathematik zwischen LERNSTAND 5 (2020) und KERMIT 5 (2019) an Gymnasien



Quelle: eigene Berechnungen

## 5 Diskussion

Die Ergebnisse der Analysen sind mehrdeutig. Auf der einen Seite gibt es im Kohortenvergleich aufseiten der LERNSTAND-Erhebungen – mit Ausnahme von Mathematik in LERNSTAND 5 – in allen Bereichen deutlich mehr Schüler\*innen, die zu wenige Aufgaben für die Berechnung eines Kompetenzwerts bearbeitet haben. Welche Gründe dazu geführt haben, dass die Aufgaben unzureichend bearbeitet wurden, ist jedoch nicht bekannt.

Eine naheliegende Vermutung ist, dass die Schüler\*innen die Aufgaben aufgrund zu geringer Kompetenzen nicht lösen konnten und deshalb die Bearbeitung abgebrochen haben. Diese Begründung würde die Ergebnisse von Engzell und Kollegen (2020) stützen. Insbesondere Schüler\*innen aus Stadtteilschulen und aus sozial stark belasteten Schulen zeigen besonders hohe Dropout-Raten.

Geht man dieser Vermutung folgend einmal davon aus, dass diese Schüler\*innen in beiden Kohorten Punktwerte auf Kompetenzstufe I erreicht hätten, und berücksichtigt sie auf diese Weise bei den Auswertungen, so verstärken sich die Unterschiede in den Kompetenzstufenverteilungen für die Primarstufe geringfügig zugunsten von KERMIT 3 (2019). Für die Schüler\*innen in LERNSTAND 5 ergeben sich etwas ungünstigere Kompetenzstufenverteilungen ausschließlich für Deutsch-Leseverstehen und hier besonders für die Belastungsgruppe 1 an Stadtteilschulen. Insgesamt errei-



chen die Kohortenunterschiede zwischen den Kompetenzstufenverteilungen jedoch auch nach der Zusammenfassung der Dropouts zu Kompetenzstufe I keine statistische Bedeutsamkeit.

Es könnten aber auch motivationale Gründe sein, die zu höheren Dropouts geführt haben. Womöglich war die wahrgenommene Belastung in der Testsituation vor dem Hintergrund der veränderten schulischen Bedingungen zu hoch. Die hier aufgeführten Gründe gelten natürlich ebenso für die Schüler\*innen, die eine ausreichende Anzahl an Aufgaben zur Berechnung ihrer Kompetenzen bearbeitet haben.

Dass in beiden Kohorten in der Testdomäne Deutsch-Leseverstehen deutlich höhere Dropoutquoten zu beobachten sind, liegt vermutlich an der Art der Testkonstruktion. Im Gegensatz zur Messung der mathematischen Kompetenz werden für die Erhebung der Lesekompetenz nicht viele in sich abgeschlossene Testaufgaben verwendet. Schüler\*innen, die eine Aufgabe in Mathematik nicht lösen können, werden mit einer höheren Wahrscheinlichkeit zur nächsten Aufgabe weiterblättern und versuchen, sie zu lösen. In Deutsch-Leseverstehen werden jedoch zwei längere Lesetexte vorgelegt, und die Schüler\*innen müssen voneinander unabhängige Fragen zu diesen Texten beantworten. Die Wahrscheinlichkeit, dass ein\*e Schüler\*in die Bearbeitung von gleich mehreren Aufgaben zu einem Lesetext abbricht, den sie\*er nicht versteht, ist vermutlich höher als in Mathematik.

Die Ergebnisse der getesteten Schüler\*innen deuten darauf hin, dass durch die Corona-bedingten Veränderungen im Unterrichtsgeschehen keine größeren Einbußen in den untersuchten Kompetenzen zu verzeichnen sind. Sowohl die durchschnittlich erreichten Kompetenzwerte als auch die Kompetenzstufenverteilungen unterscheiden sich nicht bedeutsam zwischen den Schüler\*innen, die im Jahr 2019 unter normalen Unterrichtsbedingungen getestet wurden, und denen, die sich aufgrund der Pandemie Unterrichtsausfällen und Fernunterricht gegenübergestellt sahen.

Im Hinblick auf die Vergleiche zwischen den Ergebnissen aus KERMIT 3 und LERNSTAND 4 ließe sich argumentieren, dass die Kompetenzen zu Beginn der vierten Klassenstufe etwas höher ausfallen müssten als am Ende von Klassenstufe drei. Es kann also sein, dass die Kompetenzentwicklung der Grundschüler\*innen durch die Corona-bedingten Änderungen etwas gebremst wurde. Diese Annahme wird durch die Befunde aus den Niederlanden gestützt (Engzell et al., 2020).

Andererseits liegen zwischen der regelhaften KERMIT-3-Erhebung im April/Mai und der LERNSTAND-4-Erhebung im August nur ungefähr vier Monate, von denen die Schüler\*innen eine Woche in den Maiferien und sechs Wochen in den Sommerferien waren. Zusätzlich fällt das Pensum an Unterrichtsstoff in der Zeit kurz vor und kurz nach den Ferien erfahrungsgemäß etwas geringer aus. Es ist daher in dieser Zeit nicht von allzu starken Kompetenzzuwächsen auszugehen.

Vielmehr liegt die Vermutung nahe, dass die Schulschließungen relativ gut kompensiert wurden. In Hamburg wurden unterschiedliche Maßnahmen ergriffen, um einer Abnahme der Kompetenzen der Schüler\*innen in den Kernfächern Deutsch und Mathematik entgegenzuwirken. Eine kompensatorische Maßnahme bestand in dem Angebot von insgesamt 381 zentral organisierten Lerngruppen in den Sommerferien („Lernferien“). Mehr als ein Drittel aller Hamburger Schulen beteiligte sich an diesem Angebot, das zum größten Teil (87 %) an Schulen in sozial benachteiligten Stadtteilen stattfand (Behörde für Schule und Berufsbildung, 2020). Die Lernferien wurden von Hamburgs Schüler\*innen in hohem Maße genutzt.

Eine weitere Vermutung ist, dass viele Schüler\*innen auch mit alternativen Unterrichtsformen – zum Beispiel auf digitalen Unterrichtsplattformen – gut umgehen und ihre Kompetenzen in Deutsch-Leseverstehen und Mathematik daher weiterentwickeln bzw. aufrechterhalten konnten. Zusätzlich sei hier die lernbezogene Unterstützung vieler Familien erwähnt, die kompensatorisch gut gewirkt zu haben scheint.

Dass der (digitale) Fernunterricht in Hamburg einigermaßen gut funktioniert hat, zeigt eine Online-Befragung des IfBQ, die vor den Sommerferien durchgeführt wurde. Rund 13.900 Sorgeberechtigte, 3.200 Pädagog\*innen und 3.300 Schüler\*innen wurden zu ihrer Beurteilung der aktuellen Unterrichtssituation und des Fernunterrichts befragt. Die Untersuchung kam zwar zu dem Ergebnis, dass alle befragten Personengruppen den Fernunterricht als anstrengend empfanden. Allerdings zeigte sich auch, dass die digitale Ausstattung für den Fernunterricht bei nahezu allen Schüler\*innen vorhanden ist (99,6 %) und knapp 80 Prozent von ihnen nach Einschätzung der Pädagog\*innen auch in der Lage sind, im Fernunterricht zu lernen (Brändle & Albers, 2020).

Die Erhebungszeiträume in LERNSTAND 5 und KERMIT 5 waren nicht unterschiedlich, sodass aus diesen Gründen keine Unterschiede in den Kompetenzständen der Schüler\*innen zu erwarten waren. Jedoch wurde LERNSTAND 5 im Gegensatz zu KERMIT 5 nicht von externen Testleitungen, sondern von Lehrkräften durchgeführt. Dieser Umstand kann dazu führen, dass die Kompetenzen der Schüler\*innen überschätzt werden. Denn es kann nicht ausgeschlossen werden, dass eine vertraute Person, die eventuell unabsichtlich unterstützend wirkt, zu besseren Leistungen der Schüler\*innen beiträgt. In diesem Fall wären höhere Kompetenzwerte in LERNSTAND 5 zu erwarten. Diese zeigen sich jedoch nicht.

Zum Zeitpunkt der LERNSTAND-5-Testung besuchten die Schüler\*innen jedoch erst seit wenigen Wochen die weiterführende Schule. Die Lehrkräfte waren demnach noch nicht so sehr mit den Schüler\*innen und ihren Stärken und Schwächen vertraut. Mögliche Effekte der Unterstützung durch die Lehrkräfte dürften vor diesem Hintergrund eher gering ausgefallen sein.

Die Schüler\*innen in LERNSTAND 5 erzielten insbesondere in Mathematik geringere Punktwerte, die sich am deutlichsten an sozial stärker belasteten Schulen zeigen. Ein möglicher Grund für die durchgehend schwächeren Leistungen in Mathematik könnte darin liegen, dass Sorgeberechtigte womöglich mit ihren Kindern stärker das Lesen geübt haben als Mathematikaufgaben. Eine Förderung des Leseverstehens scheint leichter umsetzbar als der Aufbau eines mathematischen Grundverständnisses. Allerdings sind die Unterschiede zwischen den Kohorten aus den Jahren 2019 und 2020 nicht besonders groß. Es lässt sich daher vermuten, dass auch in den fünften Klassen der Hamburger Schulen keine größeren Auswirkungen der Corona-bedingten Veränderungen festzustellen sind.

Dieser Befund kann zum einen dadurch bedingt sein, dass in der vierten Jahrgangsstufe – also für die Kinder, die nach den Sommerferien an LERNSTAND 5 teilnahmen – der Unterrichtsschwerpunkt auf die Kernfächer Deutsch und Mathematik gelegt wurde. Außerdem nahmen diese Kinder nach den Maiferien wieder täglich am Präsenzunterricht teil.

Ein weiterer Grund, der eventuell einen positiven Effekt auf die Leistungen in beiden LERNSTAND-Erhebungen gehabt haben könnte, ist in der veränderten Motivationslage der Schüler\*innen zu sehen. Die meisten Schüler\*innen gaben in der Online-Umfrage des IfBQ an, sich (eher) über die Schulöffnung zu freuen (Brändle & Albers, 2020). Faktoren, die sich negativ auf die Testergebnisse ausgewirkt haben könnten, sind hingegen in der veränderten Situation in der Schule (Abstands- und Hygieneregeln) oder in der als belastend wahrgenommenen Pandemiesituation zu vermuten (vgl. Engzell et al., 2020).

Insgesamt scheint es so, als hätte das Hamburger Schulsystem die befürchteten negativen Auswirkungen der Pandemie auf die Kompetenzen der Schüler\*innen in Deutsch-Leseverstehen und Mathematik zunächst gut abfangen können. Es bleibt jedoch abzuwarten, ob die Corona-bedingten Veränderungen im Unterrichtsgeschehen langfristig Folgen für die Kompetenzentwicklung der Schüler\*innen haben werden. Neben den Auswirkungen auf fachliche Kompetenzen ist ein weiterer wichtiger Untersuchungsgegenstand in der psychosozialen Entwicklung der Schüler\*innen zu sehen. Dieser Forschungsdesiderate sollten sich künftige Untersuchungen annehmen.

## Literatur und Internetquellen

- Behörde für Schule und Berufsbildung (2020, 13.07.). *Ferienunterricht. Hamburger Lernferien gestartet*. Pressemeldung. Zugriff am 18.12.2020. Verfügbar unter: <https://www.hamburg.de/bsb/pressemitteilungen/14095076/2020-07-13-bsb-hamburger-lernferien/>.
- Brändle, T., & Albers, A. (2020). *Befragung Lernen in Zukunft: Ergebnisse einer Online-Befragung von Eltern, Pädagoginnen und Pädagogen sowie Schülerinnen und Schülern in Hamburg vom 29. Mai bis 7. Juni 2020*. Hrsg. von der Freien und Hansestadt Hamburg, Behörde für Schule und Berufsbildung, Institut für Bildungsmonitoring und Qualitätsentwicklung. Zugriff am 18.12.2020. Verfügbar unter: <https://www.hamburg.de/contentblob/13982476/734c18989f94decc5fa25d5ef3d82e6d/data/bliz-ergebnisse-richt.pdf>.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2. Aufl.). Hillsdale, NJ: Erlbaum. <https://doi.org/10.4324/9780203771587>
- Depping, D. (2020). Kompetenzen von Schülerinnen und Schülern. In Behörde für Schule und Berufsbildung (Hrsg.), *Bildungsbericht Hamburg 2020* (HANSE – Hamburger Schriften zur Qualität im Bildungswesen, Bd. 18) (S. 141–154). Münster: Waxmann.
- Engzell, P., Frey, A., & Verhagen, M. D. (2020). *Learning Inequality during the Covid-19 Pandemic*. <https://doi.org/10.31235/osf.io/ve4z7>
- KMK (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland). (2015). *Gesamtstrategie der Kultusministerkonferenz zum Bildungsmonitoring*. Beschluss der 350. Kultusministerkonferenz vom 11.06.2015. Zugriff am 18.12.2020. Verfügbar unter: [https://www.kmk.org/fileadmin/Dateien/pdf/Themen/Schule/Qualitaets-sicherung\\_Schulen/2015\\_06\\_11-Gesamtstrategie-Bildungsmonitoring.pdf](https://www.kmk.org/fileadmin/Dateien/pdf/Themen/Schule/Qualitaets-sicherung_Schulen/2015_06_11-Gesamtstrategie-Bildungsmonitoring.pdf).
- Lücken, M. (2017). Kompetenzen von Schülerinnen und Schülern. In Behörde für Schule und Berufsbildung (Hrsg.), *Bildungsbericht Hamburg 2017* (HANSE – Hamburger Schriften zur Qualität im Bildungswesen, Bd. 16.) (S. 135–147). Münster: Waxmann.
- Lücken, M., Thonke, F., Pohlmann, B., Hofmann, H., Golecki, R., Rosendahl, J., Benzing, M., & Poerschke, J. (2014). KERMIT – Kompetenzen ermitteln. In D. Fickermann & N. Maritzen (Hrsg.), *Grundlagen für eine daten- und theoriegestützte Schulentwicklung. Konzeption und Anspruch des Hamburger Instituts für Bildungsmonitoring und Qualitätsentwicklung (IfBQ)* (HANSE – Hamburger Schriften zur Qualität im Bildungswesen, Bd. 13) (S. 127–154). Münster: Waxmann.
- Schulte, K., Hartig, J., & Pietsch, M. (2014). Der Sozialindex für Hamburger Schulen. In D. Fickermann & N. Maritzen (Hrsg.), *Grundlagen für eine daten- und theoriegestützte Schulentwicklung. Konzeption und Anspruch des Hamburger Instituts für Bildungsmonitoring und Qualitätsentwicklung (IfBQ)* (HANSE – Hamburger Schriften zur Qualität im Bildungswesen, Bd. 13) (S. 67–80). Münster: Waxmann.

Denise Depping, geb. 1984, wissenschaftliche Referentin am Institut für Bildungsmonitoring und Qualitätsentwicklung (IfBQ) in Hamburg.

E-Mail: [denise.depping@ifbq.hamburg.de](mailto:denise.depping@ifbq.hamburg.de)

Markus Lücken, Dr., geb. 1971, Referatsleiter am Institut für Bildungsmonitoring und Qualitätsentwicklung (IfBQ) in Hamburg.

E-Mail: [markus.luecken@ifbq.hamburg.de](mailto:markus.luecken@ifbq.hamburg.de)

*Frank Musekamp*, Dr., geb. 1978, wissenschaftlicher Referent am Institut für Bildungsmonitoring und Qualitätsentwicklung (IfBQ) in Hamburg.  
E-Mail: frank.musekamp@ifbq.hamburg.de

*Franziska Thonke*, geb. 1978, wissenschaftliche Referentin am Institut für Bildungsmonitoring und Qualitätsentwicklung (IfBQ) in Hamburg.  
E-Mail: franziska.thonke@ifbq.hamburg.de

Korrespondenzadresse: Freie und Hansestadt Hamburg, Institut für Bildungsmonitoring und Qualitätsentwicklung (IfBQ), Abteilung BQ 2, Referat Kompetenzmessung und Evaluation, Beltgens Garten 25, 20537 Hamburg